

Cómo preservar mi patrimonio digital personal

JUAN VOUTSSAS M.



La presente obra está bajo una licencia de:

http://creativecommons.org/licenses/by-nc-sa/3.0/deed.es_MX



Atribución-No Comercial-Licenciamiento Recíproco 3.0 Unported

Eres libre de:



copiar, distribuir y comunicar públicamente la obra



hacer obras derivadas

Bajo las condiciones siguientes:



Atribución — Debes reconocer la autoría de la obra en los términos especificados por el propio autor o licenciente.



No comercial — No puedes utilizar esta obra para fines comerciales.



Licenciamiento Recíproco — Si alteras, transformas o creas una obra a partir de esta obra, solo podrás distribuir la obra resultante bajo una licencia igual a ésta.

Esto es un resumen fácilmente legible del:
[texto legal \(de la licencia completa\)](#)

En los casos que sea usada la presente obra, deben respetarse los términos especificados en esta licencia.



*Cómo preservar mi patrimonio
digital personal*

COLECCIÓN
TECNOLOGÍAS DE LA INFORMACIÓN
Instituto de Investigaciones Bibliotecológicas y de la Información

*Cómo preservar mi patrimonio
digital personal*

Juan Voutssas M.



**Universidad Nacional Autónoma de México
2013**

Z701.3 Voutssas Márquez, Juan
C65V69 Cómo preservar mi patrimonio digital
personal / Juan Voutssas Márquez. - México : UNAM, Instituto de Investigaciones
Bibliotecológicas y de la Información,
2013.
viii, 280 p. - (Tecnologías de la información)
ISBN: 978-607-02-4122-2

1. Documentos Electrónicos - Desarrollo
2. Preservación Digital
3. Archivos Electrónicos-Preservación
4. Archivos Personales- Preservación
5. Seguridad Informática - Archivos Electrónicos I. t. II. ser.

Diseño de portada: Mario Ocampo Chávez

Primera edición 2013

D.R. © UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Ciudad Universitaria, 04510, México, D.F.

Impreso y hecho en México

ISBN: 978-607-02-4122-2

Contenido

INTRODUCCIÓN	1
CUESTIONARIO	1
LA EVOLUCIÓN DE LOS DISPOSITIVOS DE ALMACENAMIENTO.	3
EL CRECIMIENTO DE LA INFORMACIÓN DIGITAL	9
LA PROBLEMÁTICA DE LA PRESERVACIÓN DOCUMENTAL DIGITAL	13
LOS FACTORES DE LA PRESERVACIÓN DIGITAL	17

LA NATURALEZA DE LOS DOCUMENTOS DIGITALES. CONCEPTOS, FORMATOS Y DISPOSITIVOS

CONCEPTOS BÁSICOS	21
Documento	21
Archivo. Computación	22
Archivo. Ciencia archivística	23
Dispositivos y documentos analógicos	23
Dispositivos y documentos digitales	29
FORMATOS.	39
Formatos de texto	41
Formatos de imagen	46
Imágenes de mapa de bits	46
Imágenes vectoriales	56
Compresión de imágenes	58
Formatos de imagen más utilizados.	67
Formatos de audio	72
Formatos de video	79
OCR. <i>Optical Character Recognition</i> o Reconocimiento Óptico de Caracteres	83
ATRIBUTOS DE LOS DOCUMENTOS DIGITALES	89
Permanencia.	90
Accesibilidad	91
Calidad y valor de uso	93

Autenticidad	97
Confidencialidad	98
Versión.	98
DIGITALIZACIÓN	103
PRESERVACIÓN Y CONSERVACIÓN DOCUMENTAL DIGITAL	105
<i>COPYRIGHT</i>, DERECHOS DE AUTOR Y <i>CREATIVE COMMONS</i>	109

SEGURIDAD INFORMÁTICA

CONCEPTOS FUNDAMENTALES DE LA SEGURIDAD INFORMÁTICA	122
Recursos informáticos	122
Amenazas informáticas.	123
Riesgos informáticos	125
Vulnerabilidades informáticas	125
Impactos.	126
Principio básico de la seguridad informática.	128
METODOLOGÍAS PARA LA SEGURIDAD INFORMÁTICA PERSONAL	129
Respecto a la computadora	129
Respecto al uso	130
Respecto a las contraseñas o <i>passwords</i>	131
Respecto a los correos electrónicos.	136
Respecto a las redes	137
Respecto a las redes sociales	138
Respecto a los dispositivos móviles.	140

LA PRESERVACIÓN DE NUESTROS ARCHIVOS DIGITALES

RECOMENDACIONES PARA LA PRESERVACIÓN DOCUMENTAL DIGITAL	148
1. Establece las bases de tu plan de preservación y conservación de tus documentos digitales	149
2. Selecciona los formatos que maximicen la accesibilidad de tus archivos a lo largo del tiempo	155
3. Selecciona el equipo y programas de cómputo que puedan producir los materiales digitales en los formatos que aseguren la mayor probabilidad de permanecer accesibles a lo largo del tiempo	158
4. Organiza tus materiales digitales en agrupamientos lógicos	163

5. Identifica adecuadamente tus materiales digitales	170
6. Respalda tus materiales digitales con frecuencia y con método	175
7. Establece, controla y protege tus versiones adecuadamente	188
8. Toma acciones periódicamente contra la obsolescencia de equipo y programas	192
9. Verifica siempre que cada respaldo haya terminado correctamente.	197
10. Refina permanentemente tu sistema de respaldos.	200
 CONCLUSIONES	 203
 GLOSARIO	 209
 REFERENCIAS BIBLIOGRÁFICAS.	 217
 ANEXOS.	 223
 ANEXO 1. CÓMO CONTAR MÚLTIPLOS DE BYTES	 223
 ANEXO 2. PRODUCCIÓN MUNDIAL ESTIMADA DE CONTENIDOS ORIGINALES, ALMACENADOS DIGITALMENTE.	 224
 ANEXO 3. EL SISTEMA BINARIO	 225
 ANEXO 4. TABLA DE CARACTERES ASCII	 238
 ANEXO 5. EJEMPLO DEL TEXTO LEGAL DE UNA LICENCIA <i>CREATIVE COMMONS</i>	 241
 ANEXO 6. PRINCIPALES MODELOS E INICIATIVAS EN PRESERVACIÓN ARCHIVÍSTICA PARA ORGANIZACIONES A NIVEL MUNDIAL.	 251

Introducción

Sin excepción, todos los que hoy manejan información digital aprenderán el valor de los respaldos. Algunos lo harán por la buena, antes de que el desastre suceda. Otros lo aprenderán después.

T. E. Ronnenberg

CUESTIONARIO



Te has hallado en alguna de las siguientes situaciones?

- ❖ Repentinamente descubro que cantidades importantes de documentos que estaban en mi computadora –textos, fotos, música, etc.– han desaparecido y no se hallan más ahí.
Sí___ No___
- ❖ Me cuesta mucho trabajo navegar entre directorios y subdirectorios en mi computadora para localizar cierto archivo en especial.
Sí___ No___
- ❖ Al pasar información entre computadora de escritorio o portátil, dispositivos móviles –iPod, iPad, teléfono, Smartphone, Blackberry, tableta, etc.– descubro que cierto archivo necesario o deseado ya no existe en ninguna de ellas.
Sí___ No___
- ❖ Cuando busco cierto documento encuentro que tengo múltiples versiones del mismo y no puedo distinguir cuál es la última o la que utilicé para cierto evento.
Sí___ No___
- ❖ Tengo múltiples copias de archivos idénticos duplicados en varios discos y subdirectorios de la computadora.
Sí___ No___

Cómo preservar mi patrimonio digital personal

❖ Cuando deseo abrir un cierto archivo antiguo descubro que no tengo ya el programa que lo abre o que el que tengo no puede abrir versiones tan antiguas.

Sí____ No____

❖ El sistema operativo se ha dañado, no puedo arrancar mi computadora y por tanto no puedo acceder a ninguno de mis archivos en ella.

Sí____ No____

Y la peor de todas...

❖ El disco duro de mi computadora se ha dañado y todo lo que estaba ahí no puede ser recuperado.

Sí____ No____

Si has contestado “sí” al menos a una de estas preguntas, definitivamente este libro es para ti. Eres una de los millones de personas que se han integrado a las Tecnologías de Información y Comunicaciones –TIC– del mundo actual, y fuiste de alguna forma aconsejado y estimulado con sus enormes –e indudables– ventajas, pero sin haber sido advertido adecuadamente de sus riesgos y desventajas. A este respecto pueden hacerse reflexiones sociales, políticas, culturales, etc.; no es el caso de este libro. El propósito del mismo es repensar acerca de las reflexiones tecnológicas al respecto, pero en especial de aquellas metodologías o “buenas prácticas” que tienen que ver con la preservación o pérdida de los documentos digitales que poseemos en nuestras computadoras personales y dispositivos móviles (en casa, oficina, tableta, teléfono, etc.) así como la de aquellos documentos que de algún modo compartimos con otras personas, bien porque queremos publicar información, bien porque deseamos compartirla en redes sociales.

La destrucción, confusión o preservación de archivos digitales o electrónicos no tiene por qué ser resultado de la casualidad o la suerte y no depende en su mayor parte de productos y servicios que se adquieren. Existen principios, técnicas y metodologías que nos permiten minimizar su riesgo, destrucción y confusión y maxi-

mizar su preservación así como su correcta y eficiente administración. Como sucede con muchas otras cosas, es necesario conocer estas herramientas, familiarizarse con su uso y hacerlas un hábito para que realmente se conviertan en elementos eficientes de nuestro quehacer cotidiano en lo concerniente al manejo adecuado de nuestros archivos digitales. Su conocimiento y explotación no está reservada a corporaciones y organizaciones con áreas y personal especializados al efecto, ni se requiere de estudios especiales en informática para dominarlas. Pero una vez aprendidas requieren de forma indispensable de constancia y método –volverlas un hábito– para que funcionen y den resultado. Si sólo se aprenden pero no se siguen, o se siguen muy eventualmente, no darán ningún resultado útil. No se culpe entonces a la herramienta por su uso descuidado. En palabras de Stephen Hawking: “el mayor enemigo del conocimiento no es la ignorancia; es la ilusión de conocimiento”.

LA EVOLUCIÓN DE LOS DISPOSITIVOS DE ALMACENAMIENTO

En el futuro, las computadoras no pesarán más de una y media toneladas.

Revista Mecánica Popular, 1949

El advenimiento de las Tecnologías de Información y Comunicaciones –TIC– ha traído muchos y muy sorprendentes hechos en las últimas décadas; entre ellos, el que la información sea producida y almacenada en la actualidad mayormente en forma digital. Con ayuda de múltiples programas y accesorios al efecto, la información puede ser creada y revisada rápida y fácilmente; gracias al Internet puede ser transmitida y distribuida a una velocidad pasmosa. La ininterrumpida mejora y crecimiento en la capacidad de los dispositivos para almacenamiento de información al tiempo que la reducción de los costos han creado variados dispositivos y medios para almacenamiento de información, que ponen hoy en día al alcance de personas y organizaciones la capacidad de almacenar millones y millones de bytes de información de todo tipo: textos,

Cómo preservar mi patrimonio digital personal

imágenes, música, video, etc.; todo a un costo muy económico y que tiende invariablemente a disminuir.

El último medio siglo ha visto cómo el almacenamiento de datos –tanto magnético como óptico– ha logrado maravillas en su relación costo/beneficio de una manera casi inverosímil; a tal grado, que no podemos simplemente hablar de un fenómeno de “comodidad” tecnológica o simple “mejora de una capacidad”. Es pertinente reflexionar acerca del hecho trascendental de que pocas cosas en la historia reciente del género humano han tenido un crecimiento en capacidad y un abaratamiento en costo tan espectaculares como el almacenamiento opto-magnético. Es tan relevante este fenómeno, que ha cambiado en unos pocos años radicalmente el concepto social de lo que, en lo tocante a información, es almacenable y lo que no.

Analicemos el primer dispositivo de disco duro comercial que existió: la unidad de disco *disk drive* IBM 350, que comenzó a comercializarse en septiembre del año 1956 con la –entonces asombrosa– capacidad de 3.75 Megabytes actuales.¹ El dispositivo no era exactamente un disco duro; en realidad era un paquete o conjunto de 50 discos de 24 pulgadas –61 cm– de diámetro cada uno insertados en un eje central común, sumando 100 “caras” o superficies útiles en total. Sólo el gabinete que contenía los discos –no la computadora entera– pesaba más de una tonelada. Medía 1.73 m de altura, 1.52 m de frente y 74 cm de fondo; prácticamente dos metros cúbicos. En la época en que apareció se rentaba, no vendía, por 3,200 dólares mensuales –dólares de 1956–. Para tener una referencia, en ese entonces el Ford Fairline Sunliner Convertible Coupé, el modelo más caro de esa línea de autos familiares, se vendía por 2,359 dólares; el lujoso Ford Thunderbird se vendía ese año por 3,408 dólares. Para tener un valor de lo que hubiera sido la compra de esa unidad de disco, si sumáramos su renta a tres años obtendremos un costo total mayor de 115,000 dólares: poco más de 31,000 dólares por cada Megabyte de almacenamiento durante ese periodo.

1 1 Megabyte = 1'000,000 = 10⁶ bytes o caracteres. Véase Anexo 1.

Ya para 1980 la empresa Morrow Systems vendía un disco duro de 26 Megabytes de capacidad por “sólo” 5,000 dólares: 193 dólares por Megabyte. Un poco después aparecieron los primeros discos duros para PC, de la empresa Seagate Technologies, conocidos como “Winchester”: el modelo ST-506 tenía una capacidad de cinco Megabytes y se vendía por aproximadamente 1,750 dólares. Medía 8 x 15 x 20 cm; algo así como 2.4 l de volumen. El costo unitario era por tanto de aproximadamente 350 dólares por cada Megabyte de almacenamiento. En 1981, la empresa Apple mercadeaba sus primeros dispositivos de almacenamiento masivo de 5 Megabytes por 3,500 dólares, 700 dólares por Megabyte.

Diez años después, a principios de la década de 1990, se compraba un disco duro de la marca Western Digital de 40 Megabytes por 1,200 dólares. El costo unitario se abatía entonces a 30 dólares por Megabyte de almacenamiento. Para el año 2000 se podían comprar ya Gigabytes² –mil Megabytes– a un costo unitario de aproximadamente 1.80 dólares por Megabyte. En el año 2012 podemos comprar discos de cuatro Terabytes de capacidad, –igual a cuatro mil Gigabytes o cuatro millones de Megabytes– por aproximadamente 400 dólares, a un costo unitario aproximado de un diezmilésimo de dólar = 0.0001 dólares por Megabyte. Esto significa que podemos comprar por el mismo dinero más de doscientos sesenta millones de veces lo que se podía comprar hace 50 años. Aun quitando la depreciación inflacionaria ocurrida en estas cinco décadas –1 dólar de 1956 = 8.46 dólares de 2012– por el mismo dinero en valor real podemos comprar casi 32 millones de veces más capacidad de almacenamiento. Esa es la proporción en la que ha crecido en medio siglo la capacidad de almacenamiento a un cierto costo, sin mencionar la reducción de espacio.

Siguiendo con las tendencias actuales, esto significa que dentro de unos seis o siete años más estaremos comprando un Petabyte³

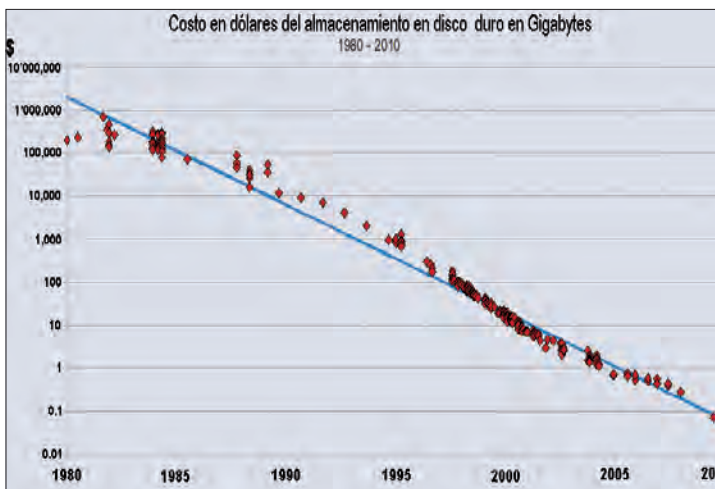
2 1 Gigabyte = 1000 Megabytes = 1'000,000,000 = 10^9 bytes o caracteres. 1 Terabyte = 1000 Gigabytes = 1'000,000 Megabytes = 1'000,000,000,000 = 10^{12} bytes o caracteres.

3 1 Petabyte = 1,000 Terabytes = 1'000,000 Gigabytes = 1'000,000,000 Megabytes = 10^{15} bytes o caracteres.

Cómo preservar mi patrimonio digital personal

–mil Terabytes o un millón de Gigabytes o mil millones de Megabytes o 10^{15} bytes o un uno seguido por quince ceros de bytes– por algo así como 500 dólares. Para visualizar esta capacidad, en un Petabyte pueden almacenarse más de 300 millones de canciones de 3 minutos cada una en formato mp3 o 30 millones de canciones en calidad CD, o 300,000 películas de 90 minutos cada una, en alta resolución, o 500 millones de fotografías de buena resolución de una cámara actual, o 500,000 millones de cuartillas de texto en formato txt.

En términos generales, la capacidad de almacenamiento se ha duplicado cada catorce meses. De hecho, se han hecho regresiones numéricas para tratar de ajustar el costo del almacenamiento a lo largo de tres décadas y puede observarse que éste se ha mantenido muy constante:⁴



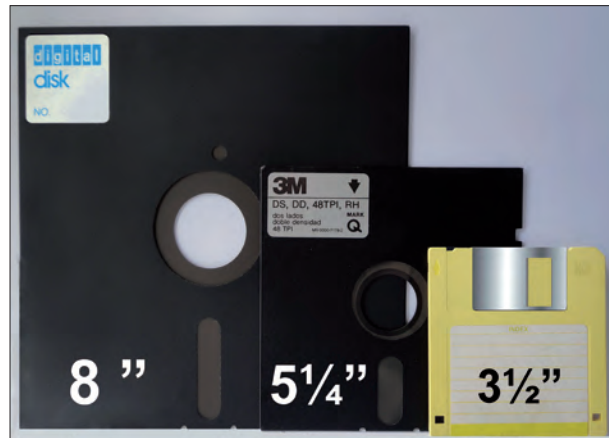
Tomado de: Matthew Komorowski, *A History of Storage Cost*.

De acuerdo con este autor, la ecuación de la regresión numérica puede definirse como:

$$\text{Costo por Gb} = 10^{-0.2502(\text{año}-1980)+6.304}$$

⁴ Matthew Komorowski, *A History of Storage Cost* [en línea].

Toda la reflexión anterior ha sido hecha utilizando los dispositivos de discos duros o masivos, propios para ser instalados dentro de las computadoras, pero conviene repetir ese análisis con los dispositivos portátiles para ver si se han comportado de una manera parecida:



Las tres generaciones de disquetes o discos flexibles.

En 1971, IBM introdujo el primer disco “portátil”; el disquete de ocho pulgadas de diámetro o simplemente 8". En esa época tenía cien Kilobytes –un décimo de Megabyte– de capacidad y costaba alrededor de tres dólares la pieza. En 1979 se fabricaron los últimos de ese tamaño los cuales ya podían almacenar 500 Kb o medio Mb. La empresa Shugart introdujo en 1976 el disquete de cinco y cuarto pulgadas mismo que podía contener cien Kilobytes; se fabricó hasta principios de los ochentas llegando a guardar hasta 1.2 Megabytes. En 1980, Sony presentó el disquete de tres y media pulgadas que en esa época contenía hasta 400 Kilobytes. Se fabricó hasta fines de la década de 1990 llegando a contener hasta 1.4 Megabytes. A fines de los años ochenta, se introdujo el CD-ROM para almacenar seiscientos cincuenta Megabytes en un disco que hoy vale veinte centavos de dólar, y a la fecha se almacenan más de cuatro Gigabytes en un DVD de cuarenta centavos de dólar.

Cómo preservar mi patrimonio digital personal

Cuarenta mil veces en treinta y cinco años por un cierto precio. Aun considerando que un dólar de 1970 es equivalente a 5.93 dólares de 2012, todavía es como siete mil veces mayor la capacidad de almacenamiento adquirida por el mismo precio. Proporciones similares existen en las cintas, cartuchos y otros dispositivos parecidos y recientemente en las memorias portátiles de estado sólido, conocidas como memorias flash USB; la tendencia se mantiene. La empresa Kingston ha anunciado recientemente sus nuevos modelos de estas pequeñas memorias de la llamada serie 300, las cuales tienen capacidades de 128 y 256 Gigabytes, en su conocido tamaño “llavero”. Esta última puede contener dentro de sí 128,000 fotografías de 6 Megapíxeles cada una, u 85,000 canciones mp3 de 3 minutos cada una, o 224 películas mpeg4 comprimidas o 21 películas en alta definición mkv.

Actualmente, el mercado de discos magnéticos vende discos cuya capacidad de almacenamiento oscila alrededor de 500 Gigabytes por pulgada cuadrada –en un disco que además sólo pesa cien gramos y ocupa menos de cien centímetros cúbicos–. Como dato adicional y curioso, esto nos da la medida actual de un bit: $5 \times 20 \text{ nm}$ –nanómetros–⁵ en los discos magnéticos más recientes. El almacenamiento magnético actual cae ya totalmente en el campo de la nanotecnología, es decir, partículas menores a 100 nm. El límite actual del almacenamiento magnético está ya determinado en la práctica por una frontera física fundamental, consistente en el límite en el que la energía térmica de los materiales excede a la energía magnética almacenada en cada pequeño grano del material sensible; si el tamaño del grano magnético disminuye a cierto tamaño, el calor creado y almacenado por el dispositivo en sí mismo puede desmagnetizarlo.⁶ Empero, ya se vislumbran desde luego algunas innovaciones en puerta que permitirán evadir los efectos nocivos de este límite, no sólo con nuevos materiales sino con

5 Un nanómetro (nm) = 10^{-9} metros; esto es la milmillonésima parte de un metro, o la millonésima parte de un milímetro.

6 Zvonimir Bandić, Dmitri Litvinov y M. Rooks, “Nanostructured Materials in Information Storage” [en línea].

nuevas maneras de estructurarlos. La empresa Seagate anunció ya su nueva generación de discos magnéticos que utilizan una nueva tecnología llamada *Heat-Assisted Magnetic Recording* –HAMR– o Grabación Magnética Ayudada por Calor, una tecnología de grabación de última generación que sustituirá a la actual denominada *Perpendicular Magnetic Recording* o Grabación Magnética Perpendicular. Se habla además entre otras teorías de los nanoláseres y del almacenamiento holográfico, lo que permitiría de entrada cuadruplicar las capacidades actuales existentes en el almacenamiento de información. Se habla también de que un bit de memoria podrá ser almacenado en el “espín” o momento angular intrínseco de un sólo electrón. Esto implicará por tanto que continúe la tendencia hacia dispositivos de mayor capacidad a la actual y a precios todavía menores por Megabyte que los actuales.

Con respecto al almacenamiento en servidores a renta o en “la nube”, un estudio de la organización International Data Corporation –IDC– reportó que el costo de almacenar un Gigabyte ha descendido de 18.95 dólares en 2005 a 1.50 dólares en 2011 y se espera se reduzca a 0.66 dólares para 2015.⁷

EL CRECIMIENTO DE LA INFORMACIÓN DIGITAL

Por todo lo anterior, y dado los enormes volúmenes de almacenamiento que se pueden comprar fácilmente hoy en día, observamos el fenómeno de grandes cantidades de información almacenada actualmente por parte de personas y organizaciones. Como se puede constatar, tal abundancia de almacenamiento aunado al mínimo espacio que ocupa y a su bajo precio ha causado un cambio radical en la forma en que en la actualidad percibimos al almacenamiento de información de todo tipo: textos, imágenes, música, videos, bases de datos, etc., pueden ser almacenados hoy en día en cantidades y a costos que eran imposibles de concebir hace una o

⁷ International Data Corporation (IDC), *The 2011 IDC Digital Universe Study* [en línea].

dos décadas. Pero de la misma manera que se ha incrementado la facilidad para almacenar información ha crecido el problema de preservarla en la misma medida que crece el riesgo de dañarla o perderla. Es de suma importancia notar y recordar que almacenar no es lo mismo que preservar.

Junto con esta facilidad para adquirir dispositivos para almacenamiento de información se han diversificado notoriamente las tecnologías para crear, capturar, transmitir, administrar y reproducir todo tipo de información: dispositivos, programas, estándares, formatos, metodologías, estrategias, personal entrenado, etc. Por lo mismo, en la actualidad se producen y almacenan ya muchos Exabytes⁸ de información cada año.

De acuerdo con los estudios de Lyman y Varian⁹ se produjo entre uno y dos Exabytes de información durante el año 2000; de acuerdo con otro estudio de estos mismos autores actualizado en 2003 la cantidad para ese año ascendía entre tres y cinco Exabytes. Un reciente estudio de John Ganz¹⁰ pronosticaba que para fines del año 2010, y como resultado de lo generado sólo en ese año, el mundo estaría produciendo ya 998 Exabytes, prácticamente un Zettabyte¹¹ de información.

En el más reciente estudio al respecto de la organización IDC se establece que en lo que va de la década de 2005 a 2015 el mundo produjo 132 Exabytes en 2005, 1227 en 2010, 1800 en 2011 y producirá casi 8,000 Exabytes en 2015. Esto es, casi 8 Zettabytes u ocho mil billones de Gigabytes o 8×10^{21} bytes en medio trillón de archivos, -500×10^{15} - archivos.¹² Y seguramente durante los años subsecuentes este número será rebasado holgadamente.

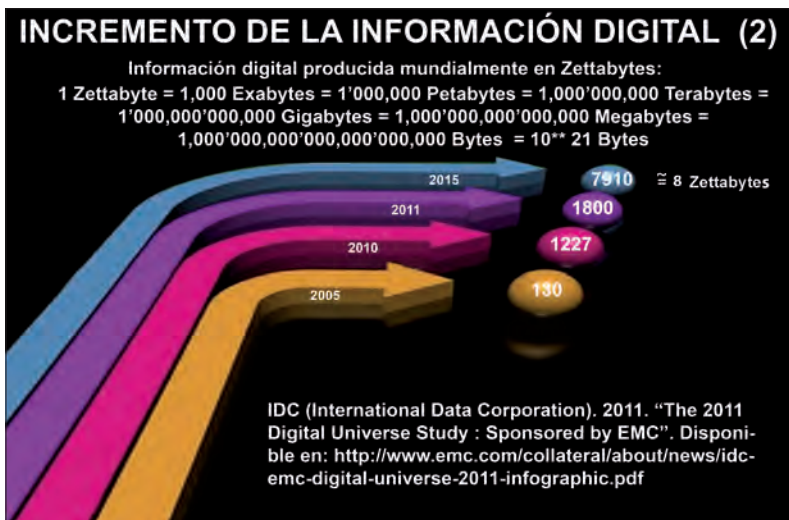
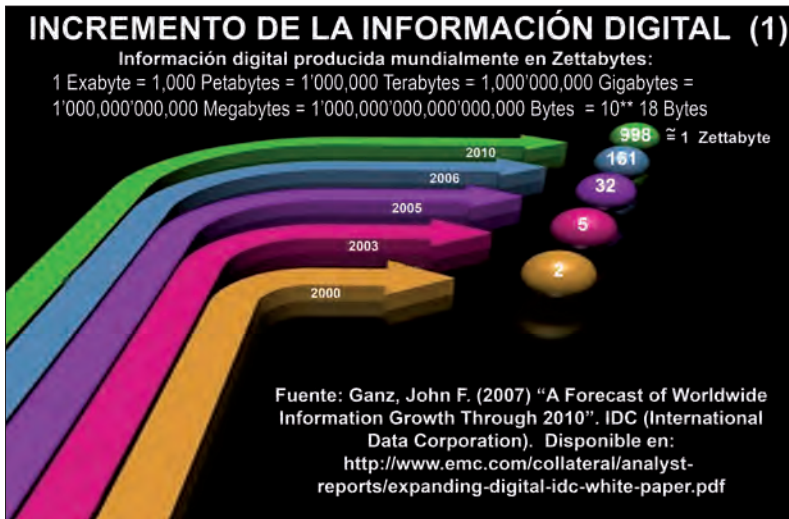
8 1 Exabyte = 1000 Petabytes = 1'000,000 Terabytes = 1'000,000,000 Gigabytes = 1'000,000,000,000 Megabytes = 10^{18} bytes o caracteres.

9 Peter Lyman y Hal Varian, "How Much Information?" [en línea].

10 John F. Ganz, *The Expanding Digital Universe. A Forecast of Worldwide Information Growth Through 2010* [en línea].

11 1 Zettabyte = 1000 Exabytes = 1'000,000 Petabytes = 1'000,000,000 Terabytes = 1'000,000,000,000 Gigabytes = 1'000,000,000,000,000 Megabytes = 10^{21} bytes o caracteres.

12 International Data Corporation, *op. cit.*

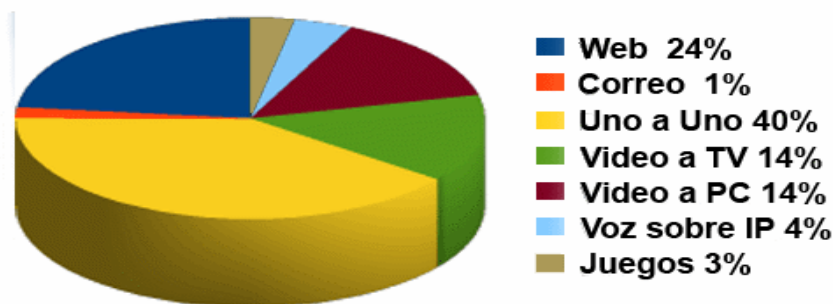


Para aquellos que son aficionados a las constantes físicas y químicas, al hablar de un Zettabyte nos aproximamos al número de Avogadro –el número de átomos de carbono contenidos en 12 gramos de carbono $12 = 6.022\,141\,79(30) \times 10^{23}$, considerado por casi dos siglos como una frontera en la naturaleza. Para tratar de ima-

Cómo preservar mi patrimonio digital personal

ginar este número consideremos que una pila de hojas de papel cuyo número fuese el número de Avogadro llegaría desde la Tierra hasta la estrella Polaris Borealis –la estrella polar–, a 430 millones de años-luz de nosotros. Igualmente, si ese número de granos de maíz fuese repartido entre todos los habitantes de la Tierra, cada ser humano sería poseedor de poco menos de 43 millones de metros cúbicos de maíz o poco más de 200,000 furgones de ferrocarril llenos de maíz hasta el tope.

La mayoría de las personas piensa que toda la información digital se encuentra en la web; pero este no es el caso. No toda la información digital que se produce en la actualidad va a dar a Internet, pero sin duda, la cantidad de información en esta red nos puede dar una idea del volumen producido. Pueden verse en la gráfica las proporciones de Internet de acuerdo con los datos suministrados por la empresa de telecomunicaciones Cisco en su estudio *Cisco Networking Index, Forecast & Methodology, 2006-2011*.



Proporción del uso de la Internet 2011

Fuente: Cisco Networking Index, Forecast & Methodology, 2006-2011

Nótese que en realidad la web representa sólo una cuarta parte de lo que se transmite por Internet. Y con respecto a la web: ¿Qué tan grande es en números absolutos? De acuerdo con la información del sitio Google, en 1998 había 26 millones de páginas; para el año 2000 se llegaba ya a los mil millones de páginas, y hoy en día ellos

mismos declaran imposible saber el número total, pero aseguran que es de varios cientos de miles de millones de páginas, tal vez un trillón.¹³ De acuerdo con el sitio The Raw Feed, la red mundial alcanzaba en febrero del 2008 155'583,825 sitios web.¹⁴ Subrayo: sitios, no páginas; el total de páginas es mucho mayor. Otros autores nos dan sus cifras: John Markoff cita un número proporcionado por Yahoo en donde afirma que en 2005 la web tenía 19,200 millones de páginas.¹⁵ Según Gulli y Signorini,¹⁶ en 2005 la web tenía 11,500 millones de páginas. Aunque los números difieran, sirven para darnos una idea más clara del tamaño y crecimiento de la web. Desde 1995 para acá, los dominios han crecido desde 15,000 hasta 350 millones.

Como conclusión de todo lo anterior es obvio entonces lo que el desarrollo tecnológico ha podido crear en las últimas décadas acerca de la generación y distribución de información. Por desgracia, desde la perspectiva de la preservación documental no se han producido avances tecnológicos ni de experiencia acumulada en el tema al mismo ritmo que la producción y el almacenamiento de información.

LA PROBLEMÁTICA DE LA PRESERVACIÓN DOCUMENTAL DIGITAL

Los requerimientos para la preservación a largo plazo llevan al límite a las tecnologías actuales de información y administración de información. Desde el punto de vista de las organizaciones preservar su información digital a largo plazo requiere de sistemas, modelos técnicos y de organización, así como personal calificado y experimentado, lo suficientemente robustos para sortear fallos tecnológicos, cambios sucesivos de plataformas de cómputo, obsolescencia de medios y formatos de almacenamiento, errores humanos, negligencia y ataques malintencionados, cambios a la misión

13 Jesse Alpert y Nissan Hajaj, "We knew the web was big..." [en línea].

14 "How Big is the web?" [en línea].

15 John Markoff, "How Big is the web?" [en línea].

16 Antonio Gulli y Alessio Signorini, *The Indexable web is more than 11.5 billion pages* [en línea].

Cómo preservar mi patrimonio digital personal

institucional de las organizaciones o fallas e interrupciones en su dirección y financiamiento, por mencionar algunas amenazas.

En la obra *Preservación del patrimonio documental digital en México*¹⁷ se establecieron las bases y definiciones de la preservación documental digital, así como sus premisas básicas. Se establecieron también los seis factores que inciden positiva o negativamente en el proceso de preservación documental digital de las instituciones. Se establecieron propuestas para ayudar en la toma de decisiones y en la elaboración de proyectos estratégicos en ese aspecto. Lo principal que cabe resaltar aquí es que ese trabajo fue hecho para las organizaciones que manejan información: bibliotecas, archivos, repositorios, etcétera.

Es pertinente continuar este esfuerzo pero ahora con un punto de vista totalmente personal. ¿Qué pasa con los materiales digitales que producen los individuos, fuera de las organizaciones? Si bien los principios y metodologías fundamentales son semejantes, es necesario enfocarse a los problemas y adaptar técnicas propias de ese tipo de materiales y su tratamiento especial, de forma tal que podamos lograr una visión personalizada, diseñada expresamente para los individuos, para la persona común, acerca de la preservación documental digital.

Haciendo de lado entonces a las bibliotecas y a los archivos pertenecientes a organizaciones así como a sus colecciones y herramientas debemos pensar en un enfoque del problema estrictamente para los individuos ¿Y por qué es esto importante? En el estudio ya mencionado del universo digital de la organización IDC,¹⁸ se establecen tres premisas muy importantes para nuestro propósito:

- 1) En la actualidad, 75% de la información digital producida en el mundo es generada por individuos.
- 2) Sólo un tercio de la información digital producida en el mundo tiene protección mínima.
- 3) Sólo la mitad de la información que debería estar protegida lo está.

17 Juan Voutssas M., *Preservación del patrimonio documental digital en México*.

18 International Data Corporation (IDC), *op. cit.*

Entonces, no sólo la información digital producida y resguardada en las organizaciones es cuantiosa e importante, también lo es la información digital personal. ¿Cuáles son los principales elementos a considerar en este aspecto?

- ❖ Cada día son más las personas que crean y guardan sus archivos documentales en formatos digitales: textos, tareas, trabajos y tesis escolares, documentos de oficina y trabajo: oficios, memorandos, textos, hojas de cálculo, presentaciones PowerPoint, contratos, etc. Copias electrónicas de documentos de trámites gubernamentales: declaraciones de impuestos, registros de firmas electrónicas, certificados, pólizas, etc. Copias de documentos escolares: inscripciones, actas, calificaciones, certificados, etc. Facturas, pagos y recibos electrónicos. Cuentas bancarias y de inversiones, préstamos y pagos, pólizas. Catálogos y manuales de productos y servicios. Compra de bienes y servicios en línea, firma de documentos electrónicamente, etc. Todos estos documentos son cada vez más cotidianos y tienden a incrementarse más y más, dejando huellas y registros digitalmente.
- ❖ Muchos profesionistas que trabajan en forma independiente –médicos, dentistas, arquitectos, abogados, contadores, etc.– cada vez con más frecuencia crean y llevan expedientes digitales de sus clientes, pacientes, casos y operaciones.
- ❖ También, cada vez son más las personas que consumen y descargan productos digitales: libros, revistas, artículos, videos y datos electrónicos; música digital, textos escolares, por citar sólo unos cuantos. De la misma forma, se descargan diariamente fotos, textos, comentarios, etc., de personas que interactúan en las redes sociales.
- ❖ Muchas personas publican en la web de forma regular textos, comentarios, fotografías, noticias, etc., ya sea a través de blogs, listas de discusión, chats, enciclopedias, “wikis”, etcétera.
- ❖ Prácticamente todas las personas usuarias de Internet reciben y envían correos electrónicos, tanto personales como de trabajo, de los cuales puede desearse preservar copia.
- ❖ Muchas personas desean guardar copia permanente de documentos importantes –textos, correos– más allá de las generaciones

Cómo preservar mi patrimonio digital personal

tecnológicas y obsolescencias de equipos, sistemas, programas y dispositivos, garantizando siempre su acceso futuro a largo plazo.

- ❖ Muchas personas tienen preocupaciones relacionadas con la seguridad de la información, ya que a veces debe ser operada, distribuida y consultada por muchos y en forma remota. Por sus condiciones de ser digital y operada en red, la información contenida en archivos de este tipo se enfrenta a riesgos de daño o pérdida. Las personas que se desempeñan en este ambiente en el que tienen información accesible por terceros vía la red, se enfrentan cada vez más con riesgos procedentes de una amplia variedad de fuentes: pérdida, alteración, copia o consultas no autorizadas de su información debido a negligencia, sabotaje o vandalismo. A muchos les preocupa la pérdida de información por desastres naturales: incendios, inundaciones, terremotos, etc. Espionaje y fraudes informáticos. El sabotaje o vandalismo se presenta en muy variadas formas: “spam”, virus informáticos, ataques de intrusión o de negación de servicio, robo de identidades, etc., son cada vez más comunes, ambiciosos y sofisticados
- ❖ A algunas personas les preocupan en especial las consideraciones legales con respecto a derechos de propiedad intelectual y de acceso a ciertos documentos. ¿Cómo publicar documentos en la red protegiendo mis derechos? ¿Cómo saber que no estoy utilizando en la red material ilegal o restringido?
- ❖ Y obviamente, los costos asociados a la preservación de este material digital.

Como parte de los inconvenientes a considerar dentro del uso de documentos digitales podemos distinguir los siguientes como los principales:

- ❖ Para poder consultarlos es indispensable el uso de una computadora o dispositivo electrónico similar.
- ❖ La computadora o dispositivo debe contar con un programa o aplicación adecuada para poder abrir cierto tipo de documento; estos programas funcionan sólo en cierto tipo de máquinas y de sistemas operativos. No son universales.

- ❖ Es muy fácil crear múltiples copias de los documentos lo cual obstaculiza encontrar la última versión de cierto documento, o la versión deseada que fue utilizada para cierta ocasión o evento.
- ❖ Con mucha facilidad los materiales pueden verse dañados o destruidos por múltiples factores –descuido, accidente, virus, etcétera–.
- ❖ Los archivos quedan con frecuencia inaccesibles al cambio de versiones de programas y sistemas.

LOS FACTORES DE LA PRESERVACIÓN DIGITAL

Si piensas que la tecnología puede solucionar tus problemas de seguridad, significa que todavía no entiendes ni esos problemas ni esa tecnología.

Bruce Schneier

Es muy común pensar que el problema de la preservación de documentos digitales es un problema exclusivamente del ámbito de la tecnología y por tanto, si dominamos este aspecto, habremos encontrado una solución. Esto es muy inexacto y tal creencia forma parte de las causas de la pérdida documental. En realidad, existen múltiples factores que inciden en la preservación a largo plazo de documentos digitales. Tales factores pueden agruparse de varias formas para su estudio. En lo personal, yo los he agrupado en seis categorías: factor cultural, tecnológico, legal, documental, económico y social. Estos seis factores y su visión relativa a las organizaciones han sido ya explicados en detalle en la obra *Preservación del patrimonio documental digital en México*. Si el lector quiere profundizar en ellos lo mejor será referirse a esa obra. Para fines de este libro, se utilizará y explicará la parte esencial de cada uno de los factores con el propósito de entenderlos de manera integral y de aprovecharlos en nuestro favor.

El factor cultural tiene que ver con la falta de sensibilidad de ciertas sociedades en ciertas épocas respecto al valor histórico de partes de su patrimonio documental, lo que las lleva a excluir la necesidad de preservar ese patrimonio, lo cual deriva a la larga en la pérdida de ese acervo documental. Esto puede hacerse extensi-

Cómo preservar mi patrimonio digital personal

vo a las personas. En nuestros tiempos, a pesar de que se afirma que vivimos en la “sociedad de la información” se tiene en forma general un enorme problema de tipo cultural con respecto a la preservación de la información electrónica, una especie de oxímoron cultural. Si no estamos conscientes de que hay un problema con la creación, manejo y preservación de nuestros archivos digitales, es poco probable que tomemos acciones tendientes a su manejo eficiente. De hecho, la frase de introducción de esta obra pretende comenzar a sensibilizar al lector acerca del problema existente. Una vez que estamos conscientes de que existe un problema con la creación, manejo y preservación del material digital de nuestra propiedad, podemos convencernos de la necesidad de comenzar a tomar medidas correctivas en este sentido y habremos comenzado a andar en la dirección correcta. Mientras pensemos que la preservación de los documentos digitales se da en automático al salvarlos o que su destrucción es debida a la “mala suerte” o al azar, seguiremos arrastrando el factor cultural en nuestra contra.

El factor tecnológico –muy conocido de todos– tiene que ver con el cambio rápido y constante de los dispositivos y aspectos tecnológicos relacionados con la información electrónica. Dentro del concepto de “obsolescencia tecnológica” se afirma hoy en día que el problema ya no es conservar los bits, sino cómo hacer que sean legibles en un futuro, dados esos cambios tecnológicos ajenos a la información documental en sí, pero indispensables para hacerla legible. Este es el factor relacionado con el entorno de esa información digital, tal como el equipo y los programas que se requieren para reproducirlos. Mención aparte requieren los sistemas operativos y múltiples formatos que se han creado para el uso y representación de diversos documentos digitales en todos sus tipos: texto, audio, imagen estática y en movimiento, etc., y que surgen y desaparecen con rapidez cada vez mayor. Dentro del factor tecnológico se encuentra además el elemento de seguridad de la información: ¿cómo se evitará que se pierdan por accidente, error o mala intención documentos que se desea preservar? ¿Cómo se garantiza su integridad para evitar que yo mismo o terceras personas sustraigan, destruyan o alteren esos documentos?

A continuación tenemos el factor de tipo legal alrededor de la preservación de documentos electrónicos: si deseo publicar documentos digitales en la red, ¿cómo puedo garantizar que esto se haga protegiendo mis derechos de autor y propiedad intelectual? ¿Cómo puedo encontrar material electrónico en la red que pueda ser utilizado en forma personal sin infringir disposiciones legales y salvando los derechos del autor? ¿Hasta dónde puede distribuirse información sin lesionar la privacidad de alguien? ¿Hay legislación vigente al efecto?

El factor documental –o metodológico– tiene que ver, en primer lugar, con la valoración de mis documentos electrónicos con el fin de tasar o establecer “oficialmente” el valor que le otorgo a los mismos con propósito de determinar el periodo y condiciones de su preservación. No todos los documentos generados digitalmente deben ser preservados y mucho menos todas sus copias y no todos por el mismo periodo. En segundo lugar, tiene que ver con los procedimientos para el registro de los documentos y los metadatos que deben ser agregados a estos documentos; de nada sirve preservar miles y miles de documentos digitales si luego no voy a poder recuperarlos. ¿Cuáles son los metadatos adecuados para cierto tipo de documentos? ¿Cómo se buscará en conjuntos no homogéneos de documentos con metadatos distintos? ¿Cuáles son los metadatos mínimos? ¿Cuáles los óptimos? ¿Cómo debe ser la interoperabilidad actual de datos, formatos y sistemas para que todo interopere en el futuro? Aunque el tema aparenta dificultad de inicio, como podrá verse más adelante, no lo es tanto.

En cuanto al factor económico, tiene que ver con que la persona que inicia un proyecto de esta naturaleza debe reflexionar tarde o temprano en materia de costos. ¿Cuánto cuesta administrar y preservar los documentos? ¿Pueden lograrse economías sin menoscabo de la calidad? Si bien los costos de preservar a nivel personal no son altos, existen, y deben establecerse minuciosamente los elementos que se verán involucrados en el desarrollo de la tarea para mantenerlos viables desde el punto de vista económico. De todas formas, debe tenerse siempre presente a la hora de costear que no preservar ni administrar saldrá a la larga todavía más caro. La

Cómo preservar mi patrimonio digital personal

pregunta de fondo es: ¿cuánto vale mi información? ¿Cuánto cuesta perder información valiosa?

Finalmente, el factor social tiene que ver mayormente con la visión de los gobiernos y las organizaciones y no tanto con los individuos, ya que está relacionado con el derecho al acceso a la información. Lo menciono aquí únicamente para tener el panorama completo de los factores. Su esencia consiste en cómo una sociedad va a poder garantizar el acceso y el empleo documental de la información que debe hacerse pública. Es necesario establecer los mecanismos para que en un futuro se pueda tener acceso efectivo y masivo a esa información por parte del público. De nada sirve preservar por preservar si nadie tendrá acceso a ella. Salvados los derechos, propiedades, privacidades y restricciones correspondientes, el reto futuro será cómo hacer llegar esa información digital a un número mayor de personas en el mundo sin crear nuevas brechas segregacionistas. Como puede observarse, éste no es el caso de nuestra información de propiedad personal.

Como ya se mencionó y en caso de tener la curiosidad intelectual, puede abundarse más en la lectura de los seis factores en la obra *Preservación del patrimonio documental digital en México*.

La naturaleza de los documentos digitales. Conceptos, formatos y dispositivos

Siempre que la ciencia resuelve un problema crea otros diez más.

George Bernard Shaw

CONCEPTOS BÁSICOS

Este libro trata acerca de la preservación de documentos digitales y por tanto es necesario comprender previamente –de manera general– la naturaleza de este tipo de documentos. Por lo mismo, antes de adentrarnos en el estudio y recomendaciones acerca de los materiales digitales, es conveniente plantear algunos conceptos y definiciones básicas que aseguren que estamos utilizando el mismo lenguaje.

Documento

En esta obra, el concepto de “documento” debe considerarse en su acepción más amplia. Documento es entonces un texto presentado como libro, revista, periódico, tesis, artículo, etc. También es documento cualquier imagen fija, sea fotografía, dibujo, grabado, mapa, plano, diagrama, partitura, placa de rayos X y más. Es también cualquier grabación de sonido: canción o pieza musical, voz grabada, etc. Lo es también una imagen en movimiento: película o video y de ahí en adelante. A estos documentos también se les llama “objetos”; si están en formato digital son entonces “documentos digitales” u “objetos digitales”. Los documentos digitales requieren de tres elementos para poder ser percibidos: la computadora o dis-

positivo electrónico, el o los programas necesarios para operarlo y leerlo y el documento digital en sí mismo.¹⁹

Archivo. Computación

Esta obra trata de “archivos” bajo esta acepción. Desde el punto de vista de las computadoras y otros dispositivos semejantes, se denomina archivo a una unidad lógica de información, generalmente indivisible. Se llama unidad lógica a la que contiene a toda la información necesaria para ser funcional por sí misma: un documento de texto, una fotografía, una pieza musical, un programa de cómputo; en la inmensa mayoría de los casos, un archivo en una computadora representa un documento de texto, una imagen, una canción, un video, etc. No obstante, en casos excepcionales puede desearse que un archivo agrupe a varios documentos a la vez, como los “empacados” o zip para que viajen siempre juntos o, si la información es muy grande, puede desearse que una unidad lógica como una película sea dividida en varios archivos separados. Todos los archivos poseen un nombre que los identifica, un “sufijo” o terminación que indica el tipo de archivo del que se trata y una ruta de acceso al archivo dentro del dispositivo.

Los archivos, salvo que se indique específicamente lo contrario, pueden ser creados, editados, movidos, reducidos, agrandados, copiados, transmitidos o borrados dentro de una computadora o dispositivo digital. Cuando un archivo se crea, el autor lo designa con un nombre arbitrario, y a menos que sea renombrado o borrado, conservará ese nombre permanentemente. El sufijo es una terminación que tienen los archivos que indica el tipo de archivo o formato con el cual está construido: así, un archivo con sufijo txt indica que se trata de un archivo de texto en formato de “texto plano”; un sufijo doc indica un texto pero en formato *document* de Microsoft; un sufijo jpg indica un archivo cuyo contenido es una imagen en ese formato; un sufijo mov indica que el archivo contiene un video o película en ese formato, y así sucesivamente.

19 Alejandro Delgado Gómez, *El Documento Electrónico en la Sociedad de la Información*.

Finalmente, la “ruta” es una secuencia en forma de árbol de las divisiones lógicas del o los dispositivos de almacenamiento en directorios, subdirectorios, sub-sub directorios, etc., en donde los archivos son almacenados. Así, un archivo nombrado “C:/mis_documentos/documentos_de_texto/mi_texto.txt” nos indica que es un archivo que puede localizarse dentro de la computadora en el disco duro denominado “C”, directorio “mis_documentos”, subdirectorio “documentos_de_texto”, documento denominado “mi_texto.txt” y cuyo formato es precisamente txt o texto plano.

Archivo. Ciencia archivística

De *file* (conjunto documental). La suma o conjunto integrado de documentos producidos o separados que participan en el mismo asunto o están relacionados con un mismo evento, persona, lugar, proyecto o materia, agregado de tal forma que pueda ser recuperado para una acción o como referencia. Son elaborados y recibidos por una persona física o jurídica en el desarrollo de actividades y preservados. También se le conoce como expediente o dossier. Este libro no trata de archivos en especial bajo esta acepción, ya que esta acepción es por lo general utilizada en organizaciones que manejan esos archivos, y dado el interés de nuestra obra, nos interesan los conceptos para las personas. No obstante, muchas de las premisas aquí enunciadas proceden de ese ambiente.

Dispositivos y documentos analógicos

Son aquellos que se basan en una semejanza o analogía para poder comprender y medir cierta realidad de la naturaleza; un ejemplo típico de ello es el reloj de manecillas: un dispositivo que nos permite observar y medir el paso del tiempo. El transcurrir del tiempo es un concepto que por sí mismo no puede observarse físicamente en forma directa con nuestros sentidos, por tanto es necesario medirlo de alguna forma alternativa y para ello se decidió desde hace mucho tiempo hacerlo por medio de una analogía o semejanza. Tal analogía consiste en la posición de las manecillas

Cómo preservar mi patrimonio digital personal

en la carátula con respecto al paso del tiempo; se estableció –arbitrariamente– que un giro de la manecilla pequeña corresponde a 12 horas –medio día– y que un giro de la manecilla grande corresponde a una hora, un veinticuatroavo de día. Observando la posición de las manecillas –algo que sí podemos hacer con nuestros sentidos– por analogía observamos y medimos el transcurrir del tiempo. Los números del 1 al 12 forman parte de una escala para medirlo –horas o minutos–. Cuando se conoce bien la analogía de las posiciones de las manecillas podemos inclusive prescindir de la escala: hay relojes con puntos en lugar de los números e inclusive algunos sin elementos excepto las manecillas; podemos de todos modos medir y observar el paso del tiempo en ellos ya que conocemos bien la analogía.

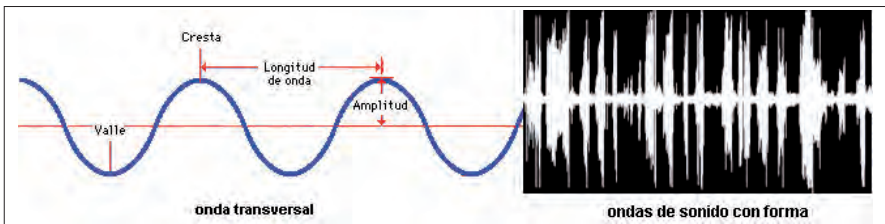
Otro ejemplo es el termómetro casero de mercurio. Sabemos que es un dispositivo para leer la temperatura, siendo ésta a su vez una medida del calor. Pero el calor no puede verse directamente y el sentirlo es subjetivo. ¿Qué es lo que realmente vemos en un termómetro? Calor no, temperatura tampoco; lo que vemos en realidad en ese termómetro analógico es el desplazamiento de la columna de mercurio dentro de un pequeño tubo que gracias a una escala podemos establecer y numerar. Por analogía del desplazamiento del mercurio gracias a su dilatación, podemos “ver” la temperatura. El desplazamiento hacia arriba o hacia abajo del mercurio nos indica el aumento o disminución de la misma y una escala numérica nos indica la cantidad de grados Celsius o Fahrenheit de la lectura.

Como estos podemos mencionar muchos otros dispositivos analógicos que permiten medir o percibir ciertas entidades de la naturaleza: el velocímetro de un auto, que nos permite medir su velocidad por medio de una analogía basada en el desplazamiento de una pequeña aguja sobre un eje, además de una escala. El multímetro, que nos permite medir entidades eléctricas: voltaje, corriente, resistencia, etc., de la misma forma por medio del movimiento de una aguja y utilizando varias escalas en una misma carátula.

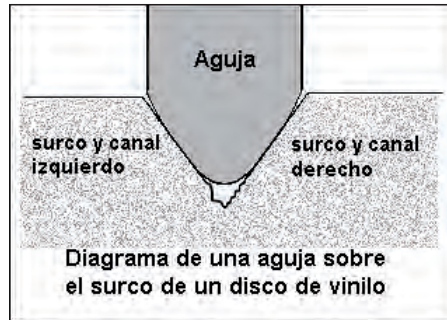
La fotografía con rollo de película es otro ejemplo de una analogía: por medio de luces, sombras y colores sobre una superficie po-

demostramos percibir un paisaje, un rostro, un objeto, un evento. Nótese que lo que vemos en esa superficie no es igual al original, no tiene volumen y es una mera representación gráfica de ese original; pero puesto e interpretado de cierta forma y bajo ciertas convenciones nos permite percibir e identificar muchas características del objeto real. Pensemos en un negativo fotográfico: en realidad es sólo una muy pequeña pieza de acetato con algunas manchas de colores en su superficie creadas con sustancias químicas y luz. Por lo general ni siquiera podemos percibir claramente en él lo que representa a simple vista. No obstante contiene en sí mismo –por simple analogía óptica y química– la capacidad de entregar múltiples copias, de muchos tamaños, de claras imágenes de personas, lugares, eventos y cualquier otro tema que sea de nuestro interés.

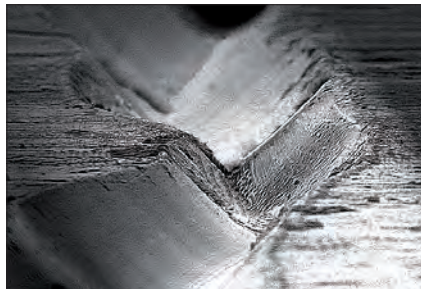
En el campo de la electrónica se desarrollaron durante sus primeras décadas dispositivos que nos han permitido percibir realidades de la naturaleza por medio de analogías que han sido construidas invariablemente en forma de ondas electrónicas. Pensemos en el sonido: un tipo de ondas mecánicas longitudinales producidas por variaciones de presión del medio en que se transmiten; estas variaciones de presión, captadas por el oído, producen en el cerebro la percepción del sonido. La forma de la onda en un instante dado nos da las cuatro propiedades básicas del sonido: altura, duración, timbre e intensidad. La altura –agudo, medio, grave– está dada por la frecuencia de la onda; la intensidad –fuerte o suave– está dada por la amplitud de la onda; el timbre –áspero, metálico, cuerda, percusión– está dado por los armónicos de la onda, esto es, su forma, y finalmente la duración –largo o corto– está dada por el tiempo de la vibración de la onda.



Cómo preservar mi patrimonio digital personal



Pensemos ahora en un disco antiguo de vinilo: su contenido es una representación de la forma de ciertas ondas de sonido grabada dentro de surcos sobre las superficies del disco. La forma física que tiene internamente el surco en cada momento varía simultáneamente con la forma del sonido y es aquí donde se encuentra la analogía: forma física del surco contra forma de la onda sonora. Una vez grabado, o más bien dicho moldeado, una aguja recoge esas variaciones en la forma física del surco y las entrega en forma de una onda electrónica la cual al ser amplificada y trasladada nuevamente a ondas sonoras por medio de bocinas nos permite percibir una reproducción muy parecida a los sonidos originales.



Surco en un disco de vinilo. Puede verse la forma variante y distinta a cada lado del surco, lo que representa respectivamente la onda de sonido en los canales derecho e izquierdo del estéreo. El ancho promedio de un surco es de 40 μm (micrómetros).²⁰

20 μm – un micrómetro = 1×10^{-6} metros = un millonésimo de metro = 1 milésimo de mm

Si analizamos un antiguo casete de audio o una cinta magnetofónica, también encontraremos una analogía con las ondas de sonido, sólo que en este caso la forma de las ondas sonoras está representada en los patrones o variaciones de flujos magnéticos grabados sobre una cinta de poliéster recubierta con un material ferromagnético –esto es, sensible a ser magnetizado, como el óxido férrico: Fe^2O^3 o el dióxido de cromo CrO_2 -. Cuando se graba la cinta, una cabeza o sensor magnético reorienta en cada punto de ella el conjunto de partículas de manera que su patrón magnético siga la forma que tiene la onda de sonido. Cuando se reproduce, la cabeza magnética recorre la cinta leyendo esas formas de variaciones magnéticas que al ser amplificadas y trasladadas nuevamente a ondas sonoras por medio de bocinas nos permite igualmente percibir una reproducción muy parecida a sus sonidos originales. La cabeza magnética se desliza sobre la cinta a una velocidad de 4.76 cm/seg, por lo que requiere de una cinta de aproximadamente 90 m para reproducir 30 minutos de sonido por cada dirección. Esta es la analogía magnética.

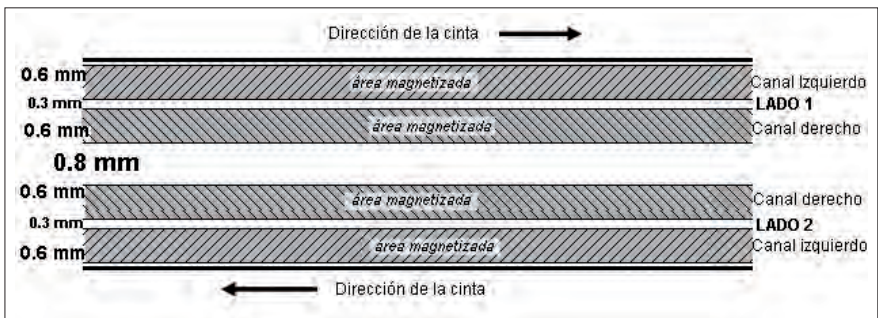
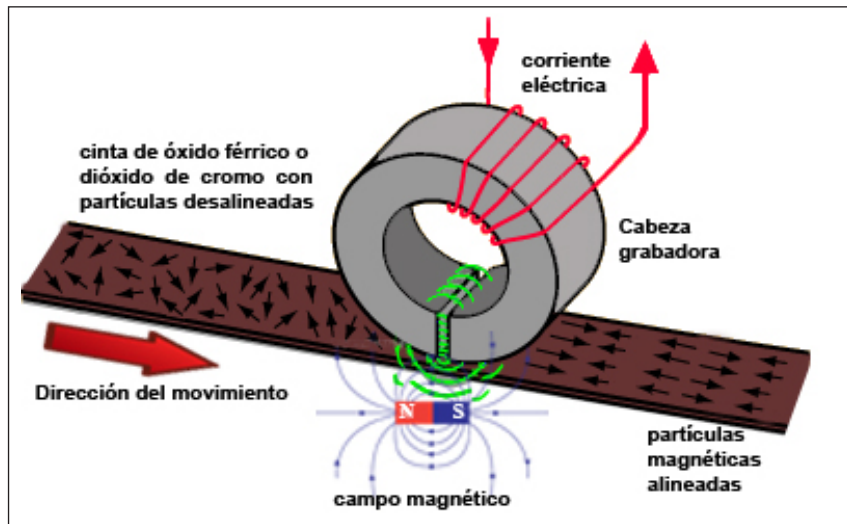


Diagrama de la superficie de un casete de audio.

Sucede lo mismo con un videocasete tradicional en formato Beta o VHS; son representaciones magnéticas de ondas sobre una cinta, sólo que en este caso, las ondas representan también imágenes además del sonido.



Como puede percibirse de los ejemplos anteriores, la variedad de dispositivos basados en analogías es inmensa. Como se ha podido observar también, no todas las analogías son electrónicas: la fotografía tradicional es un buen ejemplo de ello; también los muy antiguos gramófonos no eléctricos con sus distintivas bocinas en forma de tulipán. Los lectores de microfilme son dispositivos analógicos con componentes mecánicos, ópticos y químicos –la película– pero no electrónicos. Otra buena parte de estos dispositivos y sus analogías son en efecto parte mecánicos y parte electrónicos, como la tornamesa y el disco de vinilo; otros son en parte magnéticos y en parte electrónicos, como los casetes de audio o video, y así sucesivamente en una variedad casi infinita de combinaciones.

No obstante, en la literatura actual se tiende a basar las definiciones de datos, componentes, dispositivos, codificación, etc. –todos analógicos– exclusivamente con esta visión electrónica, como si el mundo de los dispositivos analógicos hubiese sido desde siempre electrónico. Por ejemplo, una definición actual de “codificación analógica” la establece como “el uso de patrones de señales electrónicas continuas variables para representar información”. Una definición típica de “dato analógico” es: “la unidad más pequeña de

información con significado, expresada en forma de señales electrónicas continuas y fijada en un medio analógico”. Prácticamente todas las definiciones de entes analógicos expresan claramente la idea de las señales “continuas y variables” –las señales en forma de onda ya mencionadas en los ejemplos– las cuales son por supuesto continuas y variables, pero esas definiciones no incluyen por lo general a los componentes mecánicos, ópticos, magnéticos o químicos, los cuales también forman parte de los entes analógicos. Volviendo a nuestro ejemplo del reloj mecánico, recordemos que en efecto es un dispositivo analógico, y que cuenta ya con algo así como ochocientos años en su forma moderna, muchos siglos antes de la electrónica. Esta ciencia vino a “redondear” y perfeccionar las representaciones analógicas de realidades de la naturaleza, pero debe quedar claro que existían dispositivos analógicos desde muchos siglos antes de la invención de la electrónica.

Por lo mismo, se habla de documentos analógicos como aquellos creados sobre soportes analógicos usando señales continuas y variables. Esto sólo es verdad si aplicamos esta definición con documentos tales como grabaciones de audio o video en discos de vinilo o cintas magnéticas, pero ya no lo es con documentos impresos sobre papel, negativos fotográficos, discos fonográficos, etc. No obstante, todos los documentos actuales que no son digitales tienden a ser clasificados muchas veces como analógicos por oposición, pero ello no es del todo exacto como ha podido verse.

Dispositivos y documentos digitales

“Digital” significa que una entidad de la naturaleza, en este caso un documento, está representado por dígitos; esto es, por números. Si bien en teoría podría utilizarse para ello cualquier sistema numérico, en la práctica se utiliza casi siempre el sistema numérico binario, ya que es el ideal para ser utilizado en equipos electrónicos u ópticos. El sistema numérico binario se basa en sólo dos números, 0 y 1, los cuales pueden ser representados en forma electrónica –existe corriente eléctrica = 1; no existe = 0–; en forma electromagnética –existe magnetismo = 1; no existe = 0– o en forma óptica –existe

Cómo preservar mi patrimonio digital personal

luz = 1; no existe = 0–, lo anterior entre muchas otras formas de señalización física, haciendo más simple la construcción de estos dispositivos.

De esta forma, un “documento digital” es aquel documento que ha sido creado de origen o convertido a una forma de representación basada exclusivamente en números bajo un cierto patrón arbitrario, con objeto de poder ser almacenado, transmitido o percibido por medio de dispositivos electrónicos. “Digitalizar”, por tanto, es el proceso de convertir un documento analógico en uno digital, con objeto de poderlo almacenar, transmitir o percibir en una computadora, equipo de teleproceso, etc. En este caso este proceso significa convertirlo a un formato basado en números, generalmente binarios, bajo un cierto patrón arbitrario preestablecido antes de ponerlo en su nuevo soporte, un dispositivo asociado a una computadora. A los documentos que nacen digitales se les llama “documentos digitales de generación” y a los documentos que estaban en un soporte tradicional y han sido digitalizados se les llama “documentos digitales de transformación”. Como corolario puede agregarse que –hablando estrictamente– todo documento digital es electrónico, ya que se requiere de dispositivos electrónicos para su uso, pero no todo documento electrónico es digital, ya que como vimos previamente, existen documentos electrónicos analógicos. No obstante, esto debe manejarse con cuidado, ya que en la actualidad, dado que prácticamente todos los documentos electrónicos que se crean o convierten hoy en día son digitales, los términos “documento digital” y “documento electrónico” tienden a volverse sinónimos, sin serlo del todo.

Como ya mencionamos, un registro analógico está formado por una serie de datos o valores de forma continua como los de una onda. En contraste, un registro digital es la representación de objetos o datos físicos por medio de una sucesión de valores numéricos diferenciados –discretos o discontinuos–, por lo general valores numéricos binarios –bits– de acuerdo con un código o convención preestablecidos. La primera vez que se utilizó el término “digital” para describir un registro numérico no continuo fue en 1938 al hacer una descripción de un circuito de cálculo que operaba datos

en forma “discreta”, es decir, no continua: “El emisor [...] difiere de otros emisores en que tiene doce puntos conductores digitales”.²¹ En contraste con una representación analógica de un objeto o proceso físico, una representación digitalmente codificada nunca se parece físicamente al original.

Volviendo a nuestros ejemplos de algunos dispositivos analógicos enunciados previamente, analicemos ahora a sus equivalentes digitales para redondear la idea: en un reloj digital, las pulsaciones de un oscilador de cuarzo marcan el tiempo; no vemos ninguna analogía, sólo su resultado evidente expresado en números: únicamente vemos dígitos, números que nos indican la hora y para nosotros ya no hay analogía ulterior que hacer: sólo leer la hora. En el reloj analógico, las personas necesitan aprender la analogía; si no les es explicada previamente no entienden lo que ven en la carátula. Una vez aprendida la analogía, la escala sobra, igual da un reloj de manecillas que tenga puntos en vez de números o inclusive con una carátula en blanco. En la versión digital, las personas que puedan leer números pueden leer automáticamente la hora en un reloj digital, sin explicación previa de ninguna analogía.

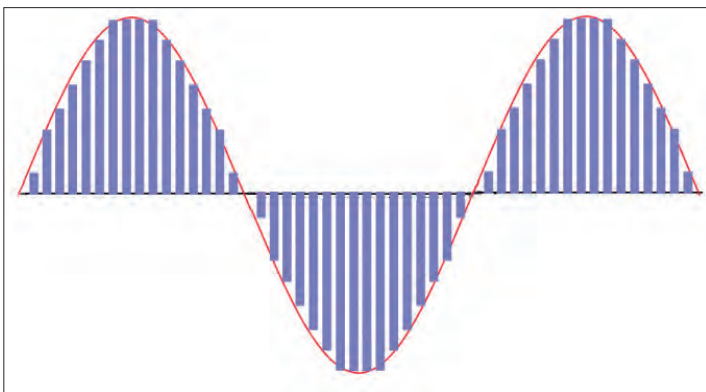
En un termómetro digital leemos sólo dígitos –números– que nos indican directamente la temperatura. No hay analogía que establecer o interpretar, el proceso se hace internamente en sus circuitos y no es visible para nosotros: sólo su resultado evidente y numérico. Los demás ejemplos son semejantes: el velocímetro digital, el multímetro digital, etc., son dispositivos en los que no hay analogía: sólo un resultado digital evidente.

En un disco musical, un CD, existe exactamente la misma información que existía en un disco de acetato o vinilo, un audiocasete, una cinta magnetofónica: como ya mencionamos, necesitamos grabar y reproducir una onda con las cuatro propiedades básicas del sonido: altura, duración, timbre e intensidad, sólo que necesitamos representar esa onda en una forma no continua, es decir, discreta o diferenciada. ¿Es esto posible?

21 Citado en University of Cornell, “Timeline of Digital Preservation” [en línea].

Cómo preservar mi patrimonio digital personal

Hemos dicho que la codificación digital implica valores discretos, discontinuos o diferenciados; esto significa que los valores vienen siempre como secuencias de números perfectamente distinguibles o diferenciables uno del otro. Aun cuando se deba representar una onda de forma continua, ello puede hacerse con secuencias diferenciadas de números. Si se analiza la figura siguiente, se observa que el valor de la amplitud de una onda puede expresarse dando valores numéricos distintos a una serie de valores numéricos o barras que representan la amplitud de la onda en un momento dado. A mayor cantidad de números o barras, la representación de la onda es más aproximada, más fina. Esto es real hasta un cierto punto.

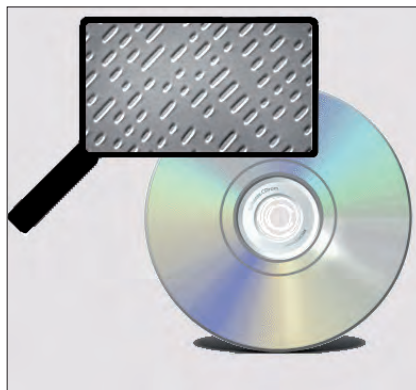


Representación de una onda en forma de valores discretos o diferenciados.

En la práctica, las barras se denominan “muestras”. El principio de Nyquist-Shannon estableció que para poder replicar con exactitud la forma de una onda –esto es, que el proceso sea matemáticamente reversible en su totalidad– se requiere que la frecuencia de muestreo sea igual al doble de la máxima frecuencia que la señal pueda portar. Todas las muestras más allá de ese número no aportan mayor calidad a la posterior reproducción de la onda. Siempre que la tasa empleada cumpla el criterio de Nyquist-Shannon el proceso de muestreo será –desde el punto de vista matemático– perfectamente reversible; es decir, su reconstrucción será exacta, no

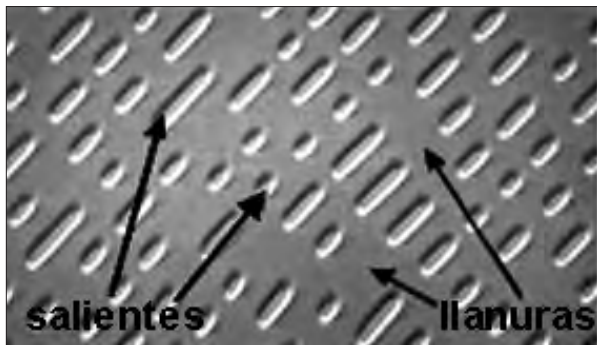
aproximada. Este teorema sirve para establecer de forma práctica el número de barras o muestras necesarias por ciclo para registrar una onda. Ejemplificando, si se muestrea una señal cuya máxima frecuencia es cinco mil ciclos por segundo –5 kilohercios o kHz– su reconstrucción y reproducción será exacta a partir de 10,000 muestras o barras, o sea el doble de la máxima frecuencia posible. Toda la información de una señal contenida en un intervalo de tiempo entre dos instantes cualesquiera estará descrita por la serie total de muestras siempre que la señal registrada sea de naturaleza periódica –como el sonido–. Si muestreáramos más allá de ese límite, sólo estaríamos consumiendo más espacio, pero la señal no mejoraría ya en calidad.

El ser humano puede escuchar frecuencias hasta 20,000 ciclos por segundo –20 kHz–; por lo mismo los fabricantes de discos musicales establecieron desde el principio que la frecuencia de muestreo para grabaciones digitales sería de 44,100 o 44.1 kHz, poco más del doble de lo que el ser humano puede oír. Ello daba la máxima calidad a las grabaciones digitales sin exceder el principio de Nyquist-Shannon. Más muestreo no hubiese agregado calidad a la grabación para el oído pero hubiera incrementado inútilmente el espacio de almacenamiento de cada grabación. Por ello se estableció que el muestreo de alta calidad en audio musical sería 44.1 kHz y así se ha mantenido desde el inicio del CD musical hasta nuestros días.



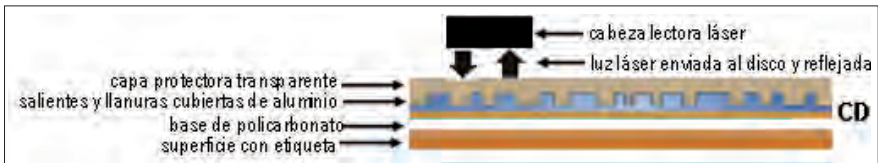
Cómo preservar mi patrimonio digital personal

En cuanto a la manera de representar la información sobre un disco, también ha habido un cambio; como mencionamos, en los discos de acetato o vinilo la forma del surco era igual a la forma del sonido. En un CD digital, la música no viene representada en forma de onda, sino en forma de una secuencia de números: las muestras o barras ¿Cómo se representa físicamente esa información?



La respuesta es simple. El disco compacto consiste en una base plástica de policarbonato en la cual se graban –por moldeo o reacción química– una serie de protuberancias, llamadas “salientes”, y se dejan zonas planas, llamadas “llanuras”, en una secuencia con forma de espiral. Las salientes y llanuras están cubiertas con una delgadísima capa de aluminio para hacerlas reflejantes y a su vez todo está cubierto con una capa plástica protectora transparente para evitar rayones. El láser, al desplazarse sobre la espiral, va iluminando cada punto del CD y con el reflejo del rayo va comprobando si hay una saliente o una llanura. El láser “sabe” siempre en qué zona se encuentra y detecta cuando no hay cambios en el reflejo, es decir, hay una secuencia de llanura o de saliente, y detecta también cuando hay un cambio de saliente a llanura o de llanura a saliente. La continuidad significa “unos” y el cambio significa “ceros”, y su secuencia va entregando el código numérico binario con la información, esto es, la información digital. Cada hendidura tiene 6 micrones –millonésimas de milímetro– por lo que no pueden ser observadas a simple vista. Un lado de un CD estándar de una hora

tiene aproximadamente 4,800 millones de posiciones o bits. La otra cara tiene por lo general el estampado gráfico del CD.



Corte transversal de un CD-ROM típico grabado en fábrica.

Como puede observarse, en estas definiciones se hace énfasis en los valores “diferenciados” –también llamados discretos o discontinuos– para marcar la diferencia con la definición de analógico, donde los valores eran continuos por provenir de representaciones tipo onda.

Si ya se sabía cómo almacenar y transmitir documentos por dispositivos y medios analógicos, ¿por qué la tendencia ha sido en los últimos años representar cada vez más los documentos con codificaciones digitales, las cuales han sustituido prácticamente a todas las representaciones analógicas? ¿Por qué se decidió el cambio hacia dispositivos y medios digitales? La respuesta se encuentra en la serie de contundentes y notorias ventajas que tienen los documentos digitales sobre sus equivalentes analógicos:

- ❖ La señal digital es más resistente a las señales erróneas ajenas, también llamadas “ruido”; por lo mismo, la señal digital es menos sensible a interferencias en su camino que la señal analógica.
- ❖ La señal digital permite la generación de múltiples copias de manera infinita sin pérdida de calidad. Si así se desea, una copia digital puede ser siempre idéntica al original, y por ende la copia de la copia, y así sucesivamente. Esto no sucede con las copias analógicas, las cuales siempre se ven degradadas con respecto a su original hasta llegar a ser casi irreconocibles con respecto a él. Con fines de preservación este es un factor crítico.
- ❖ Ante la pérdida de cierta cantidad de información, la señal digital puede ser reconstruida gracias a sistemas especiales de regeneración. También es posible agregarle con facilidad sistemas de detección y corrección de errores.

Cómo preservar mi patrimonio digital personal

- ❖ La señal digital es susceptible de comprimirse con el análisis de sus patrones de números. Esto hace que se almacene en menores espacios que su equivalente analógico. Puede ponerse una señal más rica y precisa que el equivalente analógico en el mismo espacio o tiempo. Como ejemplo compárese el espacio de las dos superficies de un disco de vinilo de doce pulgadas –en total 45 minutos sobre ambas– y el de un CD –74 minutos en una sola cara–. No obstante que ocupa menor espacio, en el CD puede almacenarse una mayor cantidad de información acerca del sonido grabado, lo que tiene como resultado una mayor calidad del mismo al reproducirlo. No es cierto que los discos de vinilo tenían más calidad que los CD.
- ❖ Los dispositivos digitales pueden construirse con mayor precisión que los analógicos; por ejemplo, un termómetro digital aproxima fácilmente hasta décimas de grado, mientras que a su equivalente analógico le es muy difícil. Un multímetro digital casero aproxima a lecturas de un décimo de voltio mientras que su equivalente analógico ofrece lecturas con dos o tres unidades de error. Y así para otros casos.
- ❖ La señal digital es muy fácil de procesar con ayuda de computadoras; puede por tanto ser cifrada o encriptada, amplificada, comprimida, etc., de manera muy sencilla. La señal analógica requiere de varios pasos previos y se vuelve sumamente estorbosa para ser almacenada en una computadora.

Las representaciones con números de un documento pueden existir en muchas formas distintas; esto es, existe más de una manera de representar un cierto documento en forma de números. En realidad, existen muchísimas maneras de representar un documento en forma numérica. Depende de una convención arbitraria o reglas de representación que se seleccionen en un momento dado. Por lo mismo, numerosas organizaciones, empresas, oficinas de estándares y corporaciones han ido definiendo diferentes representaciones a lo largo de las décadas para diferentes tipos de documentos digitales, y dada la variedad de posibilidades y de actores, existen diferentes representaciones para un mismo tipo de documento. Pensemos en

textos: estos son un ente que no se presenta en forma de ondas, pero sabemos que los textos pueden ser guardados, transmitidos y reproducidos por una computadora. Si hablamos de textos digitales entonces por supuesto su representación debe hacerse por medio de dígitos –números–.

La manera más sencilla de representar un texto en forma numérica es un “código” donde cada letra o carácter de la escritura (coma, punto y coma, punto, signos de admiración o interrogación, paréntesis, etc.) estén representados por un número diferente. Sabemos que por la construcción de la máquina debe ser un número binario: 0 o 1. Por supuesto que no podemos representar algo con sólo dos dígitos, pero sí con combinaciones de esos dos dígitos. Así, la combinación 00 puede representar una letra, la combinación 01 otra, la combinación 10 otra más y finalmente 11 puede representar otra. Como puede observarse, el número de combinaciones distintas creció a cuatro si usamos conjuntos de dos dígitos binarios. Si usáramos conjuntos de tres dígitos tendríamos ocho posibles diferentes combinaciones: 000, 001, 010, 011, 100, 101, 110, 111. En realidad, el número de combinaciones distintas en un cierto conjunto de dígitos viene dado por el número 2 –la base del sistema binario– elevado a la potencia correspondiente al número de dígitos. Así, 2^2 nos da cuatro posibilidades distintas con conjuntos de dos dígitos, los conjuntos de tres dígitos ofrecen 8, esto es 2^3 u ocho posibilidades distintas; si tomáramos conjuntos de cuatro dígitos el número de combinaciones diferentes estaría dado por 2^4 , esto es 16 distintas combinaciones, 2^5 daría 32, 2^6 daría 64 y así sucesivamente –véase el Anexo 3 para una explicación del sistema binario–.

Basado en ello: ¿cuántas combinaciones distintas se necesitan para representar todas las letras del alfabeto y los signos de puntuación? En el alfabeto anglosajón, 26 letras y algo así como 20 caracteres. Esto nos da 46 combinaciones distintas requeridas más diez dígitos decimales requeridos: 0 al 9; en total 56. Ya vimos que 2 a la sexta potencia nos da 64 combinaciones posibles, por lo que suena lógico que ése fuera el conjunto seleccionado, ya que 2 a la quinta potencia sólo nos da 32 posibles combinaciones y no alcanza. En un principio, hace varias décadas, así fue: se hizo una

tabla arbitraria donde se establecía que cada combinación de unos y ceros agrupados en conjuntos de seis correspondía a una letra, número o carácter que podían ser utilizados en las computadoras. Esto representó un problema a la larga: sólo se podían representar las letras mayúsculas. Si deseamos representar en la computadora letras mayúsculas y minúsculas forzosamente cada una de ellas debe tener una representación numérica diferente. Al aumentar el conjunto requerido de 56 a 81 combinaciones puede observarse que la convención anterior de seis posiciones no funcionaba más: era necesario agrandarlos en conjuntos de siete unos y ceros, para así tener $2^7 = 128$ posibles combinaciones distintas. De esta forma, fue posible representar las dos variantes de las letras además de los caracteres de la escritura. En un principio, cada fabricante de computadoras tenía su propia tabla de convenciones de unos y ceros, por lo que la tabla era válida sólo entre documentos intercambiados en equipos de la misma marca, ya que para cada fabricante una cierta combinación que representaba una letra no era igual a la combinación establecida por el otro. Por supuesto que con el tiempo y para poner orden se decidió establecer una tabla convencional única seleccionada para este propósito. A esta tabla se le llamó la tabla ASCII –*American Standard Code for Information Interchange* o Código Estándar Americano para Intercambio de Información.

Con los años, y dada la proliferación de computadoras en todo el mundo y de la necesidad de crear textos en otros alfabetos, hubo necesidad de realizar un nuevo cambio, una tabla ASCII con combinaciones de ocho unos y ceros, para así poder representar $2^8 = 256$ combinaciones distintas. Así, pudieron agregarse a la tabla ASCII todos los caracteres provenientes de los demás alfabetos latinos: la ñ, la ç, las vocales acentuadas á, à, ä, â, ã, å, é, è, ë, ê, í, ì, î, î, ó, ò, ô, õ, ø, ú, ù, ü, û y todas sus mayúsculas, otras letras tales como la ß o doble s alemana, æ, ý, ÿ, además de caracteres especiales de estos alfabetos, como “¿”, “¡”, o “€” llegando de esta forma a la versión de ocho dígitos que es la utilizada hasta hoy. Nótese que la tabla ASCII sólo representa los caracteres de alfabetos latinos. Para los caracteres griegos, cirílicos, árabes, japoneses, coreanos, etc., se utilizan

tablas semejantes, pero donde cada una de las 256 combinaciones de unos y ceros representa otro carácter distinto al de nuestro alfabeto. Así, existe una tabla con una convención arbitraria para cada grupo de alfabetos, sólo que esas no se denominan tablas ASCII. Cuando esta tabla pasó de conjuntos de siete dígitos a conjuntos de ocho se le denominó además *ISO-Latin-8859* convirtiéndose en norma internacional vigente a la fecha.

Gracias a esta nomenclatura ASCII o ISO-Latin es que hoy en día se define un conjunto preestablecido de ocho bits –*binary digits*– o dígitos binarios como un byte, y es equivalente a un carácter. La tabla completa ASCII puede verse en el Anexo 4.

FORMATOS

Como hemos ya establecido, todo documento digital sin importar su tipo: texto, imagen, sonido, video, etc., es una representación hecha con dígitos, o sea números –de ahí el nombre de digital– de la realidad que ese documento representa. Cuando una persona u organización preestablece las reglas por las cuales un cierto tipo de documento puede representarse con dígitos, está estableciendo un formato digital. Debe tenerse en cuenta de que existen muchas, innumerables maneras en las que cierto documento u “objeto” puede ser representado con números; por lo tanto, pueden existir muchos, innumerables formatos para representar al mismo tipo de documento y existen formatos para representar a variados tipos de documentos. Cada formato ha sido creado con un propósito y por lo mismo cada uno de ellos tiene ciertas ventajas y desventajas. No existe el formato universal que maximice todas las ventajas minimizando todas las desventajas. El conocimiento de unas y otras y de sus contextos para aprovechar las primeras y evitar las otras es lo que nos permitirá ir tomando mejores decisiones acerca de los formatos que debemos usar para el trabajo cotidiano y aquellos que vamos a usar para preservación a largo plazo de nuestros materiales documentales. Esta es la importancia del estudio y conocimiento de los formatos y no un mero ejercicio intelectual.

Cómo preservar mi patrimonio digital personal

Continuando con la idea, existen diferentes formatos para representar textos –txt, doc, odt, PDF, html, rtf, etc.–; diferentes formatos para representar imágenes –jpg, tiff, tga, gif, bmp, png, etc.–; otros más para representar sonido –CDD-Audio, mp3, midi, ra, wav, aiff, etc.–, y por supuesto otros tantos para representar cine o video –wma, mpeg, avi, etc.–. Cada uno de ellos es una representación digital distinta del documento que pretenden representar: los hay especializados en textos, en imágenes, en sonido, en video. Cada uno de ellos proviene de diferente compañía u organización y fue construido con distintos propósitos, ventajas y desventajas.

Todos los formatos digitales han sido diseñados para que los documentos creados con ellos puedan ser procesados, almacenados o transmitidos desde su dispositivo natural nativo –iPad, iPod, reproductor de sonido, cámara fotográfica o de video– hacia una computadora y viceversa. Como también ya mencionamos, si bien en teoría puede utilizarse cualquier sistema numérico para hacer esta representación numérica, en la práctica se utiliza casi siempre el sistema numérico binario y combinaciones del mismo, ya que es el ideal para ser utilizado en equipos electrónicos u ópticos; prácticamente todas las computadoras del planeta se basan en él, haciendo más simple la construcción de estos dispositivos.

Los formatos no son legibles por sí mismos dentro de una computadora o dispositivo electrónico por el simple hecho de estar en forma digital de unos y ceros. Para que una computadora o dispositivo pueda crear o leer un documento en un formato dado, es indispensable que exista ya instalado un programa o “aplicación” que sea capaz de interpretar el formato en cuestión, de otra forma le será imposible a la máquina interpretar ese documento.

Es conveniente aclarar en este punto que existen otros formatos de archivos que son electrónicos, pero no son digitales. A estos se les llama formatos analógicos y se usan en los dispositivos analógicos que ya hemos estudiado; por ejemplo los formatos Beta o VHS de los videocasetes. En las computadoras de casa u oficina y en los dispositivos móviles actuales, por su misma naturaleza, se emplean exclusivamente documentos en formatos digitales, representados con números, a los que nos referiremos de aquí en adelante.

Formatos de texto

Volviendo a los textos, analicemos los formatos más comunes que utilizamos en nuestros documentos así como sus características, ventajas y desventajas.

Hemos analizado ya el formato más simple para textos: el texto ASCII. Los documentos creados en esta forma, al momento de ser guardados en la computadora, generan un archivo en formato txt. Dado que es un formato muy simple, prácticamente todas las computadoras y dispositivos semejantes tienen instalado de fábrica una aplicación que puede crear y leer este tipo de documentos. Esta es una de sus ventajas: podrá ser leído en prácticamente cualquier dispositivo.

Un documento txt sólo puede contener texto; esto es, no pueden intercalársele imágenes ni ningún otro tipo de información, ni variantes de letra tales como negritas, itálicas, subrayado, etc.; tampoco cambios de fuente de letra: Courier, Arial, Helvética, etc. Si sólo quiere guardarse y transmitirse texto, este es un formato ideal, ya que los archivos resultantes de él al no tener “añadidos” prácticamente no ocupan nada de espacio y pueden almacenarse y transmitirse usando muy poco espacio y tiempo de transmisión. Además, pueden ser leídos por prácticamente cualquier computadora o dispositivo del mundo, ya que casi todos tienen alguna opción de fábrica para leer este tipo de documentos. Si lo que se desea es texto con características de edición y presentación muy elaboradas, o lleva imágenes y otros añadidos, este formato definitivamente no es recomendable.

El formato doc es un formato también para textos el cual ha sido creado por la empresa Microsoft. Permite la creación de textos y permite además su edición y diseño de una manera más elaborada: permite variantes de las letras, permite el uso de múltiples tipos de fuente de letra, permite cambiar el tamaño de la fuente de letra donde se desee dentro del texto, cambio de colores de letra, inserción de imágenes, tablas, corte y pegado, entre otras herramientas. Esta aplicación logra su cometido agregando “marcas” al texto ASCII. Históricamente, las marcas tienen su origen en el conjunto de ins-

trucciones que se le daban a un impresor o tipógrafo acerca de cómo debería imprimirse un cierto pasaje de un texto: las negritas, los subrayados, la tipografía, el tipo y tamaño de letra, los símbolos especiales se señalaban con anotaciones o marcas al margen del texto, de ahí el nombre. Conforme los textos se fueron automatizando, el término se fue extendiendo hasta cubrir toda clase de códigos de marcado insertados en esos textos electrónicos con objeto de incrustar una serie de características propias que el usuario desea en cada sección o parte del texto. De este concepto se desprende que el marcado es una codificación agregada al texto, gracias a la cual cada porción del mismo se va haciendo explícita y toma forma en un contexto. Los títulos, encabezados, los párrafos, los capítulos, los incisos, las sangrías, etc., se han agregado a los textos desde hace mucho tiempo para que sus lectores puedan leer explícitamente un conjunto de textos y darles un sentido. En el caso de un procesador de texto, obviamente, las marcas son ciertos conjuntos preestablecidos de dígitos binarios, las cuales no son visibles por el usuario.

En suma: un editor de textos puede trabajar prácticamente todas las funciones de esta tarea. Como todas las aplicaciones que manejan formatos, debe estar instalada previamente en el dispositivo para poder crear o leer este tipo de documentos. Es un producto comercial de la empresa que lo creó y por lo mismo sus especificaciones pueden cambiar –como ya lo han hecho– sin previo aviso. Aunque existen aplicaciones alternativas gratuitas que pueden crear y leer este tipo de formato, su aplicación principal, Word de Office, es un producto con costo.

El formato Open Document Format for Office Applications –Documento Abierto para Aplicaciones de Oficina Automática– u odt –Open Document Text– es un formato creado a nivel corporativo por OASIS, un consorcio de empresas de cómputo buscando precisamente la creación de un estándar que no estuviese atado a un producto o compañía. Fue aprobado para su publicación como estándar internacional ISO e IEC con el nombre de *ISO/IEC 26300:2006*.²²

22 *ISO and IEC approve OpenDocument OASIS standard for data interoperability of office applications.*

El formato OpenDocument es una propuesta de formato estándar para archivos de documentos ofimáticos tales como oficios, memorandos, hojas de cálculo, gráficas y presentaciones. Está basado en un esquema XML²³ inicialmente creado por la organización OpenOffice. El estándar fue desarrollado públicamente por un grupo de importantes organizaciones del cómputo a nivel mundial, entre ellas algunas tales como IBM, Sun, Novell, Intel y Adobe. Es accesible públicamente y puede ser implementado por cualquier persona o grupo sin restricciones. Este formato pretende ofrecer una alternativa abierta a los formatos de documentos propietarios de Microsoft tales como doc, xls, ppt, incluyendo el denominado Microsoft Office Open XML Format cuyos requisitos de licenciamiento impiden su empleo a diversos competidores. La motivación principal para usar formatos estándar estriba en que las personas y organizaciones que lo hacen evitan convertirse en clientes cautivos de un único proveedor de *software*, permitiéndoles cambiar de ambiente informático si su proveedor actual es expulsado del mercado o cambia su modelo de licenciamiento en términos menos favorables para el cliente.²⁴ OpenDocument es el único estándar para documentos ofimáticos implementado por distintos competidores, aceptado por organismos de estandarización independientes y susceptible de ser implementado por cualquier proveedor. Ya existen versiones de programas que usan el formato OpenDocument, siendo los más conocidos OpenWriter de la suite ofimática OpenOffice que lo utiliza como formato de archivos por omisión, actualmente se ofrece la versión 3.3.0; StarWriter de la suite ofimática StarOffice –versión 9.0– y Kword de la suite ofimática KOffice, actualmente se utiliza la versión 2.3.3. También puede usarlo Writer de Google.

Básicamente un documento tipo odt es semejante a un documento tipo doc: un archivo de computadora conteniendo texto con una serie de marcas que permiten su edición en lo relativo a su

23 XML: acrónimo de eXtensible Markup Language o Lenguaje de Marcado Extendido. Véase Microsoft, *XML para Principiantes* [en línea].

24 Organization for the Advancement of Structured Information Standards (OASIS), *OASIS Standards and Other Approved Work* [en línea].

Cómo preservar mi patrimonio digital personal

aspecto: párrafos, secciones, letra, tamaños, así como “inserciones”: imágenes u otros añadidos. Los editores de texto “de fuente abierta”, como StarOffice u OpenOffice pueden crear, leer y editar este tipo de archivos de la misma forma que lo hacen con documentos doc. Inclusive, pueden convertir los formatos entre ellos. Este tipo de editores de texto son de acceso libre o mínimo costo al no ser productos comerciales atados a una empresa.

El formato PDF o Portable Document Format surgió en 1993; es una variante de Postscript, lenguaje de programación diseñado para la definición de páginas para sistemas de impresión de alta definición y requerimientos. Por lo mismo permite representar textos de una manera muy dinámica, con variedad de tipos de letra, distribución física en la página, inserción de gráficos, etc.; además se pueden incluir gráficos vectoriales de tal forma que al ampliar la imagen no se pierda calidad y por lo mismo su acabado puede ser muy elaborado. La característica más importante del formato PDF es que conserva el aspecto original de un documento. Adicionalmente, permite realizar índices e hipervínculos y buscar palabras dentro del mismo documento. Es un formato idóneo para la impresión de documentos. Prácticamente todos los formatos de texto se pueden convertir directamente a formato PDF. Este formato se ha difundido enormemente a nivel mundial por parte de distribuidores de publicaciones ya que no permite la edición del texto ni su copiado en partes, lo cual lo hace un formato ideal para los distribuidores de información que desean que ésta se use sólo para lectura. La compañía Adobe, su propietaria, comercializa únicamente la parte que crea estos documentos, y distribuye gratuitamente la aplicación que permite su lectura en una gran diversidad de plataformas de cómputo. Como todo el mundo puede tener instalado en su computadora o dispositivo el lector de documentos para formato PDF de forma gratuita, se ha vuelto un formato muy popular para este tipo de propósitos.

Las principales ventajas de este formato consisten en:

- ❖ Conserva el aspecto original de un documento.
- ❖ Puede contener múltiples páginas en un solo archivo.
- ❖ Posibilidad de zoom.
- ❖ Opciones de desplazamiento: miniaturas, marcadores.

- ❖ Creación de hipervínculos.
- ❖ Contenido con multimedios.
- ❖ Creación de Formularios.
- ❖ Opciones de seguridad para evitar la edición.
- ❖ Texto extra reconocible y editable –incluye módulo de OCR–.
- ❖ Posibilidad de incorporar comentarios y notas.

Si bien PDF en general es un formato estándar, abierto, de gran aceptación en la actualidad para preservación de documentos tipo texto, existe en especial una versión de este formato que es recomendada particularmente para la preservación de documentos a largo plazo: la denominada PDF-1.4 estándar *ISO 19005-1:2005*²⁵ conocida popularmente como el PDF/A-1.

Existen otros ambientes institucionales que se dedican a la producción de textos profesionalmente, tales como editores de libros, revistas o periódicos; agencias de publicidad, etc. En estos ambientes los formatos más difundidos sin duda son QuarkXpress de la empresa Apple y el formato Pagemaker también de Adobe, pero a nivel de uso personal se ven muy poco por lo que no abundaré en ellos.

El formato HTML –Hyper-Text Markup Language– o Lenguaje de Marcado de Hipertextos fue creado en 1989 por Tim Berners-Lee y Anders Berglund, dos investigadores del Laboratorio Europeo de Partículas Físicas –CERN–; este es un formato basado en etiquetas para marcar documentos técnicos y hacerlos así compartibles en Internet. Desde 1992 ha llegado a ser el estándar para el manejo de información en la web. En este formato las marcas o “etiquetas” agregadas a los documentos sí son visibles por el usuario. Su objetivo es el de publicar documentos en la web haciéndolos legibles en forma semejante en todos los dispositivos conectados a la red sin importar la plataforma desde donde sean accedidos.

Estos no son todos los formatos existentes para representación de textos, pero son los principales para uso personal.

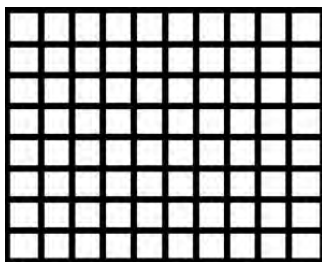
25 International Organization for Standardization, *ISO 19005-1:2005. Document management. Electronic document file format for long-term preservation. Part 1: Use of PDF 1.4 (PDF/A-1)* [en línea].

Cómo preservar mi patrimonio digital personal

Formatos de imagen

Los métodos para representar una imagen en forma digital se dividen en dos grandes grupos: las representaciones por medio de un mapa de bits y las representaciones por medio de vectores. Una imagen puede ser creada digitalmente de origen o puede ser digitalizada a partir de una imagen existente en un soporte tradicional y convertida a una representación digital.

Imágenes de mapa de bits



Básicamente, un mapa de bits es una representación digital de una imagen que puede entenderse con una rejilla matricial o *raster* la cual está compuesta de “celdas” o puntos de colores arreglados en renglones y columnas, donde cada punto es representado por un conjunto de datos numéricos que determinan la posición y el valor del color de ese punto. La secuencia total de todos los números de cada punto es la representación digital de una imagen. En el diagrama ejemplo puede verse una rejilla de diez por ocho pixeles.



Para tratar de entender mejor este concepto pensemos en un mural construido a partir de pequeños mosaicos; en él la imagen total está compuesta por muchas pequeñas piezas de mosaico de múltiples colores. Visto desde una distancia razonable no notamos los pequeños azulejos individuales: nuestro cerebro sólo ve la imagen general que representan. Conforme nos vamos acercando a la imagen empezamos a distinguir cada una de las pequeñas piezas que conforman el mural y podemos ver que cada una de ellas tiene un color propio. El concepto de “pixel” –PICture ELement– es similar; cada pixel es uno de los puntos de esa rejilla, excepto que en un escáner estos pixeles son mucho más pequeños que los mosaicos, y siempre están alineados en renglones y columnas perfectas de pequeños rectángulos o cuadrados para componer la imagen total. En combinación crean una impresión visual de líneas, sombras o matices específicos dentro de una imagen mono o policromática para ser reconstruida después en la pantalla de una computadora o sobre una hoja impresa. A esta rejilla también se le denomina imagen rasterizada, gráfico rasterizado o imagen matricial. Véase la imagen siguiente donde se representa un pez en pixeles o puntos en una matriz de 13 renglones x 16 columnas.

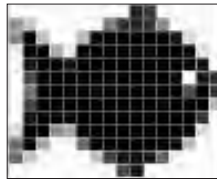


Imagen en pixeles o puntos en una matriz de 13 renglones x 16 columnas.

Pensemos en cualquier objeto documental que sea una imagen plana –fotografía impresa, grabado, mapa, diagrama, partitura, etc.– e imaginemos que podemos “romperlo” en una rejilla o imagen matricial o arreglo rectangular de puntos distribuidos a lo largo y ancho de un plano tales como el presentado previamente. Eso es exactamente lo que hace un escáner cuando “barre” una imagen puesta en él. Esto se logra por medio de un Dispositivo de Carga Acoplada o CCD, *Charge-Coupled Device*; un dispositivo sensible a la luz, en

forma de línea o barra formado por millones de celdillas que al recibir luz generan un impulso eléctrico. El escáner arroja luz sobre la imagen, la cual es reflejada y registrada por el CCD; éste “romperá” la imagen en miles y miles de pequeños puntos –píxeles– asignándole a cada uno de ellos un valor numérico para el color y otro valor numérico para la posición que el punto ocupa en el plano, creando así un renglón del mapa de bits; al desplazarse la barra del CCD del escáner a lo largo de su ventana gracias a su motor, el dispositivo va registrando renglón por renglón los valores llenando la rejilla. El valor numérico del color de cada punto se obtiene de una tabla preestablecida.

Si la imagen es sólo blanco y negro, por ejemplo una partitura, cada punto puede tener sólo dos valores numéricos posibles: 0 y 1, que serían los valores digitales asignados al color de un cierto punto. Si la imagen tiene tonos de gris, por ejemplo una fotografía en blanco y negro, cada punto puede tener un valor numérico que irá desde el blanco, pasando por una serie de valores de distintos grises, hasta el negro. En la práctica se pueden preseleccionar entre ocho posibles tonos de gris o dieciséis posibles tonos de gris; es decir, la paleta de colores posibles de tonos de gris puede seleccionarse de ocho o de dieciséis tonos; cuanto mayor sea esa paleta de grises, mejor fidelidad cromática. Por tanto, el valor numérico del color gris asignado por el escáner a un cierto punto de acuerdo a su paleta preestablecida puede variar entre cero y siete o entre cero y quince, dependiendo del rango deseado por el usuario. Sabemos que deben ser números binarios; por lo mismo para representar ocho posibles valores distintos necesitamos 3 bits ya que como vimos 2^3 nos da ocho posibles combinaciones numéricas diferentes; si fuera de dieciséis colores necesitaremos 4 bits, 2^4 nos da dieciséis posibles combinaciones distintas; con estos conjuntos de bits representará el escáner numéricamente el valor de un tono de gris de cada punto dentro de un pixel.

¿Qué pasa cuando la imagen es a colores? Aunque existen varias técnicas, explicaremos la idea con el concepto de color RGB. La teoría del color dice que cualquier color es una mezcla de tres valores de componentes rojo, verde y azul –*Red, Green, Blue*, de

donde viene lo de RGB-. Un pixel a color de imágenes digitales es entonces sólo una mezcla de valores RGB donde cada uno de esos valores puede ser representado con un número. Si representamos cada uno de esos valores con un número binario de 8 bits, tendremos $2^8 = 256$ posibles valores para cada componente rojo, verde o azul. Tres números de 8 bits para representar el color de cada pixel o tres bytes de ocho bits cada uno. A esto se le denomina profundidad de color, la cual en este caso es de 24 bits, la suma de los tres valores de 8. Por ejemplo, tres valores RGB tales como 175, 135, 234, significan rojo = 175, verde = 135, azul = 234 y juntos conforman un pixel de un cierto tono de violeta. El color negro es 0,0,0 y el blanco es 255, 255, 255.²⁶ Como puede verse también, el número de colores posibles diferentes a ser representados de esta forma es de $256 \times 256 \times 256 = 16'777,216$, o simplemente una paleta de 16.7 millones de colores.

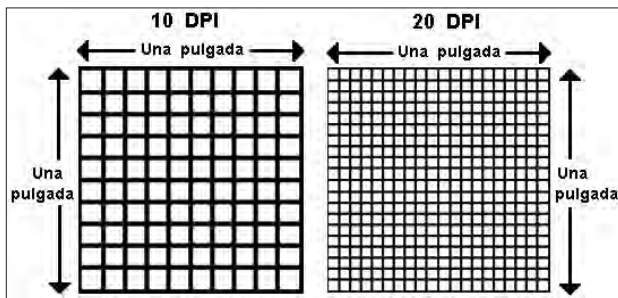
Conozcamos ahora otra característica de una imagen bajo el método de mapa de bits: su resolución, es decir, el número de puntos en que se puede “romper” la imagen al ser digitalizada, ya que este tamaño puede ser variable y se preselecciona al momento de escanear. Puntos grandes –baja resolución– hacen una imagen cuyos bordes y trazos pueden ser observados a simple vista, y por tanto burda y de poca calidad. Entre más finos sean los puntos –alta resolución– la imagen mejora su definición y por ende su calidad. La resolución se mide en puntos por pulgada, ppp, o por sus siglas en inglés, dpi, *dots per inch*, esto es, en cuántos puntos se romperá o descompondrá una cierta imagen por cada pulgada.

Debe reflexionarse en el hecho de que un pixel o punto no tiene una medida estándar fija. Su medida es un número variable que depende de lo que establecemos al momento de escanear y es relativo; una pulgada de imagen puede tener como resolución 75, 150, 300, 600, 750 o cualquier otra cantidad de pixeles o puntos, lo que nosotros definamos; por lo mismo el tamaño del punto o

26 Si tienes acceso a un editor de imágenes, como por ejemplo *Photoshop*, al tener una imagen abierta y tocar cualquier punto de la imagen con la herramienta Gotero, se abre un cuadro donde podemos ver, entre otras cosas, cuáles son los valores preestablecidos exactos RGB que componen ese punto específicamente.

Cómo preservar mi patrimonio digital personal

pixel es variable y relativo. Véase la siguiente figura: ambas rejillas representan una pulgada cuadrada de imagen, una a 10 dpi y otra a 20 dpi; como puede verse en la misma superficie hay $10 \times 10 = 100$ pixeles o $20 \times 20 = 400$ pixeles; el tamaño del pixel depende de la resolución que seleccionemos y es relativo a una pulgada cuadrada.



Píxeles en una pulgada cuadrada a 10 dpi y a 20 dpi. Como puede verse el tamaño del pixel es variable y depende de la resolución.

Así, al tener una imagen a digitalizar que mida, por ejemplo, tres pulgadas de ancho por dos pulgadas de alto, al seleccionar una resolución para digitalizar de 75 dpi significa que el escáner tomará 75 muestras de la imagen por cada pulgada, tanto a lo largo como a lo alto, y por tanto romperá esa imagen en 75 puntos o celdas por cada pulgada, tanto horizontal como verticalmente, y esto dará 75×3 de ancho = 225 puntos horizontalmente, y 75×2 de alto = 150 puntos verticalmente; dicho de otra forma, cada pulgada cuadrada de la imagen se rompe o divide en $75 \times 75 = 5,625$ puntos o muestras tomadas por el escáner. Esto es, la imagen tendrá un total de 225 píxeles de ancho x 150 píxeles de alto = 33,750 puntos o píxeles para representar la imagen completa. Siguiendo con el ejemplo, si esa misma imagen fuese digitalizada o construida a una resolución de 300 dpi, ello indicaría que esa imagen posee $300 \times 3 = 900$ puntos horizontalmente y $300 \times 2 = 600$ puntos verticalmente, esto es, 90,000 puntos o píxeles por cada pulgada cuadrada, y $900 \times 600 = 540,000$ puntos o píxeles para el total de la imagen.

Al comparar las dos imágenes resultantes, puede percibirse que la primera de 75 dpi tiene menor definición y nitidez, puesto que

está formada por puntos o píxeles más grandes y por tanto más diferenciables al ojo humano; la segunda, digitalizada a 300 dpi es una imagen menos burda y de mayor resolución o nitidez; los puntos o píxeles, al ser más pequeños, dejan de ser diferenciables al ojo humano como tales y la imagen se vuelve una forma continua. Como conclusión obvia podemos afirmar que, a mayor número de puntos por pulgada los puntos son más pequeños y se perciben menos a la vista y por tanto la imagen es más fina, tiene mejor calidad y nitidez: en suma, tiene mejor resolución. Véase la siguiente figura donde presentamos una imagen rasterizada en forma de mapa de bits con píxeles muy grandes y diferenciables, esto es, baja resolución; luego una imagen rasterizada con puntos más finos, apenas diferenciables, esto es, mediana resolución, y finalmente una imagen rasterizada en alta resolución, con puntos todavía más finos y prácticamente no diferenciables a simple vista. La imagen es la misma, sólo que su resolución cambia en cada una de ellas.



Representación de una imagen en baja, media y alta resolución.
Puede verse como el tamaño de los puntos o píxeles disminuye y la imagen se vuelve más fina.

De esto suena muy fácil inferir entonces que siempre debe buscarse la mayor resolución posible, pues eso incrementa sin duda la calidad. Esto es real, pero existe un fenómeno que es fácil soslayar: el tamaño de la imagen crecerá exponencialmente a la resolución. La imagen que comentamos en el ejemplo en 75 dpi tiene en total 33,750 puntos; es decir, el archivo que guarda esta información será una secuencia de 33,750 conjuntos de números de valores que representan el color de cada punto más 33,750 números que guardan

Cómo preservar mi patrimonio digital personal

la posición de cada punto dentro la imagen. En el ejemplo de 300 dpi, la imagen tiene en total 540,000 puntos; esto es, conjuntos de valores numéricos de color y posición. La imagen es la misma, pero debido a la resolución con que ha sido construida o digitalizada la diferencia en espacio ocupado es enorme, y no multiplicado por 4 –la relación entre 75 y 300 dpi–; la diferencia en espacio ocupado entre 33,750 y 540,000 es exactamente 16 veces: el cuadrado de 4.

El total de valores de puntos del archivo, sea 33,750 o 540,000, se traduce en una enorme cantidad de conjuntos de números a guardar en un archivo en algún dispositivo de almacenamiento. Dado que los escáneres actuales pueden lograr típicamente estas resoluciones y llegan fácilmente hoy en día hasta 600, 1200, 2400, 3600 dpi y todavía más –imágenes ultrafinas–, el tamaño de los archivos resultantes puede ser numéricamente grande, de muchos Megabytes y como los escáneres siguen mejorando cada día, se corre el riesgo de que el espacio ocupado por cada imagen sea desproporcionadamente grande dentro de los discos o dispositivos electrónicos. Como podemos ver, si pasáramos la resolución de esa imagen de 300 a 600 dpi, el archivo resultante no ocupará el doble de este archivo, sino el cuadrado de éste. Al multiplicar el número de píxeles de una imagen por el número de bits requeridos para cada pixel, el número de bytes de una imagen puede llegar fácilmente a varios cientos de miles o millones de bytes por cada imagen.

Como puede verse, por fidelidad cromática o por resolución no hay impedimento a la hora de crear una imagen digital; el problema surge cuando –a mayor resolución y paleta cromática seleccionadas– el número de puntos a representar se vuelve muy grande y los valores de la paleta de cada punto crecen a números grandes, mismos que representados en binario ocupan muchos bits; es decir, mucho espacio de almacenamiento. Por lo mismo la resolución y la riqueza cromática son un arma de dos filos: una imagen de alta resolución proveniente de una hoja, digamos tamaño carta –8.5 x 11 pulgadas = 93.5 pulgadas cuadradas– va a estar formada por miles y miles de puntos cada uno de los cuales a su vez requiere de un número más o menos grande de bits cada uno dependiendo de su paleta cromática.

Puede verse entonces que una imagen carta de alta resolución y buena paleta tiene millones de puntos con valores numéricos grandes cada uno. Esto implica un gran número de bits consumidos para esa imagen digitalizada; almacenarla requiere de un gran espacio en un disco duro u otro dispositivo similar. Si la cantidad de imágenes a digitalizar es grande, el espacio requerido para su almacenamiento puede llegar a ser enorme, lo cual conlleva un costo oneroso. Si bien existen técnicas para compresión de estos archivos con objeto de que ocupen menos espacio en la vida real, debe considerarse que la compresión puede implicar pérdida de calidad de la imagen. Además del espacio consumido, debe considerarse además que entre más grande es una imagen, más tiempo toma a un escáner digitalizarla, más tiempo se toma en la red para transmitirla o recibirla, etc. Otras tareas se ven obstaculizadas también por un tamaño excesivo de un cierto archivo.

Existe entonces una importantísima relación de proporción y balance en este sentido: a menor resolución y riqueza cromática, menor calidad final de la imagen, pero es más económica en el espacio ocupado. A mayor color y mayor resolución, la imagen resultante tiene mayor calidad, pero ocupa más espacio y cuesta más. ¿Qué criterio debe prevalecer entonces, calidad o economía? Para poder responder a esta pregunta, el propósito de ese proceso de digitalización se vuelve muy importante. Si deseo preservar esas imágenes para el futuro, por periodos largos, debe prevalecer el criterio de calidad; si simplemente son imágenes que utilizo para la creación y uso temporal en alguna actividad entonces debe prevalecer el criterio de la economía. De este principio puede extraerse ese criterio necesario para tomar la decisión. Si nuestro propósito en ese momento es preservar a largo plazo un documento único o raro para nosotros y por ello muy valioso, y queremos obtener un nuevo “original” de él para preservarlo, debe hacerse el gasto; debe digitalizarse con la mayor calidad posible poniendo en segundo plano el costo de almacenarlo. Al fin y al cabo, los documentos raros o únicos no se dan por millones. Entre más valioso y único para nosotros sea un documento, mayor calidad debe imprimirsele en el proceso de digitalización y preservación. Por ejemplo, la ima-

Cómo preservar mi patrimonio digital personal

gen digitalizada de nuestro título profesional, o la digitalización de las fotos antiguas familiares que están en papel y que son tesoros irremplazables.

Por el contrario, si el propósito principal es digitalizar para uso en actividades temporales y poner al alcance de otras personas algún documento, el criterio principal es la economía, sobre todo si el documento no se considera raro o escaso. Por ejemplo, las fotos que distribuyo en las redes sociales; no es necesario enviar originales de ocho o diez Megapíxeles, ni fotos digitalizadas que ocupen 10 Megabytes cada una. Si utilizo ciertas imágenes como decoración o complemento de un blog u otro tipo de publicación personal, no necesito crear imágenes ni conservar copia de ellas en alta resolución. Pero debe valorarse con cuidado: no tiene caso tomar con la cámara imágenes de 10 Megapíxeles para después convertirlas en la computadora a una única copia de 1 Megapíxel o menos.

Por supuesto, el criterio de la economía no debe llegar a niveles que reduzcan tanto la calidad que hagan el material ilegible o no apto para su uso. Debe haber una calidad mínima con la cual el material es útil al usuario y no obstante sigue resultando económico de almacenar. Abundaré en ello más adelante.

Además de los escáneres, tenemos a las cámaras digitales como dispositivos para crear imágenes digitales de origen o para digitalizar imágenes ya existentes en papel. En términos generales, las cámaras se comportan de forma semejante a los escáneres, ya que se basan en el mismo principio de la imagen rasterizada; es decir, pueden tomar imágenes con mayor o menor resolución al igual que un escáner rompiendo la imagen que entra por la lente en un número mayor o menor de puntos. La diferencia esencial estriba en que la cámara no trabaja con una unidad de pulgadas cuadradas ni con dpi como el escáner, el cual recorta el tamaño de la imagen que se desea escanear. Una cámara digital posee su ventana del visor que es siempre igual, típicamente rectangular en una proporción de 5 a 4 en relación de la distancia horizontal con la vertical. En esa ventana rectangular se alojarán un cierto número de puntos que el usuario preselectiona y que forman una imagen para una fotografía dada gracias también a un Dispositivo de Carga Acoplada

o CCD. En comparación, sería el rollo de película de una cámara fotográfica antigua. Cada celdilla del CCD está recubierta con un filtro sensible a las tonalidades azul, verde y rojo cuyos valores pueden guardarse en forma de números binarios. Al número total de millones de puntos de celdillas en esa ventana que es siempre de la misma superficie se le conoce como Megapíxeles en vez de dpi. Un Megapíxel es poco más de un millón de puntos.

Como puede deducirse de lo anterior, la diferencia consiste entonces en que el escáner refiere sus unidades a puntos por pulgada cuadrada de imagen tomada, el cual deberemos multiplicar después por el número total de pulgadas cuadradas que conforman esa imagen para obtener el total de píxeles que ocupan su superficie total. En la cámara digital la superficie siempre es la misma, por lo que las unidades refieren el número total de puntos que conforman esa imagen, pudiendo variar solamente el tamaño de esos puntos –su resolución–. En nuestro ejemplo visto anteriormente, el de las imágenes a 75 y 300 dpi, si las midiéramos en unidades de cámara fotográfica, vemos que ocupan 33,750 y 540,000 puntos respectivamente, esto es, la primera es una imagen de aproximadamente 33 Megapíxeles y la segunda es una imagen de aproximadamente 540 Megapíxeles. Así, cuando una cámara “rompe” una imagen en su ventana en, digamos, 1280 por 1024 puntos, el total de la imagen fotografiada obtenida es de 1’310,720 puntos; el producto de los dos números anteriores. En números redondos una imagen de 1.3 Megapíxeles.

Cuando una cámara nos indica que tiene 5.2 Megapíxeles máximos nos dice que puede tomar fotografías seleccionando de entre varias resoluciones, pudiendo llegar hasta un máximo de 5.2 millones de puntos totales por cada imagen en su CCD; esto es, el máximo posible en esta cámara son imágenes de 2560 x 2048 puntos = 5’242,880, y la proporción seguirá siendo 5 a 4 entre la horizontal y la vertical. En la actualidad hay cámaras instaladas en telescopios que alcanzan los 1000 Megapíxeles –un Gigapíxel– pero son todavía sumamente voluminosas. En el ramo de las cámaras personales existen en la actualidad hasta de 80 Megapíxeles totales por imagen, lo cual equivale a una resolución de aproximadamente 10,000 x 8,000 puntos. Es sumamente alta, pero siguen mejorando; como

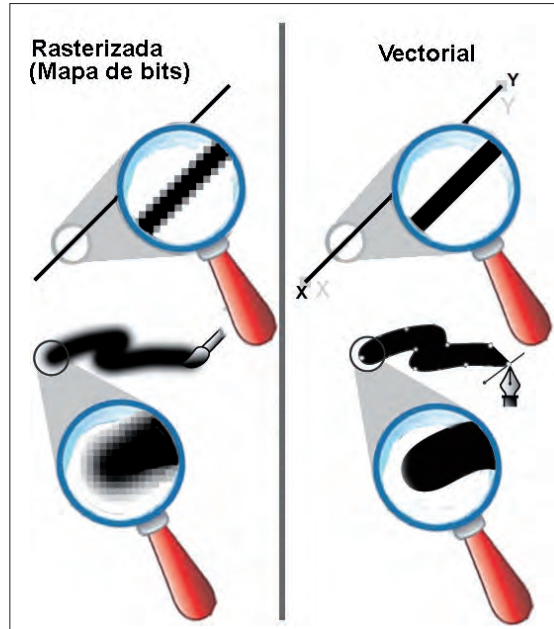
Cómo preservar mi patrimonio digital personal

referencia, el ojo humano promedio tiene alrededor de 127 millones de células cromáticas o 127 Megapíxeles. Obviamente, puede pedirse a la cámara que use menor resolución respecto a su máximo, bajando la resolución y por tanto el espacio requerido por cada imagen, pero por supuesto no pueden tomarse imágenes con esa cámara más allá de su resolución máxima estipulada de fábrica. Debe tenerse en mente que mayor resolución en la fotografía implica que pueden tomarse menos fotografías con una cierta dotación fija de memoria interna, su rollo digital por así decirlo, y viceversa. Una cámara con una tarjeta de memoria de 8 Gigabytes puede obtener alrededor de 4000 fotos de baja resolución con esa memoria, o 1600 de resolución media, o 500 de alta resolución. Otra ventaja de la cámara es sin duda su portabilidad, mucho mayor que la de un escáner.

Imágenes vectoriales

La segunda técnica para representar imágenes digitales además del *rastering* o malla realizada por escáneres y cámaras se conoce como técnica de *rendering* o de meta-vectores –del inglés *render*: interpretar, verter, presentar–. Esta técnica tiene su origen y es ampliamente utilizada por los programas que permiten “dibujar” o crear imágenes en la computadora. En este tipo de representación una imagen no es una matriz de puntos contiguos o mapa de bits; consiste en vectores de líneas y colores almacenados en fórmulas matemáticas. Los vectores se usan para crear líneas, formas, polígonos; prácticamente cualquier forma imaginable. El programa estipula cada uno de ellos en función de una fórmula matemática. Por ejemplo, si deseamos una línea a 45° desde un punto X hasta un punto Y para una distancia de 50 mm y con dos 2 mm de ancho ello se almacenará como una fórmula y no como una secuencia de puntos. Su enorme ventaja es que esta técnica es escalable sin mayor pérdida de calidad. Si la línea de 50 mm se escalase a 500 mm o a 5000 mm la proporción de todo lo demás se ajusta automáticamente. Todo esto permite las acciones típicas de estos programas

de cambiar escalas de imágenes, girarlas, proyectarlas en ángulo, realizar efectos cromáticos complejos, etc., propias de un programa de este tipo y sin pérdida de calidad de la imagen resultante. También pueden contener información embebida de datos cromáticos, de textura y apariencia, los metadatos de la imagen, o mezclas de metadatos y vectores, de ahí el nombre.



Ejemplos de líneas de malla y vectoriales. Nótese la diferencia al ampliar o escalar las imágenes.

En términos generales, y sin que sea una regla absoluta, las imágenes representadas con mapas de bits tienden a ser imágenes provenientes de la realidad: fotografías provenientes de una cámara, imágenes digitalizadas a partir de un escáner. Las imágenes vectoriales tienden a ser creadas en una computadora a partir de un programa al efecto: dibujos, animaciones, logotipos, diseños gráficos, etcétera.

Prácticamente todos los formatos utilizados hoy en día para la digitalización de imágenes obedecen a los dos principios y técnicas

Cómo preservar mi patrimonio digital personal

enumerados anteriormente. Como ya mencionamos, existe más de una manera de representar numéricamente esos mapas de bits o esos vectores. Cada convención distinta es un formato, el cual obedece a los propósitos de personas u organizaciones que los crean, al tipo de material representado, etcétera.

Compresión de imágenes

Hemos analizado cómo ciertos factores como la resolución, la paleta de color, el tamaño de la imagen original, etc., están estrechamente entrelazados e influyen sensiblemente sobre el tamaño final, también llamado “peso” de un archivo con una imagen. Este no es un problema nuevo; por ello, en las últimas décadas se han invertido muchos esfuerzos en investigación y desarrollo con el propósito de comprimir cada vez más los documentos tratando de que ocupen menos espacio y tiempo de transmisión. Este propósito ha influido en buena forma en la creación o uso de formatos de imágenes y algunos de ellos son fruto directo de esas investigaciones.

Seguramente todos hemos usado alguna vez un archivo “.zip” o “.rar” que teníamos que descomprimir previamente para obtener el o los archivos que se encontraban “empacados” dentro del mismo. Este es un ejemplo de compresión de archivos, sólo que es un caso explícito; muchos archivos llevan la compresión en forma inherente, de otro modo no podrían existir de forma práctica. El advenimiento de los documentos multimedia en forma de imágenes, audio o video así como de la transmisión masiva de información vía Internet desencadenó un auge inusitado en la investigación y desarrollo de las técnicas de compresión de la información, ya que como hemos visto la cantidad de bytes o caracteres de un archivo de información cualquiera incide sensiblemente en el espacio de almacenamiento requerido en un cierto dispositivo así como en el tiempo en que ese archivo tarda en ser enviado o recibido en la red; si no existieran estas técnicas, los teléfonos celulares no serían prácticos, ni tampoco el audio o video transmitido entre

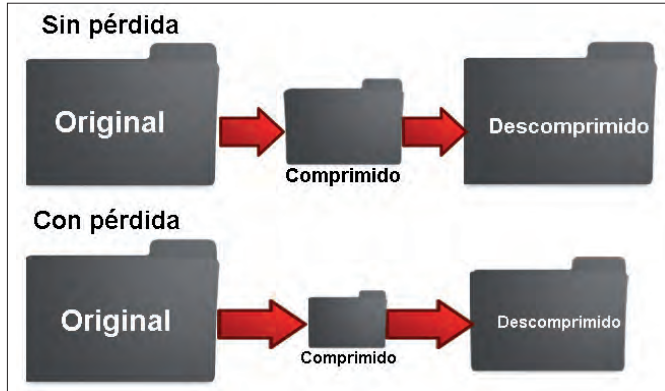
computadoras. Hoy en día, las películas en DVD, las transmisiones de televisión digital y las de televisión satelital no existirían sin las técnicas de compresión. Por ejemplo, la representación digital de un segundo de video sin compresión, usando el formato CCIR 601, ocupa más de 20 Megabytes, el equivalente a veinte millones de caracteres de texto. Si consideramos el número de segundos de un video, digamos de 10 minutos o 600 segundos, significaría un consumo de 12,000 Megabytes, o doce Gigabytes, o doce mil millones de caracteres sólo para diez minutos de video; justo la capacidad actual del DVD más amplio por los dos lados. Sin compresión, el video digital no sería práctico. Con este ejemplo puede verse el porqué se requiere forzosamente de técnicas de compresión al momento de crear y guardar un archivo de este tipo. La representación digital de un minuto de música en formato CDDA no comprimido de cualquier disco compacto musical comercial –44,100 muestras por segundo, 16 bits por muestra– ocupa más de 9 Megabytes, o 9 millones de caracteres. Tratar de descargar música o video de la red con estos tamaños de archivo sería absolutamente impráctico por el tiempo que el proceso consumiría o requeriría de hacerlo siempre por banda ancha.

Como usuario personal no tienes por qué conocer al detalle esas técnicas, pero es importante que sepas de su existencia, de su importancia, de sus formatos asociados y sobre todo, de sus consecuencias y efectos en el manejo y preservación de documentos y estar al tanto de sus posibilidades y avances. Ello te permitirá evolucionar con estos desarrollos y te permitirá tomar mejores decisiones sobre cuál formato usar para construir una colección documental dada. Recordemos que los formatos existentes en la actualidad no durarán para siempre; van evolucionado rápidamente: nacen y desaparecen y es necesario por tanto moverse junto con ellos, ya que ninguno ha prevalecido a lo largo de muchos años. A manera de resumen examinemos, aunque sea someramente, cuáles son las técnicas de compresión más estudiadas y desarrolladas en la actualidad y sobre qué principios están fundamentados algunos de los formatos más utilizados hoy en día y algunos de los que fundamentarán los formatos del futuro cercano.

Cómo preservar mi patrimonio digital personal

De inicio, es conveniente mencionar que prácticamente todas las técnicas de compresión se basan en principios y algoritmos matemáticos aplicados a un proceso de la vida real. Un algoritmo es un método basado en un conjunto ordenado y finito de pasos y operaciones bien definidas que permite hallar la solución de un problema en una secuencia cíclica o repetitiva. Tiene ciertas características: 1. *Definibilidad*: el algoritmo debe poder definirse de forma precisa para cada paso, sin ninguna ambigüedad, mediante un lenguaje formal, ya sea matemático o de programación de computadora. 2. *Finitud*: el algoritmo debe acabar tras un número finito y razonable de pasos. 3. Tiene estado inicial y datos de entrada bien definidos y produce datos de salida bien definidos. En este caso el problema que pretende solucionar un algoritmo de este tipo es la compresión de un archivo digital.

Existen dos técnicas fundamentales de compresión de datos: sin pérdida *-non-lossy-* y con pérdida *-lossy-*. En la primera, cualquier proceso de compresión de datos se basa en el principio fundamental de que en el algoritmo o proceso de compresión al cual entra un documento A y obtenemos un documento comprimido B que ocupa menos espacio que el original, al aplicar el proceso inverso y someter por esa técnica el documento comprimido B deberemos obtener como resultado el documento original A idéntico. Esto es, el procedimiento es reversible al 100%. Si esto se cumple, se denomina a esta técnica sin pérdida *-non-lossy-*. En el segundo tipo de técnica, el archivo resultante, ya comprimido, ha perdido algo de su calidad original, y al aplicar un proceso inverso no se tendrá el archivo original del cual se partió; a cambio de ello se logran mayores niveles de compresión. El archivo ya comprimido no contiene la misma información que el original, y por tanto en un proceso inverso no se podrá tener el archivo idéntico del cual se partió, sino sólo una aproximación.



Técnicas de compresión y descompresión de archivos.

El que una técnica de compresión sea con o sin pérdida no tiene que ver con la aceptación o rechazo de ella; depende de su aplicación, de su contexto; ninguna técnica es mejor que otra por sí misma. Por ejemplo, al comprimir la voz humana en un teléfono podemos eliminar las frecuencias que no son audibles para el oído humano, así que su eliminación al momento de comprimir no tiene mayor efecto sobre la reconstrucción de un original no idéntico para la persona que escucha y por lo tanto no importa que la reconstrucción no sea idéntica al original: en este caso, la compresión con pérdida es perfectamente aceptable. En el caso de un texto es indispensable a la hora de reconstruir el archivo que el texto descomprimido sea idéntico al original; la pérdida de partes del texto sería inaceptable. Por lo mismo, en algunos casos se evalúa un algoritmo de cierta manera y en otros de forma distinta, dependiendo del contexto y uso del documento a comprimir.

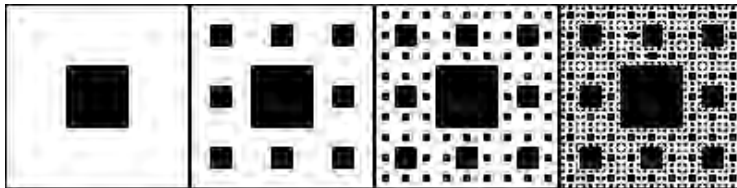
Cada técnica o algoritmo –a cambio de ningún o cierto nivel de pérdida de fidelidad con respecto al original– entregará factores de compresión de mayor o menor grado; es decir, factores de compresión de dos a uno, de cuatro a uno, dieciséis a uno, etc., respecto al original, lo cual se convierte en un elemento de medición y calificación del algoritmo. Otras variables que sirven para calificar una técnica dada son la cantidad de memoria que consume en la computadora a la hora de efectuar el proceso, la velocidad a la

Cómo preservar mi patrimonio digital personal

que se puede comprimir masivamente un conjunto de documentos, la velocidad de descompresión, etc. Al existir entonces numerosas técnicas propuestas para comprimir un tipo de documento en particular es necesario estar consciente de las ventajas y desventajas de alguna técnica en especial empleada en función del documento a comprimir. Entre las técnicas más utilizadas y estudiadas hoy en día distinguimos:²⁷

- ❖ Compresión fractal
- ❖ Codificación de Shannon-Fano y Huffman
- ❖ Compresión LZW –Lempel-Ziv-Welch–
- ❖ Compresión RLE –Run Length Encoding–
- ❖ Codificación aritmética
- ❖ Código predictivo
- ❖ Cuantización escalar y vectorial
- ❖ Algoritmos de compresión tipo JPEG
- ❖ Grupo 4

Esta no es ni con mucho una lista exhaustiva, simplemente es un ejemplo para ilustrar algunos de los métodos más representativos de las técnicas de compresión, sus ventajas y desventajas, y para conocer un poco acerca de las técnicas más utilizadas en relación a los documentos, sean textos, imágenes, audio o video.



Fractal: "Alfombra" de Waclaw Sierpinsky.

La compresión fractal es una técnica que básicamente retira elementos redundantes de la representación de una imagen. En ella, en vez de ir almacenando para cada pixel o punto su color y posición se crea un conjunto de instrucciones que permita reconstruir

27 Khalid Sayood, *Introduction to Data Compression*.

toda una sección de una imagen. La teoría de los fractales no es reciente; sin embargo su uso moderno con imágenes data de 1977.²⁸ Un fractal es una figura geométrica cuya estructura básica, fragmentada o irregular, se repite a diferentes escalas, y esta repetición es lo que la hace ideal para comprimirse. Su importancia es tal que hasta hoy en día una gran cantidad de juegos electrónicos generan sus escenografías y personajes virtuales basados en técnicas de fractales. De manera inversa, una imagen real puede ser reducida a ciertos grupos de instrucciones, aunque sea parcialmente, y no a descripciones punto por punto. En el ejemplo, la Alfombra de Sierpinsky, cada figura está compuesta por elementos de la que está a su izquierda; la figura básica es un cuadrado. Esto puede ser representado matemáticamente para no tener que guardar el dato de cada punto, ocupando así mucho menos espacio. Esta técnica, hasta cierto nivel de compresión, resulta sin pérdida; más allá de ese nivel se logra mayor compresión, pero comienza a tener pérdida.

La técnica de compresión conocida como Shannon-Fano fue desarrollada por Shannon en los Laboratorios *Bell* y por Ugo Fano en el Instituto Tecnológico de Massachusetts –MIT– presentada en el año 1948.²⁹ A partir de sus obras, Claude Shannon³⁰ es considerado como el primero que unió los trabajos anteriores en algo coherente que fue denominado la teoría de la información, de la cual parten la mayoría de las técnicas actuales. Se basa en una técnica de medición y compresión de datos en forma estadística la cual brinda una reducción considerable en datos de tipo alfabético. Esta técnica fue perfeccionada por David Huffman en los cincuentas y es ampliamente utilizada hoy en día como algoritmo de compresión de textos así como base para compresión de otro tipo de datos; se usa principalmente en redes de cómputo. Básicamente, esta técnica se

28 Benoit Mandelbrot, *La geometría fractal de la naturaleza*. Un fractal es un objeto semigeométrico cuya estructura básica, fragmentada o irregular, se repite a diferentes escalas. Al ser repetitivo, ese objeto puede guardarse sólo una vez; las restantes pueden quitarse, dejando sólo una “marca” de que deben reinserirse después, de acuerdo con la muestra de la primera.

29 Aaron D. Wyner, *The Significance of Shannon's Work* [en línea].

30 Claude Shannon, “A Mathematical Theory of Communication”.

basa en la probabilidad de frecuencia de ocurrencia de un carácter dentro de un texto en una lengua dada. Esto se ha logrado contando la incidencia de letras en múltiples textos de un idioma, y se le asignan códigos binarios más reducidos en número de bits que los ocho usuales del código ASCII.

La técnica de compresión sin pérdida LZW –Lempel-Ziv-Welch– fue ideada originalmente por Jacob Ziv y Abraham Lempel, y publicada en los anales del Instituto de Ingenieros Eléctricos y Electrónicos –IEEE– en 1977;³¹ fue refinada por Terry Welch en 1984 y de ahí tomó el nombre completo. Dependiendo de la época, han sido conocidas como técnicas LZ-77, LZ-78 y así sucesivamente. El algoritmo LZW es sorprendentemente simple: sustituye cadenas de bits que se repiten, –llamadas “cuerdas”– con códigos más simples y construye progresivamente un diccionario de equivalencia de las cadenas originales y los códigos que las sustituyeron. Al ser estos códigos más simples y menores en tamaño que las cuerdas a las que sustituyen, ocurre la compresión. Esta técnica ha sido ampliamente utilizada y dió origen, entre otros, al formato GIF para imágenes. Fue patentada como tal por la empresa Unisys, por lo que el formato PNG, formato libre sucesor de GIF, no lo utiliza, quedándose con la versión LZ-78.

La técnica de compresión RLE o *Run-Length Encoding* es una forma muy simple de compresión de datos sin pérdida en la que secuencias de puntos con el mismo valor son sustituidas con el valor del dato más el número de veces que ocurre; por ejemplo, si existen 50 puntos consecutivos del mismo color no se guardan cada uno de los valores de esos 50 puntos, sino que se sustituye por la información de un punto y se indica que éste se repetirá por 50 veces. Es muy útil en datos que contienen muchas de estas secuencias de un solo color; por ejemplo, gráficos sencillos con áreas de color plano como logotipos, partituras, mapas, etcétera.

La codificación aritmética es una versión más avanzada de las técnicas de Huffman también sin pérdida. La diferencia básica con-

31 Jacob Ziv y Abraham Lempel, “A Universal Algorithm for Sequential Data Compression”.

siste en que en esta última se requieren potencias de números enteros de bits para poder hacer la representación de un carácter mientras que en la compresión aritmética se usan potencias no enteras, es decir, con números reales, lo cual hace todavía más eficiente el algoritmo de compresión, y se logra usar todavía menos caracteres para la representación; su desventaja es la complejidad. Mientras que el algoritmo de Huffman es bastante simple y por lo mismo fácilmente programable, la compresión aritmética resulta mucho más compleja en su desarrollo y programación.

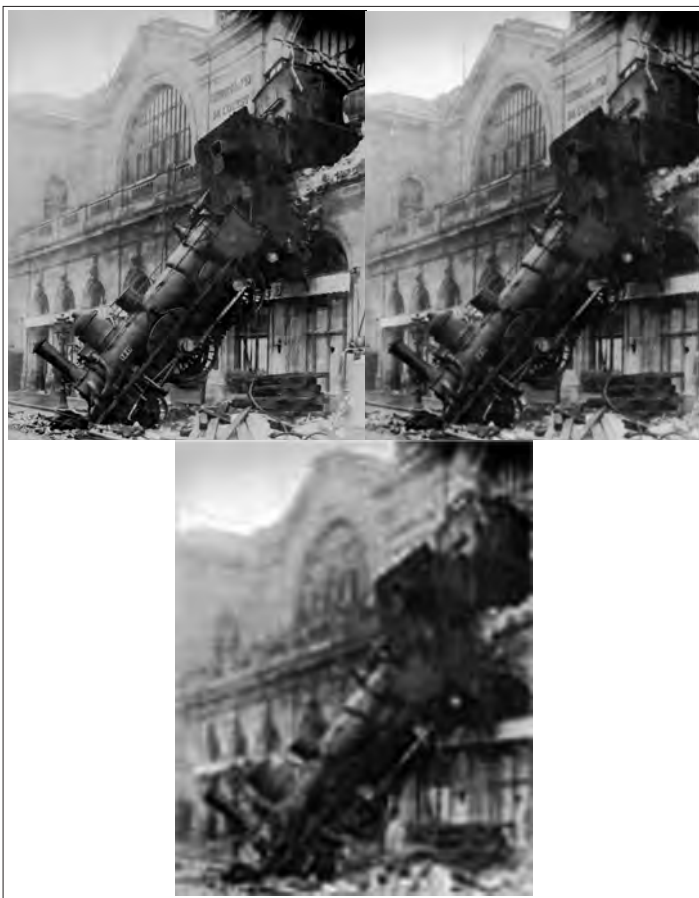
La técnica del código predictivo se basa en ocurrencias pasadas de algún tipo de datos para estimar la probabilidad de que ese dato se presente de nuevo; es decir, se comprime en mayor proporción lo que más se ha presentado. Ello permite una compresión muy eficiente. Entre estas técnicas se distinguen la predicción con coincidencia parcial –*Prediction with Partial Match* o PPM–, las transformadas de Burrows-Wheeler o BWT, la técnica CALIC –*Context Adaptive Lossless Image Compression*– o compresión de imágenes sin pérdida por contexto adaptativo.

La técnica de cuantización escalar se basa en el proceso de representación de un conjunto de valores de palabras-código que a su vez representan un conjunto aún mayor de valores de palabras fuentes. Esta es una técnica sumamente eficiente, pero es del tipo con pérdida. La cuantización vectorial es semejante pero trabaja con bloques de datos a la vez, llamados vectores.

La técnica de compresión tipo JPEG fue desarrollada por el *Joint Photographic Experts Group*, un comité que definió ese estándar. Se diseñó expresamente para la compresión de imágenes fijas: fotografías, mapas, carteles, ya sea a color o en blanco y negro. Es una técnica con pérdida, pero que logra sensibles factores de compresión de una imagen hasta un séptimo del tamaño del original en el volumen de bytes, con sólo una pérdida de 25% en calidad. Fue diseñada para distribución de imágenes vía la red, donde la pérdida para ese propósito y por ese medio era bastante aceptable. Si comparamos los tamaños de una misma imagen en un formato, digamos, tiff, el cual es sin pérdida, o jpg, con mínima pérdida, podemos observar la enorme diferencia entre el volumen de uno y

Cómo preservar mi patrimonio digital personal

otro, mientras que la calidad no disminuye en la misma proporción. Este formato jpg es por tanto muy recomendable para distribuir materiales en Internet, pero no tan recomendable con fines de preservación de imágenes a largo plazo a menos que se usen pérdidas aceptables.



Resultado de tres imágenes al aplicar sucesivamente a un original un algoritmo de compresión-descompresión con pérdida considerable.³²

32 *Accidente ferroviario en la estación Montparnasse, 1895, París.*

Otro estándar de compresión internacional adoptado con amplitud en la actualidad tiene que ver con el envío de datos a través de la red. El más utilizado en la actualidad es el conocido como Grupo 4 –*Group 4 for 1-bit images*–. Este estándar fue introducido originalmente en 1988 para módem y fax por el CCITT –*Comité Consultatif International Télégraphique et Téléphonique* o *International Consultative Committee on Telecommunications and Telegraphy*–. El CCITT cambió su nombre en 1993 a ITU-T –*International Telecommunications Union–Telecommunications Standardization Sector*–. Las especificaciones han evolucionado naturalmente a las necesidades de la época.

Como pudo verse en este apartado, la compresión y tipo de técnica a utilizarse se selecciona de acuerdo con un criterio que logre el adecuado balance entre el propósito que busca el formato y su eficiencia al almacenarse y distribuirse. No existe la técnica perfecta de compresión que maximice todas las ventajas y por ello debe ser analizada y seleccionada de antemano de acuerdo con un propósito definido y un criterio. Lo importante para el usuario es estar siempre al tanto de cuál es el formato –y por tanto compresión– en que cada una de sus colecciones es digitalizada y preservada con objeto de prevenir pérdidas no deseadas que en un futuro puedan resultar contraproducentes. Abundaremos en la aplicación práctica de compresión y formatos más adelante.

Formatos de imagen más utilizados

Una vez que hemos entendido las maneras en que se hace la representación gráfica de las imágenes en un archivo digital así como las técnicas de compresión asociadas, podemos estudiar una reseña de los principales formatos usados en la actualidad y que tienen relación con la preservación y la distribución documental. Algunos de estos formatos han sido creados para máxima calidad, resolución y colores; de más está decir que consumen mucho espacio. Otros han sido creados para optimizar el espacio requerido y tiempo de transmisión, pero su calidad es reducida.

❖ *tiff –Tagged Image File Format–* es un formato de mapa de bits muy flexible y distribuido que es soportado ya por una amplia

variedad de aplicaciones de procesamiento de imágenes y plataformas. Entre sus principales características se encuentran: soportar varios tamaños de pixel y por tanto paleta variable; un archivo puede contener varias imágenes; puede usar distintos algoritmos de compresión, un sólido conjunto de metadatos embebido, etc. Dada la alta calidad de imágenes que produce y que es el estándar de los equipos Apple diseñados con enormes ventajas específicas para la creación y manejo de imágenes, así como su aceptado uso en equipos PC, se ha convertido en el estándar más ampliamente aceptado para la preservación de imágenes.

- ❖ *bmp –Bitmap–* es un formato para imágenes creado por las empresas Microsoft e IBM para MS-DOS y Windows para computadoras PC. Fue uno de los primeros formatos en este tipo de documentos y su estructura es muy sencilla. Admite los modos de color RGB, color indexado y escala de grises y una profundidad de color de hasta 32 bits por canal de la imagen. Por lo general no tiene compresión de datos, aunque en algunas imágenes de 4 y 8 bits se puede especificar compresión tipo RLE. El uso más común de este formato es generar imágenes de poco peso para su uso en la red y en imágenes reducidas a 24 bits.
- ❖ *tga –Truevision Targa–* es un formato de mapa de bits desarrollado por la empresa Truevision para sus tarjetas Targa y Vista, tanto para PC como para Apple; permite guardar imágenes monocromáticas –2 bits–, escala de grises –4 bits– y cromáticas con diferentes paletas de profundidad de color –8, 16, 24 y 32 bits– con muy alta calidad. Permite almacenar archivos comprimidos o sin comprimir, aunque la mayoría de programas que lo soportan sólo pueden abrir archivos tga sin compresión, siendo entonces que el tamaño de los archivos resultantes es muy grande. Este formato es especialmente utilizado para editar diseños gráficos profesionales que se vayan a reproducir en pantallas, debido a que su amplia paleta de colores produce un efecto muy realista y sumamente elaborado. También es muy útil cuando se trabaja con escáneres de alta calidad y para la exportación de imágenes a edición profesional de video. Para impresión es poco usado, ya que con profundidades de color de 16 bits o menos las imágenes

pierden mucho detalle. Su principal desventaja es el tamaño de los archivos, al ocupar bastante más espacio que otros formatos de igual calidad.

- ❖ gif –*Graphic Interchange Format*– es un formato de representación vectorial desarrollado originalmente por la empresa CompuServe, cuya patente pertenece hoy a la compañía Unisys. Ha sido durante mucho tiempo uno de los formatos de imagen más populares en Internet, sobre todo debido al tamaño compacto de sus archivos. Es ideal para imágenes creadas por computadora: logotipos, pequeños iconos de navegación o adorno de páginas web, diagramas simples e ilustraciones donde se requiere precisión o gráficas con bloques grandes de un solo color, tanto fijos como animados. El formato es sin pérdida, significando que no se hace borroso o sucio. Los archivos gif usan una técnica simple de compresión LZW para reducir el tamaño de las imágenes encontrando patrones repetidos de píxeles, sin que esta compresión degrade la calidad de imagen. Maneja una paleta reducida de hasta 256 colores. Su principal desventaja es que al ser un formato propietario los desarrolladores de aplicaciones deben pagar a su titular por el uso lo cual hace que cada vez más vaya cayendo en desuso en favor del formato png, igual en características pero con acceso libre.
- ❖ png –*Portable Network Graphics*– es un formato recién inventado expresamente para la web en respuesta al esquema de licencias introducido por CompuServe que significó a los creadores de cualquier programa que soportara el formato gif el pago de altos derechos por su uso; por lo mismo, el uso de png está libre de regalías y actualmente está avalado por el Consorcio Mundial para el desarrollo de la World Wide Web, conocido también como W3C. Una versión del formato, PNG-8, es similar al formato gif en sus funciones y también maneja una paleta de 256 colores, pero con mayor eficiencia en la compresión y por tanto en el tamaño del archivo resultante. El PNG-24 es otra variante que maneja una paleta de colores enorme semejante al jpg, aunque sin la eficiencia de compresión de aquél. Una característica distintiva de png y que lo hace muy atractivo sobre los

otros es su manejo de imágenes desde totalmente transparentes hasta totalmente opacas, lo cual permite sobreponer imágenes, lograr “marcas de agua”, tapices, etc., de manera interesante y atractiva en las páginas web. Es muy pronto todavía para decir si sustituirá a jpg para la edición en web, pero sin duda ha ido sustituyendo cada vez más a gif.³³

- ❖ *svg –Scalable Vector Graphics* o Gráficos Vectoriales Escalables– es un formato avalado por el Consorcio Mundial para el desarrollo de la World Wide Web para describir imágenes. Al igual que gif y png, puede representar tanto imágenes estáticas como animadas. Aunque está hecho principalmente para representar imágenes de tipo vectorial, también se pueden embeber imágenes de tipo mapa de bits. Se ha ido popularizando recientemente pero es muy pronto todavía para decir si sustituirá a gif y a png para la edición en web.³⁴
- ❖ *jpg o jpeg –Joint Photographic Expert Group* o Grupo Unido de Expertos en Fotografía– es un formato de mapa de bits diseñado por esa asociación con fines de distribuir fotografía digital.³⁵ Es capaz de desplegar millones de colores en su paleta y maneja perfectamente la mezcla compleja en cuanto a matices en imágenes fotográficas. Como tiene una capacidad muy grande de compresión se pueden obtener archivos muy pequeños; por esta razón, aunado al hecho de que su uso es libre de regalías se ha difundido enormemente para uso en páginas web, dado el efecto aligerador sobre ellas; fue de hecho un estándar *de facto* para distribución de imágenes en la red antes de volverse un estándar de norma *ISO/IEC 10918-1:1993(E)*. Este formato usa un algoritmo de compresión complejo que puede ser aplicado sobre una escala móvil; la compresión es conseguida “olvidando” ciertos detalles sobre la imagen, que el formato tratará entonces de lle-

33 *Portable Network Graphics (PNG) Specification (Second Edition)*. Information technology. Computer graphics and image processing. Portable Network Graphics (PNG): Functional specification. ISO/IEC 15948:2003 (E) [en línea].

34 World Wide Web Consortium, SVG, 2001.

35 Joint Photographic Experts Group (JPEG). Sitio web oficial: <http://www.jpeg.org>

nar más tarde cuando la imagen se despliegue. Puede salvarse una imagen jpg con la compresión del 0% para una imagen perfecta con tamaño de archivo grande, o con compresión del 80% para una imagen pequeña pero perceptiblemente degradada. En el uso práctico, un ajuste de compresión aproximadamente del 40 al 60% llevará al equilibrio óptimo entre calidad y tamaño con fines de distribución en web. El formato jpeg actual muy probablemente será sustituido en algún punto en el futuro por el formato jpeg2000 actualizado.³⁶

- ❖ jpeg2000 también está producido por el *Joint Photographic Expert Group* como una versión perfeccionada del primer jpeg. Es un formato muy avanzado: posee alta calidad para preservación pero ocupa menor espacio que tiff o RAW. Puede contener diferentes resoluciones en el mismo archivo, por lo que evita tener que guardar varias copias en diversas resoluciones para diversos propósitos. Tiene muy buenos algoritmos de compresión con mínima o ninguna pérdida. Soporta metadatos en XML, y además posee equivalencias con Dublin Core. Sus especificaciones están publicadas por lo que se considera *software* abierto. Está aprobado como estándar *de facto* y *de jure*.
- ❖ PCX fue creado por la empresa Z-Soft como su formato principal para el programa de edición gráfica PC Paintbrush. Es uno de los formatos de mapa de bits más conocidos; soporta imágenes de hasta 24 bits de profundidad de color –16.7 millones de colores– sin limitación respecto al tamaño de las imágenes. Usa un algoritmo de compresión RLE de codificación sencilla y alta velocidad de descompresión, orientado más a la rapidez de acceso que a la reducción de tamaño de los archivos. Su principal ventaja es la compatibilidad, ya que un gran número de aplicaciones gráficas de los programas de edición y tratamiento de imágenes –*desktop publishing*– lo soportan.

36 Joint Photographic Experts Group (JPEG), *Information Technology. JPEG 2000 Image Coding System: Code Coding System* [en línea] e International Organization for Standardization (ISO) *Information Technology. JPEG 2000 Image Coding System: Extensions* [en línea].

Cómo preservar mi patrimonio digital personal

- ❖ PICT de *Macintosh Quickdraw* es sumamente utilizado en la industria editorial para formar los originales de libros y revistas. Este formato comprime muy eficazmente las imágenes con tonos parecidos sin perder calidad. Trabaja con 16 o 32 bits de color.
- ❖ RAW –del inglés *raw material*, materia prima– es un formato de imágenes que contiene la totalidad de los datos de la imagen tal como ha sido captada por el sensor digital de un escáner o cámara. No hace manipulaciones para estabilizar o perfeccionar la imagen y no lleva compresión alguna. Por ello, y ya que contiene la totalidad de los datos de la imagen captada por el dispositivo, alta profundidad de color –30 a 36 bits por pixel–, los archivos resultantes ocupan mucho espacio. Por lo general lo usan quienes están interesados en guardar imágenes tal como han sido captadas por el sensor CCD, sin ningún tipo de tratamiento, para poder procesarla posteriormente en la computadora mediante un programa editor de imágenes. Su gran inconveniente es la falta de estandarización; cada fabricante produce su formato, lo que convierte a la mayoría de las versiones en *software* propietario y por ello no se ha popularizado como formato de preservación.

Finalmente, cabe mencionar que existen muchos otros formatos para almacenamiento de imágenes fijas: ai, clp, dcx, fpx, img, mac, msp, pct, ppm, psd, rle, wpg, rif, etc. Estos son sólo algunos de los acrónimos que representan otros tantos formatos para almacenamiento de imágenes. Con el advenimiento de los programas de computadora conocidos como editores gráficos pero sobre todo con las cámaras digitales toda una colección de ellos ha aparecido. La presentada en los párrafos anteriores no fue una lista exhaustiva, simplemente una reseña de lo más aceptado en el momento para el manejo de las imágenes a nivel personal.

Formatos de audio

Ya hemos visto con anterioridad cómo es que se representa una onda sonora en forma analógica y en forma digital. Recordemos que el proceso de digitalización del sonido se logra representando

la forma de la onda en un instante dado con un valor numérico lo que en conjunto nos da las cuatro propiedades básicas del sonido: altura, duración, timbre e intensidad. Tomando muestras de la onda a intervalos de tiempo iguales y preestablecidos podemos describir esa onda sonora en todo momento en forma numérica.

Un concepto igualmente importante en los archivos de audio es el principio de la resolución que ya analizamos en los formatos de imágenes; la resolución en imágenes tiene su equivalente en el sonido: se mide en tasa de bits *-bit rate-* por segundo; esta es la tasa o número de muestras de la onda sonora que se obtiene por unidad de tiempo. Dentro de los límites del principio de Nyquist-Shannon, a mayor tasa de bits muestreada, mayor cantidad de información para los circuitos, canales y bocinas y el sonido tendrá por lo mismo mayor calidad, pero conlleva a la vez mayor espacio consumido para almacenamiento y transmisión del archivo. Inversamente, a menor tasa de bits la cantidad de información que representa al sonido disminuye decreciendo con ello el espacio requerido para el archivo pero también su calidad. Los formatos creados y utilizados al efecto siguen estos principios básicos.

Una pieza de audio grabada en forma numérica bajo un cierto patrón o formato contiene típicamente tres partes: 1. La “envoltura” *-wrapper-*, en la cual se registran características generales y complementarias al archivo: licenciamiento de uso, capacidad de reproducción en-línea del sonido *-streaming capability-*, el intérprete, álbum, la letra o texto de la pieza musical, etc. 2. Un encabezado *-header-*, el cual almacena información técnica acerca del archivo en sí, tal como la resolución o tasa de muestreo del sonido, tipo de compresión utilizada, códec utilizado, etc. 3. Finalmente vienen los datos de información propios del sonido de la pieza de audio bajo una cierta estructura conocida como códec. Un códec *-acrónimo de codificador/decodificador-* es simplemente una estructura o técnica específica para la codificación y decodificación de ciertos datos, reduciendo su tamaño al mismo tiempo. Por ejemplo, entre los tipos de codificación o códecs más utilizados se encuentra el denominado PCM *-Pulse Code Modulation-* o código de muestreo de la señal de audio. Dado que es un código genérico, puede ser leído por la ma-

Cómo preservar mi patrimonio digital personal

yoría de los sistemas de audio, de manera similar a la que un archivo de texto plano tipo txt puede ser leído por todos los programas de procesamiento de texto. PCM es utilizado como códec base para los reproductores de discos compactos musicales y cintas digitales de audio tipo DAT o *Digital Audio Tapes*. PCM es también el códec que usan formatos para computadora muy comunes como aiff –*Audio Interchange File Format*– y wav –*Wave Form Audio Format*–. Puede existir más de un códec dentro de un formato. Posteriormente el archivo se graba con mayor o menor compresión.

Al igual que con las imágenes, los formatos más utilizados para preservación o almacenamiento de largo plazo para documentos de tipo audio tienen por lo general altas tasas de muestreo del sonido y poca o nula compresión; la calidad del sonido en estos formatos es por lo tanto alta pero consumen por lo mismo grandes cantidades de espacio en el archivo y su consecuente almacenamiento y enormes tiempos de transmisión por red. Cuando la intención es almacenamiento de corto plazo, distribuir sonido por la red, presentar pequeños trozos de audio, almacenar para fines de reproducción o entretenimiento y no está en juego la preservación documental, utilizamos formatos con alta tasa de compresión y mínimo espacio de almacenamiento.

Derivado de lo anterior, es pertinente revisar los formatos de audio más utilizados en la actualidad para ambos propósitos:

❖ CDDA o CDA –*Compact Disc Digital Audio*–. Es el formato nativo de todos los discos compactos musicales comerciales que vemos en las tiendas de todo el mundo, con frecuencias de muestreo de 44.1 kHz, 16 bits de muestreo y dos canales. Tiene alta calidad pero por lo mismo ocupa mucho espacio, entre 8 y 10 Megabytes por minuto de grabación. Un disco compacto musical promedio tiene como máximo capacidad de 74 minutos de grabación. Debido a que tiene una estructura muy diferente a la de los archivos de computadora se utiliza mucho en la preservación de sonido en reproductores domésticos de sonido pero poco para preservación de archivos en computadoras; cuando éste es el propósito y medio, por lo general es transformado hacia formatos wav, aiff o MPEG-4 versiones SLS, ALS, o DTS.

- ❖ wav –*Wave Form Audio Format*–. Formato nativo y estándar de sonido creado para ser almacenado en las computadoras con sistema operativo Windows. Aunque no usa compresión, pueden preseleccionarse desde tasas bajas hasta altas tasas de muestreo: desde los 8 kHz hasta 192 kHz por segundo, por lo que puede ser utilizado para archivos de poca calidad hasta archivos con muy alta calidad, aún mayor que la del CDDA. Muy utilizado en instituciones para preservación y distribución dentro de empresas, ya que con frecuencia tiene embebidos altos controles de licenciamiento y copiado, pero por lo mismo no es aceptado totalmente por los usuarios personales y no ha podido posicionarse como estándar universal. No es muy usado para compartir archivos vía la red, ya que existen otros formatos de audio sin pérdida que reducen mucho más el tamaño de los archivos.
- ❖ aiff o aif –*Audio Interchange Format File*–. Formato de alta calidad para preservación sin compresión muy utilizado en el ambiente de computadoras marca Apple. Soporta hasta 44.1 kHz y 32 bits de muestreo. Usa el códec de impulsos codificados PCM. Por ser sin compresión, sus archivos ocupan mucho espacio, aunque también existe una variante estándar conocida como AIFC que sí posee compresión.
- ❖ RA o RM –*Real Audio*–. Creado por la empresa Real Networks. Puede lograr muy buena compresión y descompresión, pero por lo general su calidad no es adecuada para aplicaciones de preservación a largo plazo; es muy aceptado en la distribución de señales de audio a través de internet en vivo –llamado en tiempo real o *streaming*–. Debido a este uso sus archivos pueden programarse con opciones que inhiben su copia.
- ❖ wma –*Windows Media Audio* o *MS Audio*–. Creado por Microsoft para competir con mp3 y RA. pueden preseleccionarse desde tasas bajas hasta altas tasas de muestreo: desde los 8 kHz hasta 192 kHz por segundo, por lo que puede ser utilizado para archivos de poca calidad hasta archivos con muy alta calidad. No ha podido desplazar a mp3 o MPEG.
- ❖ FLAC –*Free Lossless Audio Codec*– es otro códec de compresión sin pérdida, y consigue reducir el tamaño de un archivo de so-

nido original de entre la mitad hasta tres cuartos del tamaño inicial. El formato FLAC se suele usar para la venta de música por internet, y como alternativa al mp3 para compartirla cuando se desea reducir el tamaño que tendría un archivo wav-PCM sin perder calidad, ya que con este tipo de compresión se pueden reconstruir perfectamente los datos originales del archivo. También se suele usar para realizar copias de seguridad de CD de audio; admite tasas de bits variables entre 1 y 65 kHz según la complejidad de la transmisión de audio y por tanto tiene un rendimiento semejante a MPEG-2.

- ❖ AAC –*Advanced Audio Coding*–. Formato de audio digital estándar desarrollado por el Instituto Fraunhofer, Sony, AT&T y los laboratorios Dolby. Es una extensión de MPEG-2 comprimido con pérdida bajo el estándar internacional *ISO/IEC 13818-7*. Ofrece bastante más calidad que MP3 para una misma tasa de muestreo y tamaño de archivo aunque no ha podido sustituirlo en cuanto a aceptación. Su método de codificación adapta automáticamente la tasa de bits necesarios en un momento dado en función de la complejidad de la transmisión de audio en ese momento. AAC soporta tasas de muestreo hasta 96 kHz, que es calidad “estudio de grabación” y hasta 48 canales distintos, lo que lo hace indicado para sonido envolvente –*surround*– como el 5.1, 6.1, etc., y sonidos polifónicos. Es compatible y estándar para una amplia variedad de dispositivos: iPod, iTunes, Winamp, Walkman, Ahead Nero, MP4, Nintendo DSi, etcétera.
- ❖ mp3, MPEG-1 Audio Layer-III o MPEG-1 Capa 3. Por el lado de los formatos para distribución, de menor calidad pero con mucho menor consumo de espacio, el estándar *de facto* indiscutible del mercado es el formato denominado MPEG-1 Audio Layer-III o MPEG-1 Capa 3, más conocido como mp3. Fue creado por Thomson Multimedia y el Instituto Fraunhofer de acuerdo con la norma establecida por el Grupo de Expertos de Imágenes en Movimiento –*Moving Picture Experts Group* o MPEG– para ser el estándar de audio digital en los archivos de video MPEG-1. Soporta tasas de muestreo entre 16 Hz y 48 kHz y dos canales –estéreo–. Es ampliamente utilizado a nivel mundial para inter-

cambio de audio por la red ya que puede obtener factores de compresión –con pérdida– de hasta doce veces con respecto a aquellos formatos para preservación con sólo una pérdida de aproximadamente 25% en la calidad; ello se basa en el principio de eliminar las frecuencias que no son audibles por el oído humano. Este formato tiene varias calidades de salida según el muestreo utilizado, pero en promedio puede consumir alrededor de un Megabyte por minuto de grabación con una calidad razonable; esto lo hace muy aceptado con fines de distribución de información tipo audio, sobre todo en la red. Su sucesor, quien también maneja video, parece ser alguno de entre el mp3 Pro, MPEG-1 capa 4 o MP4, MPEG AAC o el MPEG-2.

- ❖ MIDI o MID –*Musical Instrument Digital Interface*–. Este formato en sí es un protocolo o norma para transferir información entre dispositivos musicales como sintetizadores, guitarras eléctricas, violines electrónicos y *software* diverso. Un archivo MIDI indica tonos, pulsos, variaciones de intensidad, sostenidos y duraciones, que pueden modificarse en cada instrumento. Es muy útil para los artistas que desean generar sus producciones combinando varios dispositivos con el apoyo de computadoras. Es muy conocido también por ser el que se utilizaba para la música sencilla de juegos de computadoras, tonos de celulares, etc. En la práctica no es de alta calidad y tiene muchas limitaciones, como la de reproducir voces humanas.

Estos no son los únicos formatos para audio digital, pero son los más utilizados en el mundo de las redes y la información. Una lista exhaustiva puede estudiarse en el sitio web de la Audio Engineering Society.³⁷ Finalmente, a manera de resumen presento una tabla de lo enunciado anteriormente:

Formato	Extensión	Ventajas	Desventajas
AIFF 16 bits(Apple)	.aif, aiff	Formato abierto - calidad	Archivos muy grandes

³⁷ Audio Engineering Society. Sitio web oficial: <http://www.aes.org>

Cómo preservar mi patrimonio digital personal

Formato	Extensión	Ventajas	Desventajas
AIFF 8 bits (Apple)	.aif, aiff	Formato abierto - calidad	Archivos grandes
Wave Form 16 bits	.wav	Formato abierto - calidad - Windows y Apple	Archivos muy grandes
Wave Form 8 bits	.wav	Formato abierto - calidad - Windows y Apple	Archivos grandes
U-law	.au	Formato abierto - calidad escalable - Windows - Unix - Apple	Archivos grandes a muy grandes
CD-audio (CDDA)	.cda	Alta calidad	Sólo para reproductores domésticos
MPEG AudioLayerIII	.mp3	Calidad escalable - poco espacio - abierto	Puede degradarse fácilmente a poca calidad
Advanced Audio Coding	.aac	Calidad escalable - poco espacio - abierto - mejor rendimiento que MP3	Poca aceptación
Windows Media Audio	.wma	Calidad escalable	Propietario (Microsoft)
QuickTime 16 bits	.qt	Formato abierto - calidad- Windows - Apple	Archivos muy grandes
QuickTime 8 bits	.qt	Formato abierto - calidad- Windows - Apple	Archivos grandes
Real Audio	.ra, .ram	Audio en red en tiempo real	Propietario - Real Networks
Wave Form 16 Bits	.wav	Formato abierto - Windows y Apple	Archivos muy grandes
Wave Form 8 Bits	.wav	Formato abierto - Windows y Apple	Archivos muy grandes
MIDI	.mid	Espacio muy reducido	Calidad muy baja

Formatos de video

En esencia, el video es una representación de imágenes consecutivas más una representación de audio. Partimos del hecho básico del cine en el cual una serie de 24 imágenes por segundo se vuelve una imagen continua para el ojo humano; esto significa que para un segundo de video necesitamos representar al menos 24 imágenes más el audio correspondiente. Ya hemos visto que la representación de una imagen fija en donde se desea una buena calidad requiere de muchos bits para lograrla. Si multiplicamos la cantidad de bits necesarios para representar un segundo de video por el número de segundos a representar podemos percibir fácilmente que se requieren cantidades realmente enormes de números para esta representación digital. Pensemos en una película típica de 90 minutos; esto es, 5400 segundos: el total es de $24 \times 5400 = 129,600$ imágenes fijas más la o las correspondientes pistas de audio. Al igual que con los tipos anteriores de documentos, se han desarrollado formatos especiales para este tipo de materiales, pero las técnicas de compresión se vuelven especialmente críticas, en particular en video de alta definición donde obviamente la calidad de las imágenes debe ser muy alta además de sonido calidad 5.1.

En general, los formatos de video son semejantes a los de audio en cuanto contienen un cierto tipo de códec; esto es, el algoritmo de compresión usado para codificar/decodificar los datos de la película digital dentro del archivo; el mismo formato puede contener opcionalmente diversos códecs. El códec a usar variará de acuerdo con la naturaleza del contexto del video y sus dispositivos asociados: hay técnicas y formatos especiales para la transmisión y recepción de televisión; para las cámaras digitales, para los celulares y otros móviles, para la distribución de películas en DVD y para el almacenamiento, distribución y preservación entre computadores vía la red. Por supuesto algunas de estas técnicas y sus consecuentes formatos tienen características que los hacen atractivos para la preservación de documentos digitales a largo plazo y son los que más nos interesarán en este momento.

Cómo preservar mi patrimonio digital personal

Existen diversas técnicas para lograr la representación digital con una adecuada compresión: las técnicas estadísticas de compresión de video se basan en los principios que ya hemos estudiado de compresión de puntos repetidos en una imagen, como por ejemplo la técnica RLE *–Run Length Encoding–*. Las técnicas denominadas redundancia espacial observan la similitud entre un pixel y sus vecinos adyacentes; si son muy parecidos a la percepción que pueda tener el ojo humano, se reducen a una información promedio. Por ejemplo, un punto en cierto verde claro y otro punto en verde un poco más claro serán reducidos a un sólo punto verde claro estándar; esta es una técnica con cierta pérdida. Las técnicas de redundancia temporal operan sobre un pixel o pixeles que se repiten constantemente en una secuencia de imágenes, fotograma a fotograma, reduciendo esa información al mínimo. Por ejemplo, en una escena donde todo es estático y lo único que se mueve son los labios del actor. Esta también es una técnica con pérdida, pero ambas logran muy buenos niveles de compresión.

Existen otras técnicas para reducir el volumen del archivo que no tienen que ver con la compresión, sino con el despliegue: por ejemplo, reducir el número de fotogramas de 24 por segundo a 20, 18 o hasta 16. Así se reduce el volumen del archivo pero por supuesto el ojo podrá percibir “brincos” o discontinuidades en el movimiento. También se utiliza la técnica de usar imágenes de muy poca resolución pero desplegando “ventanas” de video muy pequeñas sobre la pantalla para evitar que el ojo humano alcance a distinguir los pixeles. Obviamente estas técnicas son usadas solamente para distribución; no son usadas para preservación documental dada su poca calidad.

He aquí una lista de los principales formatos de video digital usados en la actualidad:

- ❖ AVI o AVI2 *–Audio Video Interleaved–*. El formato AVI fue desarrollado por Microsoft; tiene una estructura simple, almacenando la información por capas intercaladas, alternando una de video con una de audio; por tanto, puede almacenar simultáneamente un flujo de datos de video y varios flujos de audio y como consecuencia cada sección del archivo contiene en sí misma la infor-

mación necesaria para ese segmento del video. La estructura de los datos depende del códec utilizado, como por ejemplo AC3, DivX, Xvid, etc. El reproductor para visualizarlo viene de forma nativa en el sistema operativo Windows y existen versiones gratuitas para las principales plataformas. Permite la creación de video con toda una gama de calidades, desde las más básicas hasta calidad profesional, con o sin pérdidas.

- ❖ WMV o ASF –*Microsoft Windows Media Video*– Este formato es una de las últimas propuestas de Microsoft para Windows Media Player, el programa reproductor para video y sonido de esta empresa. Viene integrado al sistema operativo en las últimas versiones de Windows. Aunque existen versiones para otras plataformas, tienen todavía muchas limitaciones. Permite la creación de video con toda una gama de calidades, desde las más básicas hasta calidad profesional, con o sin pérdidas.
- ❖ RV, RAM, RM, RMVB –*Real Video*–. Desarrollado por la empresa Real Networks, ha sido y es uno de los más usados para *streaming* o transmisión “en tiempo real” vía la red. Requieren de su propio reproductor, el Real One del cual existe una versión sencilla gratuita, sólo para ver el video, y una profesional, que permite crear y editar videos en dicho formato. Usa una excelente técnica de compresión sin pérdida.
- ❖ MOV –*QuickTime Movie*–. Formato multiplataforma desarrollado por Apple. Permite la creación de video con toda una gama de calidades, desde las más básicas hasta calidad profesional, incluyendo el video 3D. Como todos estos formatos, requiere de un reproductor especial, el Quicktime Player para visualizarlo, del cual existe una versión sencilla gratuita, sólo para ver el video, y una profesional, que permite crear y editar videos en dicho formato, con o sin pérdidas.

Todos estos formatos se basan en alguna de las versiones del estándar denominado MPEG –*Moving Picture Experts Group* o Grupo de Expertos en Imágenes en Movimiento–. Este es un grupo de trabajo del Instituto Mundial de Estándares ISO creado en 1988 con el

Cómo preservar mi patrimonio digital personal

propósito de crear estándares mundiales para archivos de audio y sonido digitales. Las variantes más importantes son:

- ❖ MPEG-1. Primer estándar de esta serie para compresión de datos de video y sus dos canales de audio asociados para el sonido estéreo. Permite almacenar video a una velocidad de 1.5 Megabits por segundo o Mbps con calidad aproximada al de una cinta VHS en un soporte de CD denominado VCD –Video Compact Disc–. Su estándar para el sonido dio lugar al popular formato mp3.
- ❖ MPEG-2. Estándar diseñado para la televisión digital –HDTV, televisión de alta definición– por cable y satelital; usado también para videos en DVD. Ofrece alta calidad a una velocidad de hasta 40 Mbps y 5 canales de sonido envolvente 5.1. Este formato absorbió a MPEG-3.
- ❖ MPEG-4. Estándar diseñado para permitir la codificación de datos multimedia en forma de objetos digitales con el fin de lograr una mejor interactividad; esto lo hace especialmente adecuado para la web y para los dispositivos móviles –iPods, iPads, tabletas, celulares, etc.–. Va sustituyendo gradualmente a MPEG-2. Su variante MPEG4-H.264 AVC también conocida como MP4-H.264 AVC se considera un formato ideal para preservación de archivos de video.

Es obvio que en este aspecto de almacenamiento de imágenes en movimiento sucede lo mismo que en los casos de imágenes fijas y sonido: existe un principio de resolución que nos dice que a mayor calidad de imágenes o sonido, mayor espacio de almacenamiento así como mayor tiempo de transmisión en red requeridos por el archivo resultante y viceversa. Este principio se mantiene en este tipo de imágenes pero dados los enormes volúmenes asociados a este tipo de documentos, la resolución y los mecanismos inherentes de compresión se vuelven particularmente críticos.

Como puede verse del estudio de los formatos documentales en sus diversas variantes, cada formato tiene características de propiedad o apertura, mayor o menor compresión, mayor o menor calidad, menor o mayor cercanía a estándares *de jure* o *de facto*,

etc. Cada formato ha sido creado con un propósito y por lo mismo tiene ciertas ventajas y ciertas desventajas. No existe el formato universal que maximice todas las ventajas minimizando todas las desventajas. El conocimiento de unas y otras y de sus contextos es lo que nos permitirá ir tomando mejores decisiones acerca de los formatos que se deben usar para el trabajo cotidiano y aquellos que vamos a usar para preservación a largo plazo de nuestros materiales documentales.

OCR. *Optical Character Recognition* o Reconocimiento Óptico de Caracteres

En esencia, esta es una técnica o procedimiento de cambio de formato. Nos sirve para transformar caracteres tipo imagen a caracteres tipo texto, lo cual se logra por lo general a través de un programa o aplicación al efecto.

Analicemos la diferencia entre un carácter tipo texto y un carácter tipo imagen. Cuando nosotros creamos un texto a partir de un editor con un programa ofimático –digamos por ejemplo Word– tecleando un cierto texto, obtenemos uno que tiene como características básicas que es editable y buscable. Editable significa que nuestro texto, al estar visible, puede ser modificado, agregándole nuevas palabras, borrándolas, sustituyéndolas, cambiando los párrafos de lugar, su distribución física, el tipo o tamaño de letra, etc. Un ejemplo típico de esto es un texto en formato doc o txt. Nosotros podemos aplicar todas estas funciones en ellos gracias a nuestro editor de texto ofimático, modificándolo una y mil veces hasta que esté a nuestra entera satisfacción. Por otro lado, muchos programas, incluyendo nuestro editor ofimático, pueden buscar un cierto texto en el interior de estos archivos para hallar cierta palabra o frase deseada, cuantas veces lo necesitemos.

Esto es posible ya que la computadora en realidad busca en su interior una secuencia de caracteres que nosotros estipulamos –llamada cadena de caracteres o simplemente cadena–; la máquina hace esto no buscando letras, sino de acuerdo con los valores ASCII de cada letra del texto, que es lo que realmente tiene grabado en su

Cómo preservar mi patrimonio digital personal

interior. Por ejemplo, cuando nosotros deseamos buscar la palabra Tiempo la computadora sabe que ese texto está formado por los bytes consecutivos “T” = 01010100, “i” = 01101001, “e”= 01100101, “m” = 01101101, “p” = 01110000, “o” = 01101111. Véanse los valores de la tabla ASCII en el Anexo 4 para comprobar estos datos. Esto es, la computadora buscará dentro del texto una cadena o secuencia de bytes exactamente igual a 010101000110100101100101011011011011000001101111, que es la secuencia binaria de esos caracteres. La computadora puede hacer esto a velocidades vertiginosas, encontrando esta cadena de caracteres rápidamente cuantas veces sea necesario dentro del texto. Por consecuencia todo texto cuya base sea el ASCII: doc, odt, txt, html, etc., será editable y buscable.



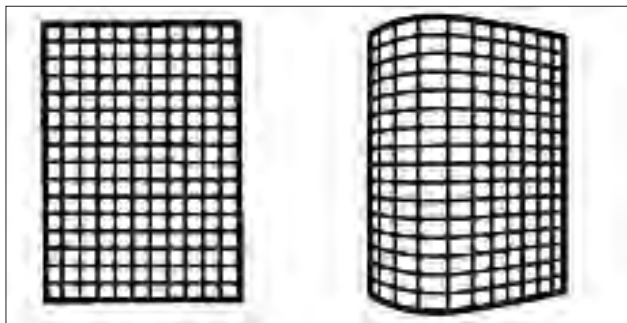
Cuando nosotros escaneamos una página de texto –esto es, que tenga impresos caracteres– y la guardamos en un formato de imagen, por ejemplo jpg, en realidad no tenemos un texto; tenemos una secuencia de puntos que la computadora ha guardado exactamente como eso, un conjunto de puntos consecutivos y que en realidad simulan o aparentan ser un texto. Si nosotros desplegamos ese archivo en una pantalla o lo imprimimos, los puntos de la matriz de puntos vuelven a quedar alineados y nosotros podemos ver los caracteres, formados por esa secuencia de puntos; para nosotros será legible, pero si queremos que la computadora busque o edite esa secuencia de caracteres, no podrá hacerlo, ya que no puede distinguir, dentro de miles y miles de puntos en una secuencia lineal, cuáles forman un cierto carácter. Ese texto dentro de ese archivo no es buscable ni editable.

Trate de imaginar el lector esta imagen, no como una rejilla, sino como una única línea de puntos consecutivos formada por cada renglón uno a continuación del otro, que es como en realidad se encuentra guardada en el archivo dentro de la computadora. Podemos darnos cuenta de que en este caso, su forma no sería reconocible como una letra. Al analizar esta imagen en forma de matriz de puntos con renglones y columnas ordenados puede observarse una imagen ya legible para nosotros como una letra, pero sigue sin ser legible para la computadora, al menos, no sin ayuda; ésta consiste en un programa OCR. La función de todo programa OCR precisamente es la de tratar de identificar a una letra “dibujada” con pixeles por medio de cada uno de ellos según la forma en que se acomodan para formar la imagen dentro del archivo, intentando darle forma y reconocer así una letra dada, e intercambiarla por su byte equivalente en ASCII, como si hubiera sido tecleada. Esto puede parecer sencillo de inicio –y lo es para los humanos– pero para las computadoras es mucho más complicado de lo que al principio se cree.

Existen innumerables configuraciones o patrones de puntos posibles para formar una letra. Consideremos de entrada que existen muchos, muchísimos tipos distintos de ellas, las denominadas fuentes: Courier, Arial, Thorndale, Times New Roman, Gótica, etc. Docenas y docenas de tipos distintos de fuentes, en mayúsculas y minúsculas, y variando en tamaño, los denominados puntos de altura de la letra, desde los muy pequeños hasta los realmente grandes. Reflexionemos en el hecho de que algunas letras tienen “patines”, las pequeñas barritas horizontales que van en la base de la letra y que sirven para alinear el texto: “f”, “h”, “m” y algunas otras fuentes no los tienen: “f” “h”, “m”. Algunas fuentes tienen variantes significativas de las letras: “a” y “ɑ”, “e” y “e”, “g” y “g”, etc. Otras tienen “curvas” inusuales y van abajo de la línea base: “f” “h” “m”. Existen además las variantes de negrita y cursiva que cambian el grosor o inclinación de las letras y por tanto su patrón de puntos dentro de la matriz. El buen o mal contraste que exista entre el papel y las letras impresas incide favorable o negativamente en el proceso de identificación de las letras; tintas muy desvanecidas tende-

Cómo preservar mi patrimonio digital personal

rán a crear confusión. Papeles muy amarillentos o sucios con pobre contraste afectarán el reconocimiento de las letras; lo mismo sucederá con papeles con manchas, ya sea de humedad, óxido, hongos u otra clase. Papeles delgados donde se transparenta el texto que existe al reverso de la página, produciendo sombras en el frente, también tienden a confundir al OCR.



Hoja plana

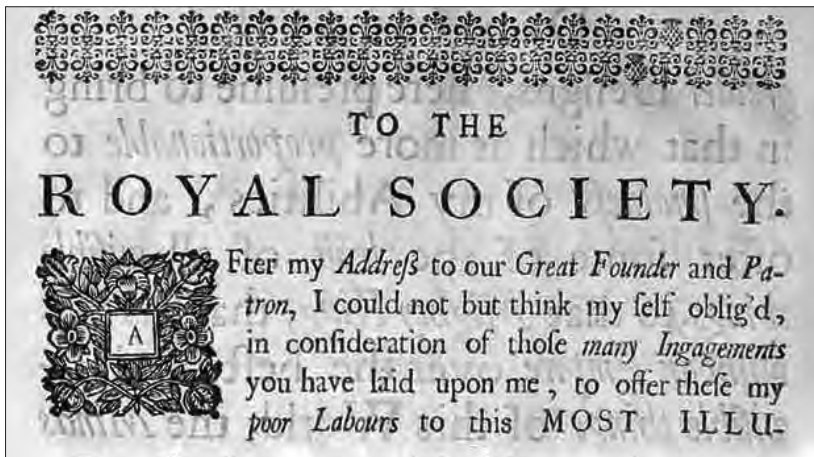
Hoja curvada

Las letras que se encuentran cerca del lomo o zona de encuadernación del libro y que ya no se encuentran totalmente planas respecto a la cama del escáner se curvan haciendo más difícil su reconocimiento. Letras acentuadas donde los acentos se empastan con el carácter sin espacio intermedio entre ellos causan también gran confusión al programa. Obviamente diagramas, plecas, letras capitulares, etc., no podrán ser identificadas por el programa OCR.



Vocales con acento "empastado"

El funcionamiento básico de este tipo de programas consiste en tratar de identificar el patrón de puntos que observa en cada letra, compararlo contra patrones ya preestablecidos que contiene de fábrica y elaborar, letra por letra, un texto ASCII que propone al usuario para que este pueda revisarlo y, en su caso, corregir los errores tecleando las letras correctas en donde exista una propuesta errónea de carácter. Como el texto resultante es ASCII puede ser guardado como se hace normalmente con el editor de textos, en formato doc, txt, html, etc. Ese archivo nuevo con ese texto es por tanto ya editable y apto para búsquedas.



Tipografía antigua y reverso de la página que se transparenta.

Cuando escaneamos textos de libros antiguos, muchas veces vienen impresos en tipografías que estuvieron de moda en ciertas épocas, pero que hoy no se usan más, y el programa OCR no las tiene dentro de sus patrones estándares de letras. Esto hace que se le dificulte mucho acertar al carácter correcto cuando compara los puntos con sus patrones preestablecidos. En estos casos, los programas OCR de cierta calidad tienen rutinas que le permiten “aprender” estos nuevos patrones con ayuda del usuario, el cual

Cómo preservar mi patrimonio digital personal

le va indicando al programa a qué carácter corresponde una cierta secuencia de puntos no identificada por él.

Dependiendo entonces de todas esas características descritas de fuentes y tamaños de las letras, época de las tipografías, contraste de letras contra papel, y todo lo demás mencionado en el párrafo anterior, podemos hacer procesos de reconocimientos óptico de caracteres muy exitosos, con cerca del 100% de éxito desde la primera vez y donde las correcciones a teclear son mínimas. Esto sucede por ejemplo cuando trabajamos con textos que provienen de archivos con textos en imagen que fueron creados digitales de origen, y en donde todas las variables mencionadas están a nuestro favor. Por el contrario, cuando tratamos de aplicar este proceso a textos antiguos, sucios, de poco contraste, curvados, etc., en donde muchas de estas variables estén en nuestra contra, el proceso produce tal cantidad de caracteres equivocados que muchas veces resulta mejor y más rápido volver a teclear el texto que tratar de arreglar los errores introducidos por el programa OCR. Esto deberá ponderarse con pruebas previas siempre que se utilicen estos programas.

Con fines de uso y sobre todo de preservación, es importante que reflexionemos cuál es la necesidad y la utilidad de poseer un texto en formato editable y cuando no es necesario. Por ejemplo, sería un error grave crear un documento extenso, como una tesis, en formato editable doc, para que al final la convirtiéramos a PDF, más presentable pero no editable, y borráramos el archivo doc pensando en que como ya tenemos una mejor versión de ese documento en PDF no necesitaremos más el anterior. En efecto, la versión PDF tendrá mejor acabado y presentación, pero este formato no es editable y habremos destruido la capacidad de poder hacer esto en un futuro si no preservamos también la versión doc. Cuando usamos programas que buscan ciertos textos dentro de un archivo PDF y los encuentran, en realidad estamos viendo un archivo PDF que se despliega ante nuestra vista, pero no vemos una versión txt que está oculta detrás de éste y donde en realidad el programa busca y encuentra ese texto. Hay obviamente un proceso apuntador que indica al programa cuál página del documento PDF mostrar al usuario con el texto encontrado en el txt.

Dependiendo del contexto de nuestro trabajo y del documento, a veces necesitaremos sólo la imagen de ese texto sin ninguna versión editable; otras ocasiones requeriremos sólo de la versión texto, por lo que podemos prescindir de la versión imagen una vez convertida a texto, y otras veces necesitaremos guardar ambas versiones, ya que trabajan juntas dentro de nuestro propósito. Por lo mismo es importante conocer esta diferencia entre texto-imagen y texto-ASCII editable para tomar las decisiones correctas acerca de cuál o cuáles tipos de archivos de texto-imagen debemos crear y preservar. Debe recordarse también que un documento guardado en forma de imagen consume mucho más espacio que el mismo documento en ASCII.

ATRIBUTOS DE LOS DOCUMENTOS DIGITALES

Hemos mencionado ya que existe el factor tecnológico como elemento que incide en la preservación documental; de hecho, este es el factor más conocido de todos aunque como ya mencionamos, no es el único. Lo que más tiene que ver con preservación documental en lo relativo al factor tecnológico es el cambio rápido y constante de los dispositivos y aspectos tecnológicos relacionados con los documentos digitales.

La razón de ello es que un documento digital requiere de varios elementos tecnológicos para poder ser percibido y por tanto utilizado. Todo documento digital está conformado por dos partes básicas: contenido y estructura. El contenido es la información del documento en sí: texto, imagen, música, etc. La estructura es todo el entorno que esa información requiere para poder ser utilizada: soporte o medio, equipos requeridos, sistema operativo, programas o aplicaciones, formatos, etc. Dentro del concepto de obsolescencia tecnológica parte del problema consiste en conservar los bits del contenido sobre un soporte y otra parte en cómo hacer que sean legibles en un futuro, dados esos constantes e incesantes cambios tecnológicos que se dan sobre la estructura y que son ajenos a los documentos digitales en sí, pero indispensa-

Cómo preservar mi patrimonio digital personal

bles para hacerlos legibles. Por lo mismo este factor tecnológico es un ente complejo que está estrechamente ligado a su vez a una serie de elementos tecnológicos más simples: equipos y dispositivos que se requieren para reproducir esos documentos; soportes o medios; sistemas operativos, programas, formatos que se han creado para la representación de diversos documentos digitales en todos sus tipos: texto, audio, imagen estática y en movimiento, etc., y que surgen y desaparecen con rapidez cada vez mayor y que ya hemos analizado. Dentro del factor tecnológico se encuentra además el elemento de seguridad de la información: ¿cómo se evitará que se pierdan por accidente, error o mala intención documentos que se desea preservar? ¿Cómo se garantiza su integridad para evitar que yo mismo o terceras personas sustraigan, destruyan o alteren esos documentos? Estas preguntas serán analizadas más adelante.

Continuando con el factor tecnológico, pasemos a la exposición y análisis de los conceptos fundamentales para entenderlos. Es necesario conocer ciertas características que poseen intrínsecamente todos los documentos digitales las cuales son llamadas atributos. Cada documento puede poseer cada uno de esos atributos en mayor o menor grado. Su relevancia radica en que para cada persona y cada tipo de documento, el valor e importancia de cada atributo variarán. Cada uno de nosotros debe establecer en qué grado deseamos que cada atributo exista en nuestros diversos tipos de documentos. Dependiendo de cómo se crean, cómo se manejan, cómo se conservan esos documentos esos atributos existirán en mayor o menor grado. De ahí la importancia de conocerlos y comprenderlos para que así podamos establecer y lograr los atributos que deseamos en cada tipo de documentos digitales que poseemos. En forma breve, esos atributos se detallan a continuación.

Permanencia

Se le llama así al hecho de que un documento digital siga existiendo a lo largo del tiempo, esté disponible por un lapso considerable. Este concepto está asociado con su presencia, estabilidad

y su seguridad. Depende primordialmente de que las cadenas de bits que representan la información –su contenido– se mantengan ininterrumpidamente estables e inmutables sobre su soporte físico, el cual por supuesto debe durar en buen estado en forma continua, es decir, permanece.

Accesibilidad

La disponibilidad y usabilidad de la información, en el sentido de la capacidad o facilidad futura de que habiendo tenido permanencia, esa información pueda ser consultada, reproducida y por tanto usada nuevamente. Esto quiere decir: su estructura.

Pensemos que tenemos en nuestras manos un disco fonográfico de acetato de 78 rpm, una microficha de 4 x 6 pulgadas, un casete con una película en formato VHS, una película de cine en 8 mm, un archivo de hoja de cálculo en Supercalc en un disquete de 8 pulgadas; un archivo de texto en Word 5.1 en una cinta magnética de computadora de media pulgada. Asumamos que todos estos documentos y sus soportes se encuentran en perfecto estado de conservación. Tenemos por tanto un conjunto de documentos sobre su soporte que han permanecido a lo largo del tiempo: existen sin duda, sus cadenas de bits están intactas y en perfecto estado. Podemos afirmar por tanto que esos documentos han tenido permanencia.

Pero existiendo, habiendo permanecido en buen estado, debemos ponderar nuestra capacidad de “abrirlos” y poder analizar su contenido. Para cada uno de ellos requerimos un dispositivo tecnológico del cual dependemos para lograr esa acción. En el caso del disco fonográfico, la microficha o la película lo único que necesitamos es físicamente un equipo, un reproductor del disco o fonógrafo, un lector de microfichas y un proyector de cine de 8 mm, respectivamente; contando con el equipo el problema del acceso está resuelto. En el caso de la cinta de video necesitamos dos equipos: una videocasetera VHS y un monitor tipo televisión; un poco más complejo pero siguió siendo sólo un problema de equipo. Estamos observando un problema tecnológico, pero sólo es

Cómo preservar mi patrimonio digital personal

de equipo. Resolviendo el acceso al equipo, y asumiendo que el medio está en buen estado, podremos revisar inmediatamente el documento.

En el caso del archivo Supercalc –una versión antigua de la actual hoja de cálculo– necesitamos obviamente el equipo: una computadora, que además incluya un lector de disquetes de 8 pulgadas, pero además requerimos del programa Supercalc para que pueda abrir ese archivo. Por si eso no fuera suficiente requerimos del sistema operativo para el equipo bajo el cual operaba Supercalc, esto es CP/M.³⁸ Es decir, aquí tenemos múltiples requerimientos de equipo y de programas. Lo mismo sucede con el archivo en Word almacenado en una cinta magnética: necesitamos la computadora, además de un dispositivo lector de ese tipo de cinta así como el sistema operativo y el programa que puede abrir e interpretar el archivo. Aquí hay que considerar que el sistema operativo en el cual ese procesador de texto operaba era MS-DOS, no Windows, así que requerimos de una computadora con un sistema operativo compatible al de la aplicación o al menos un “emulador” de aquel sistema operativo. Nuestra capacidad de acceder a estos documentos ya no está sólo en función de contar con todos esos dispositivos, sino además necesitamos contar con sistemas operativos y programas.

Como puede observarse de estos ejemplos, puede o no haber accesibilidad, independientemente de que exista permanencia. Dicho de otra forma, el que los contenidos de los documentos tengan permanencia, esto es, sus cadenas de bits estén en buen estado sobre sus soportes, no garantiza que tengamos accesibilidad hacia ellos: son dos conceptos distintos. Ambos son necesarios para poder disponer de un documento preservado. Uno sin el otro no

38 CP/M. Control Program for Monitor o Control Program for Microcomputers: sistema operativo desarrollado por Gary Kildall para los primeros microprocesadores 8080 y 8085 de Intel y Z80 de Zilog, de los primeros utilizados para la construcción de computadores personales a fines de la década de 1970. Se distribuía en disquetes de ocho pulgadas. Fue el sistema operativo más usado en las computadoras personales de esa década y su éxito se debió a que era fácilmente portable.

tiene ninguna utilidad. Y como vimos, la accesibilidad puede tener requerimientos complejos en mayor o menor grado en cuanto a programas o equipo; la estructura.

La permanencia de un documento tiene que ver en mayor parte con la duración del soporte y su superficie o medio de grabación, y la accesibilidad tiene que ver en mayor medida con la estructura de su contenido. Por lo mismo, el problema de permanencia es más fácil de resolver que el de la accesibilidad.

Calidad y valor de uso

Los conceptos de calidad y valor de uso de los documentos están estrechamente relacionados. Partamos de una premisa teórica: el principio de la calidad es directamente proporcional al grado en que el documento digital emula al documento representado. Ese documento representado puede ser un documento original en soporte tradicional del cual obtenemos una copia digitalizada o un original digital. El principal elemento de la calidad es su apariencia con respecto al documento original; es decir, qué tanto se parece el documento digital a lo que pretende representar. Esta apariencia a su vez está definida por elementos tales como la resolución o grado de minuciosidad de los elementos que conforman el documento, tono, color e inclusive textura; su escala –de preferencia uno a uno–, su secuencia original, su integridad, etcétera.

Pero la apariencia debe tener una relación práctica con la realidad del usuario y del documento; no podemos decir inflexiblemente que mayor resolución siempre dará mayor calidad a un cierto documento; si así fuera, siempre estaríamos escaneando a 1200 o 2400 dpi y usando una paleta de 256 millones de colores. Esto rara vez es necesario, ya que es indispensable contextualizar la calidad desde un punto de vista práctico. ¿Cuánta calidad es necesaria en un documento?

La respuesta está en establecer un valor de uso para ese documento en particular. Cuando necesitamos obtener información para realizar alguna tarea o trabajo y la encontramos en una biblioteca –digamos en una revista– obtenemos ahí una fotocopia del artículo

Cómo preservar mi patrimonio digital personal

de nuestro interés. Reflexionemos en el hecho de que esa copia no es ni con mucho de la misma calidad del original, pero si es suficientemente nítida, nos servirá perfectamente para nuestros propósitos de obtener información. Sólo en muy contados casos requeriremos una fotocopia a color o algo parecido. Si la fotocopidora tiene una calidad razonable no hay problema con la copia en blanco y negro; tiene la calidad suficiente para nuestro propósito: obtener información. Aun con pequeñas fallas como nitidez baja o rayas en la hoja, si la copia es legible, nos sigue siendo útil para nuestro propósito. Si la máquina se descompone y entrega copias ilegibles, ese material es inútil. Por tanto, hay una calidad disminuida del documento donde nos sigue siendo útil a pesar de no ser tan bueno como el original, y hay una calidad donde deja de ser útil.

Por lo mismo y al igual que con esa fotocopia, en el ámbito digital es muy importante que cada persona defina cuál es la calidad suficiente que un cierto material digitalizado debe tener para su propósito y adecuado uso, reduciéndose a la vez el espacio requerido y con ello, los costos.

Es necesario entender este concepto del valor de uso perfectamente: debemos estar conscientes de que un documento digital no es idéntico a un documento fijado en un soporte tradicional: puede ser muy parecido, puede ser muy semejante a su antecesor tradicional, pero no es idéntico. Lo mismo sucede si a un documento originalmente digital le reducimos los parámetros de resolución. Podemos hacer que la semejanza sea poca o podemos hacerlo sumamente semejante al original; depende de técnica, esfuerzo y trabajo, y es una decisión de la persona que lo crea. ¿Qué tanto conviene hacerlo muy semejante al original? El hacerlo así conlleva una mayor inversión de recursos, generalmente espacio, tiempo y dinero. Hacerlo poco semejante al original es más rápido y económico, pero obviamente ello va en detrimento de la calidad del documento. Para establecer la decisión de calidad suficiente es necesario siempre referirse al valor de uso del documento; esto es a los motivos, necesidades y expectativas del usuario al momento de usar en un futuro un cierto documento. ¿Qué tan útil me será como usuario el documento digital resultante? ¿Qué tanto sustituye al original?

Cuando en nuestra respuesta a estas preguntas afirmamos que un documento es más satisfactorio y útil decimos que tiene más calidad de uso y viceversa. Obviamente la definición de la calidad que podemos hacer es muy amplia y depende en gran medida de nuestras expectativas como usuarios de ese documento. Habrá usuarios y ocasiones donde las expectativas de calidad sean mínimas: se requiere simplemente que un documento de texto sea legible o que un archivo de audio sea inteligible o que un archivo de video sea visible; será suficiente entonces con documentos de calidad mínima, que nos habrán resultado muy económicos. Pero habrá también usuarios y ocasiones que deseen textos o imágenes de una alta definición, con estricta paleta de colores; audio o video de la mejor calidad, hipervínculos, etc. Requeriremos entonces de documentos a la altura de ese valor de uso esperado.

Si digitalizamos un artículo de periódico para distribuirlo en la web, de nada sirve una copia donde la resolución sea tan baja que las letras pequeñas nos resulten ilegibles –72 dpi–, como tampoco sirve digitalizarlo a 600 dpi: tal finura de resolución es irrelevante en un periódico, no mejora la imagen y sólo estaremos desperdiciando espacio inútilmente en el almacenamiento y transmisión de ese archivo. En este ejemplo vemos dos documentos con distinta calidad; la primera está por debajo de lo aceptable y la segunda por encima de lo necesario. El valor de uso nos diría que la calidad ideal estará alrededor de los 300 dpi para ese tipo de documento –periódico– y para ese tipo de uso –distribución en la web–.

Como un segundo ejemplo, pensemos en una imagen digitalizada de baja resolución –digamos 72 dpi– la cual pensamos imprimir ampliándola para usarla como fondo para un cartel. La imagen será pobre y serán muy notorios a simple vista cada uno de sus puntos –a esto se le llama una imagen pixelada–; por tanto, esta imagen será inútil para nuestro propósito, no tiene la calidad suficiente. Si esa misma imagen simplemente la queremos para usarla como una pequeña viñeta ornamental en un sitio web, esa imagen de 72 dpi sirve perfectamente para nuestro propósito: tiene la calidad suficiente.

El valor de uso nos define entonces si una cierta calidad está por debajo de lo esperado, a nivel o sobreespecificada. El documento

Cómo preservar mi patrimonio digital personal

tiene la calidad suficiente cuando representa adecuadamente a un cierto contenido para un propósito específico, independientemente de que el documento sea igual o no a un original, ya que representa adecuadamente lo que contiene y necesitamos para ese propósito específico.

Entonces, la calidad suficiente de un cierto tipo de documento está definida por su valor de uso para nosotros. ¿En qué medida el documento digital en cuestión representa suficientemente bien al documento que contiene de acuerdo con un propósito? Como podemos observar, la respuesta a cuánta calidad es necesaria en un cierto tipo de documento está estrechamente ligada con el propósito y uso de ese tipo de documento. Entonces, una vez que establecemos el valor de uso de un documento de acuerdo a nuestros propósitos y necesidades podemos entonces establecer cuál es la calidad mínima suficiente o aceptable que ese documento debe tener.

Cuando a un documento que tiene calidad suficiente le agregamos identificadores del documento –sus metadatos– tenemos además lo que se conoce en el medio como un documento funcional. Un documento funcional es por tanto un documento digital que cumple con nuestras expectativas de valor de uso ya que tiene calidad suficiente y tiene además identificadores del documento o metadatos. Ampliaremos esta idea más adelante, pero conviene aquí introducir la premisa de que lo que queremos preservar siempre son documentos funcionales.

Conviene también aquí introducir la idea de que muchas veces desearíamos obtener dos copias de un mismo documento con dos calidades suficientes distintas, ya que obedecen a dos valores de uso distintos. Ello es perfectamente válido. Pensemos en una fotografía que es importante para nosotros: la de una graduación, boda, festejo, viaje, etc., que tiene una particular importancia para nosotros. Será conveniente preservarla en una alta calidad por su relevancia, pero es conveniente tener además una copia de baja resolución para compartirla, por ejemplo, en las redes sociales. Se requieren por tanto dos copias de calidades distintas que obedecen a dos valores de uso diferentes. Más adelante abundaremos en cómo manejarlas y preservarlas bajo esta premisa.

Autenticidad

La característica de autenticidad es sumamente importante. Para ciertos tipos de documentos –en especial documentos que pertenecen a archivos institucionales– es uno de los elementos más importantes de la preservación. Tiene que ver con la confianza de un documento de archivo como tal; esto es, la cualidad de un documento de archivo de ser lo que pretende ser sin alteraciones o corrupciones. Los documentos auténticos son los que han mantenido su identidad e integridad al paso del tiempo.³⁹ Este concepto es directamente proporcional al grado en que el documento digital refleja al original, no en su apariencia física, sino en su esencia, su espíritu, su intención. Un documento auténtico es el que refleja totalmente la esencia del original; es decir, no ha sido corrompido en su contexto: alterado, mutilado, interpretado, aumentado, recortado, deformado, censurado, etc.: es confiable y por tanto aceptable. Su mensaje, autoría, fechas asociadas, lugares, etc., son en realidad las consignadas en el documento desde siempre; en suma: es auténtico. Aunque hubiese cambiado físicamente, en su esencia refleja de manera completa lo que se estableció en el documento original.

Un documento no tiene que ser idéntico al documento que le dio origen para ser auténtico; de hecho, es perfectamente natural que los documentos electrónicos sean modificados de tiempo en tiempo para actualizar su formato, versión, sistema operativo, código de caracteres, etc. Es necesario por tanto garantizar de alguna forma que aunque su estructura física cambie, su contenido, su apariencia, su esencia sigan siendo los mismos. Si esto se cumple, sigue siendo auténtico e íntegro. En la medida que podamos garantizar esa continuidad de su esencia podremos afirmar que todo nuestro documento sigue siendo auténtico a lo largo del tiempo, al margen

³⁹ InterPARES Project, *The International Research on Permanent Authentic Records in Electronic Systems. Glosario de Preservación de Archivos Digitales* [en línea].

Cómo preservar mi patrimonio digital personal

de los cambios y adaptaciones tecnológicas sufridas. Se abundará en estas ideas al tratar las migraciones tecnológicas de documentos.

Confidencialidad

La característica de la confidencialidad tiene que ver con el hecho de que los registros documentales deben estar disponibles siempre, pero sólo para las personas autorizadas, durante las circunstancias y bajo condiciones válidas y preestablecidas. No deberá ser posible obtener ninguna información de los archivos fuera de esas condiciones. De ella se desprende la privacidad: el hecho de que ciertos documentos no puedan ser accedidos por ninguna persona fuera de las que tienen el acceso confidencial del documento.

Versión

Con cierta frecuencia, podemos poseer más de una copia de un documento, no necesariamente iguales una de la otra; esto es, tenemos más de una versión de ese documento. En algunos casos nos da lo mismo cualquier versión, mientras tengamos el documento, pero en otras ocasiones es necesario para nosotros tener diversas versiones de un documento y debemos poder distinguir entre una y otra para no confundirnos. Reflexionemos pues en lo que significa una versión de un documento. Para ello pensemos en una conceptualización en cuatro partes de un cierto documento: obra–expresión–manifestación–ítem.⁴⁰ Esto puede aplicarse tanto a productos literarios como a artísticos.

Obra es un contenido intelectual, literario o artístico, una creación diferenciada de otras; es un concepto general, amplio, y por tanto no necesariamente preciso.

Expresión es la realización literaria o artística de una obra en un cierto momento y lugar.

⁴⁰ International Federation of Library Associations and Institutions (IFLA), *Functional Requirements for Bibliographic Records. Final Report* [en línea].

Manifestación es la materialización física de la expresión de una obra; esto es, un proceso de “fijado” o escritura sobre un soporte físico, tradicional o digital.

Ítem es cada ejemplar individual de una manifestación.

Dicho de otra forma:

- ❖ Una obra es realizada mediante una expresión.
- ❖ Una expresión es materializada en una manifestación.
- ❖ Una manifestación es ejemplificada por un ítem.
- ❖ Ítem es por tanto un ejemplar individual de una manifestación.⁴¹

Es mucho más fácil entender estos conceptos con ejemplos; utilizaré para ello *La vuelta al mundo en 80 días*, de Julio Verne, como ejemplo de una obra literaria y el *Concierto de Brandenburgo n° 2* de Johann Sebastian Bach, BWV 1047, como un ejemplo de obra artística.⁴² Ambos son contenidos intelectuales diferenciados, es decir, distinguibles entre otros de su tipo, y dicho de esa forma: *La vuelta al mundo en 80 días* de Verne, *Concierto de Brandenburgo n° 2* de Bach, son genéricos, no precisados; simplemente las obras de esos autores, mencionadas en forma general, sin entrar en más detalle de su edición.

Si tenemos la versión original de *Le Tour du Monde en Quatre-Vingts Jours* en francés y una traducción de la misma al castellano, *La vuelta al mundo en 80 días*, tenemos dos expresiones distintas de una misma obra. *Around the World in Eighty Days* en inglés es por tanto una tercera nueva expresión de esa misma obra. De hecho, cada traducción hecha a otro idioma es una nueva expresión de la misma obra. Si existen 60 traducciones de esa obra a otros tantos idiomas tenemos por tanto 60 expresiones distintas de una misma obra. Si una persona tradujo esa obra al castellano en el siglo XIX y otra lo hizo de nuevo hace unos pocos años serán similares pero no idénticas y por tanto tenemos dos expresiones distintas

41 *Glosario para la IME ICC (IFLA Meeting of Experts on an International Cataloguing Code/Reunión IFLA de Expertos sobre un Código Internacional de Catalogación)* [en línea].

42 La clasificación y la ley de propiedad dividen las obras en literarias y artísticas.

Cómo preservar mi patrimonio digital personal

de esa obra en castellano y así sucesivamente. La obra sigue siendo sólo una a pesar de todas sus expresiones.

Cada vez que una expresión de una obra se fija en un soporte, tenemos una manifestación de esa expresión. Por lo mismo cada edición impresa de una obra es una manifestación de esa obra ya que se ha fijado en un soporte de papel. Cada uno de los ejemplares de una edición es un ítem de esa manifestación. La copia que tal vez poseemos en nuestro librero o que vemos en la librería es por tanto un ítem –o ejemplar– de una cierta manifestación o edición en particular de esa obra. La edición de “Sepan Cuántos...” de Porrúa de 1987 es una manifestación de una traducción y por tanto de una expresión; la edición de 2006 de Editores Mexicanos Unidos es otra manifestación distinta ya que corresponde a una traducción distinta a la de Porrúa. Nótese que la obra es siempre la misma. Nótese también que aunque los ítems por lo general son idénticos, puede haber alguno de ellos que sobresale y se convierte en muy especial debido a una cierta característica del mismo: ejemplar firmado por el autor, ejemplar proveniente de una primera o rara edición, ejemplar con un defecto destacable, etcétera.

Además de estas versiones impresas en papel, podemos encontrar en el sitio de libros electrónicos del proyecto Gutenberg –<http://www.gutenberg.org>– varias versiones de esta obra: en francés, en forma de texto en formatos electrónicos HTML, EPUB, Kindle, Plucker, QiOO Mobil, UTF-Texto, y además en formato de audiolibro en formato Ogg Vorbis Audio, Apple iTunes Audiobook, mp3 Audio, Speex Audio. Como podemos observar, de una cierta expresión de la obra –su versión original en francés– podemos tener varias manifestaciones, es decir varios fijados o escrituras de la expresión sobre un cierto soporte: en forma de textos digitales, como las HTML o en forma de audio, como en mp3. De su expresión en inglés igualmente han sido fijadas o grabadas y por tanto tenemos diversas manifestaciones de esta expresión sobre un cierto soporte electrónico: en forma de textos, como las HTML, o en forma de audio, como en mp3. Cuando nosotros descargamos una copia de una de esas manifestaciones, tenemos en nuestro poder un ítem o

ejemplar de una manifestación –el fijado en HTML– proveniente de una cierta expresión –la versión en inglés– de esta obra de Verne.

En el sitio de Wikisource de Wikipedia podemos encontrar además dos versiones en español –por tanto una expresión fijada en una manifestación– en formato HTML, como otra expresión con otra manifestación, en formato de audio libro en formato ogg.⁴³ Nótese que algunos de los ejemplos que hemos dado son textos impresos en papel mientras que otros son textos en soporte electrónico; no importa, al fijarse en un soporte todas son manifestaciones. Algunas son textos y otras son audiolibros. Simplemente algunas de ellas son manifestaciones digitales.

En el otro ejemplo –el de una obra artística– cada interpretación del *Concierto de Brandenburgo n° 2* de Bach que se hace con distinta orquesta y director es una nueva expresión de esa obra. Por ejemplo, la interpretación del *Concierto de Brandenburgo n° 2* con Herbert Von Karajan al frente de la Orquesta Filarmónica de Berlín y la dirigida por Carlos Miguel Prieto y la Orquesta Sinfónica de Minería son dos expresiones distintas de la misma obra de Bach. Sólo la de Von Karajan fue grabada y por tanto se convirtió en una manifestación de la obra; la otra no. La primera se consigue grabada en CD en Deutsche Gramophon⁴⁴ y además puede descargarse en el mismo sitio vía Internet en formatos FLAC y mp3: dos posibles ítems de esa manifestación de esa expresión en particular. La segunda, que escuchamos en la sala de conciertos Nezahualcóyotl, en efecto fue una expresión de la obra, existió como tal, pero como no fue grabada en un soporte no fue materializada en una manifestación, y por lo mismo no hay ítems de ella.

Resumiendo, una obra es un producto literario o artístico diferenciable uno del otro y que se conceptualiza en forma general, sin mucha precisión. Una expresión es una versión, edición, traducción o interpretación de esa obra que se hace en un cierto momento y lugar y no forzosamente tiene que haber sido materializada en un

⁴³ http://es.wikipedia.org/wiki/La_vuelta_al_mundo_en_ochenta_días

⁴⁴ J.S. Bach, *Brandenburgische Konzerte. Nos. 1, 2, 3. Recordings 1964-1965*, Berliner Philharmoniker, Herbert von Karajan [CD].

Cómo preservar mi patrimonio digital personal

soporte físico; esto es muy obvio en las obras artísticas como las representaciones teatrales o las interpretaciones musicales. Cuando cierta expresión de una obra es fijada sobre un soporte físico o electrónico entonces existe además una manifestación y representa a “todas las copias que provienen esencialmente de la misma copia maestra distribuidas por la misma entidad”. La expresión tiene que ver más con el contenido literario y artístico de una obra y excluye lo que no tiene que ver con esos contenidos como el tipo de letra o paginación en los contenidos de textos o la separación de pistas si el soporte es disco de acetato o CD o casete en lo relativo a un contenido musical. La manifestación tiene que ver más con los aspectos físicos del contenedor o soporte de la obra; obviamente ha sido fijada sobre un soporte, digital o no, y siempre contendrá una expresión en particular de una obra.

Por lo anterior podemos hacer notar que una cierta expresión agrupa todas las manifestaciones de una obra cuyo contenido literario o artístico es igual aunque el soporte no sea igual: por ejemplo, una misma grabación de una orquesta y director presentada en disco de vinilo, CD, audiocasete o descarga vía la red; si el contenido artístico es igual, tenemos varias manifestaciones de una misma expresión. Si fuese otra grabación de la misma obra, con esa orquesta o director pero en otra ocasión, o el mismo director con otra orquesta, o la misma orquesta con otro director, tendríamos otra expresión distinta de esa obra. De cada manifestación puede existir al menos un ítem o ejemplar.

Resumiendo, nosotros poseemos en nuestra computadora ítems –es decir, ejemplares– de ciertas manifestaciones –esto es, copias fijadas sobre un soporte electrónico– de ciertas expresiones –o sea versiones– de ciertas obras. Algunas veces deseamos sólo una versión de una obra, algunas otras, deseamos varias versiones de esa obra a la vez; esto es perfectamente válido. Lo importante es no tener versiones inútiles, duplicadas y ociosas que sólo nos creen confusión. Conservemos por tanto las versiones útiles y desechemos las inútiles.

DIGITALIZACIÓN

Este no es un libro pensado en cómo enseñar a digitalizar documentos. Como ya mencionamos, esta es una obra para saber cómo preservar material que nace digital: textos, sonido, imágenes, etc. No obstante y dado que algunos lectores pueden tener esta inquietud, conviene hacer algunas reflexiones rápidas al respecto.

En primer lugar, conviene recordar que ya establecimos que digitalizar es el proceso de convertir un documento analógico en digital con objeto de poderlo almacenar, transmitir y percibir en una computadora, equipo de teleproceso, etc. En este caso este proceso de hacerlo digital significa convertirlo a un formato basado exclusivamente en números –generalmente binarios– bajo un cierto patrón arbitrario preestablecido antes de ponerlo en su nuevo formato y soporte, un dispositivo asociado a una computadora. A los documentos que nacen digitales se les llama documentos digitales de generación y a los documentos que estaban en un soporte tradicional y han sido digitalizados se les llama documentos digitales de transformación.

Hemos mencionado también ya que existe un importante factor documental –o metodológico– que tiene que ver primordialmente con la valoración de mis documentos que están sobre soportes analógicos o tradicionales con el fin de tasar o establecer oficialmente el valor que le otorgo a los mismos con propósito de determinar para cada material si debe ser digitalizado o no, y si es el caso, el periodo y condiciones de su preservación. No todos los documentos existentes en soportes tradicionales deben ser digitalizados y de los que lo fueron, sólo algunos deben ser preservados a largo plazo, y no todos por el mismo periodo.

Partiendo entonces del hecho de que no es recomendable digitalizar toda nuestra colección de materiales en soportes tradicionales, sino sólo algunas partes de ella, es importante establecer perfectamente y de antemano cuáles deben ser los criterios para que un cierto material que poseo sobre esos soportes tradicionales: textos, fotos, documentos sobre papel; discos y cintas analógicos, etc., deba ser digitalizado. Las preguntas cruciales aquí deben ser: ¿qué debe digitalizarse y con qué prioridades? ¿Por qué y para qué

Cómo preservar mi patrimonio digital personal

quiero hacerlo? De otra forma puedo estar emprendiendo un proyecto tremendamente costoso en tiempo, espacio y dinero y que a la larga puede tener un beneficio reducido o nulo. Para contestar estas preguntas, es muy conveniente usar listas de criterios ya establecidas y, de entre ellas, algunas que sean sencillas y ya adaptadas a las necesidades de una persona, como por ejemplo, los *Criterios de selección para la preservación por reformateado digital* emitidos por la División para la preservación por reformateado de la Biblioteca del Congreso de los EUA.⁴⁵ En ellos se menciona:

La selección de materiales para la preservación por reformateado digital se basa en el valor, uso, condición, características del documento original y conveniencia de las reproducciones digitales para uso y acceso.

- ❖ Valor. Se dará prioridad a los materiales de alto valor personal o profesional, raros, únicos y en riesgo. El uso de copias digitales de este tipo de información es altamente deseable; en parte para servir como preservación preventiva, así como por su seguridad, lo que se pretende lograr al reducir el manejo de los originales.
- ❖ Condición. Los documentos que no puedan usarse ya adecuadamente por daño o fragilidad son candidatos para la digitalización, ya que son documentos que se encuentran actualmente en soportes inestables.
- ❖ Uso. Los materiales originales que tengan una tasa de demanda muy alta o frecuente de nuestra parte son fuertes candidatos para digitalización.
- ❖ Características del original. Los originales en distintos formatos físicos y con características diferentes son elegibles para reformateado digital, incluyendo documentos encuadernados o sueltos, fotografías, negativos, folletos, carteles, trípticos, etcétera.
- ❖ Calidad y valor de uso del documento digital resultante. El documento digital resultante del proceso de digitalización debe cumplir los requerimientos que hayamos preestablecido al respecto.

⁴⁵ Library of Congress Preservation Website, *Selection Criteria for Preservation Digital Reformatting* [en línea].

Una vez que se ha hecho la valoración para decidir cuáles documentos vale la pena digitalizar, es necesario establecer sus periodos de retención, los formatos en que serán digitalizados, su calidad y valor de uso para nosotros. Hecho esto, puede procederse a la digitalización. Posteriormente, estos materiales pueden ya ser tratados como materiales digitales y deberán entonces establecerse sus demás atributos y características: metadatos, respaldos, versiones, etcétera.

PRESERVACIÓN Y CONSERVACIÓN DOCUMENTAL DIGITAL

Antes de continuar conviene establecer con toda claridad qué entendemos por preservación documental digital y por conservación documental digital, de otra forma estaremos discutiendo un tema sin tener claro el objeto de estudio. Y aunque a primera vista pareciera que los términos son fáciles e intuitivos de entender, podremos observar que en realidad no lo son tanto, y que tienen sus detalles y particularidades muy especiales, específicamente en el campo de los documentos digitales. Muchos piensan que preservar es sinónimo de conservar o que es sinónimo de almacenar.

El problema principal en la actualidad acerca de estos términos es poder hacer la distinción entre soportes físicos y soportes digitales. Esto se debe a que –históricamente hablando– las definiciones de preservación y conservación se remontan a principios y mediados del siglo xx provenientes de conceptos de preservación y conservación de edificios históricos y objetos provenientes de sitios arqueológicos; estos principios y conceptos fueron establecidos en la *Carta de Atenas*,⁴⁶ de 1931, y posteriormente en la *Carta de Venecia*,⁴⁷ de 1964. Después estos conceptos fueron simple y lla-

46 International Council on Monuments and Sites (ICOMOS), *The Athens Charter for the Restoration of Historic Monuments* [en línea].

47 International Council on Monuments and Sites (ICOMOS), *The Venice Charter. International Charter for the Conservation and Restoration of Monuments and Sites* [en línea].

Cómo preservar mi patrimonio digital personal

namente extrapolados a la conservación de documentos sobre soportes en papel o semejantes –microfichas, filmes, discos, etcétera–.

Las definiciones de esa época reflejan ese contexto edificios-objetos arqueológicos-papel:

Conservación tradicional:

Lo referente a las acciones tomadas para prevenir posteriores cambios o deterioro en objetos, sitios o estructuras. –ARTES–

El conjunto de medidas administrativas y prácticas, tales como el guardado en cajas, manejo cuidadoso, limpieza y control ambiental que aseguren la supervivencia de los documentos sin la intervención de especialistas. –ARCHIVÍSTICA–

En el ámbito de los archivos y bibliotecas, son todas aquellas medidas destinadas a proteger adecuadamente los documentos, con el fin de prolongar su utilización en condiciones óptimas durante el mayor tiempo posible. –BIBLIOTECAS–

Preservación tradicional:

Procesos y operaciones involucrados en asegurar la supervivencia intelectual y técnica de colecciones y registros documentales a través del tiempo. –BIBLIOTECAS Y ARCHIVOS–

Conjunto de consideraciones administrativas y financieras, las que incluyen almacenaje y acomodo, personal, políticas, técnicas y métodos involucrados en preservar materiales de bibliotecas y archivos así como la información contenida en ellos.⁴⁸

⁴⁸ International Federation of Library Associations (IFLA), *Principios para el cuidado y manejo de material en bibliotecas* [en línea].

Las actividades asociadas al mantenimiento en buenas condiciones del material de archivo y biblioteca, para su utilización en forma original.⁴⁹

Al integrar una resultante de todos estos conceptos, podemos concluir de manera sucinta que el fin ulterior y a largo plazo de estas acciones era el de preservar; es decir, asegurar la permanencia de los documentos a largo plazo. Para ello, debemos conservar, esto es, protegerlos y resguardarlos anticipada, adecuada y permanentemente; en caso de deterioro o daño los debemos restaurar. No obstante, nótese que bajo este enfoque siempre se establecieron y trataron de controlar los factores internos y externos que afectaban la durabilidad de los objetos físicos. Entre los primeros factores –los internos– se establecieron parámetros que mejoraran su estructura intrínseca: la acidez del papel y su desadificación; cinta de celuloide como soporte a filmes y sonido con calidad archivística superior; óxido férrico de alta duración para cintas y discos magnéticos, etc. En los factores externos se establecieron parámetros para el manejo de esos materiales: atmósferas controladas en cuanto a oxígeno, microorganismos, humedad, luz, temperatura, etc. Como puede verse, durante muchos años, en el mundo de los medios tradicionales hemos tratado de conservar primordialmente los soportes para así preservar las obras registradas en ellos.⁵⁰

Con el advenimiento de los documentos sobre soportes digitales esto ha ido cambiando paulatinamente; hoy en día en efecto se sigue pretendiendo conservar los soportes, pero se trata con mayor esfuerzo de preservar los contenidos. El soporte se vuelve ya un accesorio secundario.

Por lo mismo, los conceptos actuales de conservación digital y preservación digital han evolucionado también adaptándose a ese nuevo contexto:

49 Glosario de la American Library Association (ALA), *The ALA Glossary of Library and Information Science*.

50 IFLA, *op. cit.*

Cómo preservar mi patrimonio digital personal

Preservación documental digital:

La totalidad de principios, políticas y estrategias que controlan las actividades destinadas a asegurar la estabilización física y tecnológica, así como la protección del contenido intelectual de materiales –datos, documentos o archivos– y cuyo fin ulterior y a largo plazo, es el de asegurar la permanencia y acceso del contenido de documentos digitales confiables a lo largo del tiempo y las tecnologías, independientemente de su soporte, formato o sistema.⁵¹

Conservación documental digital:

Acciones tomadas para anticipar, prevenir, detener o retardar el deterioro del soporte de obras digitales con objeto de tenerlas permanentemente en condiciones de usabilidad, así como la estabilización tecnológica, la reconversión a nuevos soportes, sistemas y formatos digitales para garantizar la trascendencia de los contenidos.⁵²

Si se analizan con cuidado estas definiciones actuales, puede verse que ya incluyen este contexto moderno donde el soporte es importante, pero lo es más todavía el contenido. El soporte puede ser desechable y se cambiará con el tiempo; el contenido obviamente no.

Puede notarse también que –en el mundo digital– conservación es un conjunto de acciones y medidas para “anticipar, prevenir, detener o retardar el deterioro del soporte de obras digitales”. Como puede verse, su alcance es el corto plazo, y su ámbito es el soporte. También se le conoce como mantenimiento digital.

Preservación es a la vez un objetivo y un resultado: como objetivo, preservación es en esencia “el aseguramiento a largo plazo de la permanencia y acceso del contenido de documentos digitales confiables a lo largo del tiempo y las tecnologías”. Al aplicar y lograr

51 InterPARES Project, *The International Research on Permanent Authentic Records in Electronic Systems. Glosario InterPARES de Archivística* [en línea].

52 *Idem.*

este objetivo se obtiene como resultado –se logra– la preservación de los materiales, esto es: se obtiene la “protección del contenido intelectual de materiales –datos, documentos o archivos–”. Su alcance es el largo plazo y su ámbito es el contenido.

Con estas definiciones en mente, podemos reiterar ahora que el propósito de esta obra es aprender cómo lograr la preservación a largo plazo del contenido de nuestros materiales digitales de manera confiable, aplicando la conservación; esto es, un conjunto de acciones y medidas para anticipar, prevenir, detener o retardar el deterioro del soporte de obras digitales, a corto y mediano plazo.

Para no olvidarlo, recuérdese entonces este pequeño resumen:

Preservación — del contenido, a largo plazo.

Conservación — del soporte, a corto y mediano plazo.

COPYRIGHT, DERECHOS DE AUTOR Y *CREATIVE COMMONS*

El *copyright* o derecho de copia forma parte de los derechos de propiedad intelectual. Es un conjunto de disposiciones legales que, si bien varían en cada país, en términos generales otorgan derechos y prerrogativas al creador de una obra científica, artística o literaria para controlar cómo será utilizada esa obra. En teoría, el propósito básico del *copyright* es promover el progreso del conocimiento dando al autor de cada obra un incentivo económico para crear nuevas obras.

Dado que en esencia este principio fue creado y perfeccionado en la época de las publicaciones en soportes tradicionales, para lograr que los derechos de copia fueran efectivos el derecho de copia ha estado basado en un principio básico, el de “todos los derechos reservados”. Esto es: para copiar, difundir, distribuir o modificar una obra, insoslayablemente consulte con el autor. De esta forma el *copyright* ha sido por casi dos siglos el mecanismo para garantizar a un autor sus derechos y prerrogativas para decidir qué quiere hacer

Cómo preservar mi patrimonio digital personal

con su obra. Lo opuesto sucede cuando una obra pasa al dominio público y salvo el derecho de ser reconocido por su obra, el autor no puede imponer ninguna restricción al uso de su obra. Este es un principio de “ningún derecho reservado”.

El *copyright* tuvo su origen en los países anglosajones; tiene su equivalente en los países latinos en el derecho de autor. Aunque muchos piensan que son sinónimos, no es así: el concepto jurídico del derecho de autor es diferente del concepto jurídico del *copyright*. El derecho de autor proviene del derecho romano-canónico, típico de nuestros países iberoamericanos y está conformado a su vez por derechos morales y derechos patrimoniales. Los derechos morales consagrados en estas legislaciones tienen su origen en el Humanismo y la Revolución Francesa, de donde se extienden a España y de ahí a los países latinoamericanos y a algunos otros de Europa y tienen que ver más con el aspecto subjetivo del acto de creación de una obra por un autor dado; la esencia de estos derechos reside en que la obra forma parte integrante de la personalidad del autor, es una creación de su espíritu, el fruto de su pensamiento; de manera que no puede ser disociado enteramente de la persona, por lo que incluso cuando ha cedido sus derechos patrimoniales sobre la obra ésta continúa asociada a él y en cierta medida bajo su dependencia. En resumen, es un derecho del ser; los derechos patrimoniales son los derechos de explotación comercial de la obra y son los que en realidad se parecen al *copyright*.

En contraste, el *copyright* proviene del derecho consuetudinario anglosajón –*common law*–. En este, los derechos morales no son primordiales y su esencia está más intrínsecamente ligada al derecho de copiar o explotar una obra; es decir, su enfoque es más comercial. De hecho, los derechos morales se introducen en la ley inglesa hasta 1988 y no son perpetuos; en la legislación de los EUA aún no se han reconocido ni se vislumbra la intención de hacerlo. Ello se debe a que en esos países el derecho de copia es básicamente un derecho de propiedad, de manera que una obra intelectual es en principio un artículo de consumo, algo con lo que se puede comerciar bajo el control de la persona o institución que tiene los derechos sobre él. En resumen, es un derecho del “tener”.

Aunque son usados como sinónimos, puede observarse que existen ciertas diferencias de fondo. Resumiendo, el derecho de autor se basa en que el autor tiene un derecho natural sobre su obra y su comercialización es parte de éste; el *copyright* se basa en que el autor y la sociedad establecen una negociación en la que ambos buscan un beneficio mutuo.

Si bien las diversas legislaciones del mundo han tratado de adaptar sus disposiciones legales a la era de las publicaciones digitales, la verdad es que aún después de varias décadas todas sufren todavía de muchas lagunas legales, parches y situaciones forzadas que no acaban de conciliar el mundo de los derechos de las publicaciones tradicionales con el de las digitales.

La facilidad de publicación, acceso y distribución de obras a través de Internet ha hecho que un sinnúmero de autores publiquen al instante en la red, en todo el mundo, sin necesitar de los servicios de una empresa editorial que se encargue de las autorizaciones de derechos. Por lo mismo, son ya millones de personas que han creado obras en una amplia variedad y que han decidido distribuir las vía la red. Esto ha requerido de mecanismos que permitan de una manera más rápida y expedita la amplia utilización de esos materiales al tiempo que se respetan los derechos de autor.

El concepto del *copyleft* tuvo su origen en el sistema de comercialización o uso de los programas de computadora o *software*. Bajo los esquemas típicos de adquisición comercial, el usuario no adquiere la propiedad del programa, sólo el derecho a usarlo; por lo mismo, no puede mejorarlo, modificarlo o adaptarlo en forma alguna; de hecho en la mayoría de los casos se le prohíbe ver hasta el código fuente que lo origina. A este esquema se le ha denominado por décadas *software* propietario y ha estado basado en un estricto uso del *copyright*. Como contraposición a este esquema desde hace varias décadas se creó un esquema de *software* libre bajo el cual numerosos autores de aplicaciones de programación han puesto a la vista sus programas, permitiendo no sólo su uso, sino su distribución y modificación. Pero, ¿cómo evitar que alguna persona o institución, al hacer algunos cambios, no decidiese registrar toda la obra a su nombre y ponerla bajo el régimen de protección del

Cómo preservar mi patrimonio digital personal

copyright, evitando así que esta obra siguiese su camino en el uso generalizado y público? La respuesta fue el concepto de *copyleft* –como opuesto a *copyright*– popularizado por Richard Stallman, presidente de la *Free Software Foundation* y principal promotor mundial del concepto de *software* libre.

El propósito del concepto del *copyright* ha sido el de mantener privado y bajo un propietario el uso de una obra, en este caso del *software*. Como contrapeso –y de ahí su nombre– el propósito del *copyleft* ha sido el de mantener libre a una cierta pieza de *software*, evitando su privatización. Bajo este principio, la declaración del *copyleft* en una obra es en apariencia la misma que la del *copyright*, informando del nombre del autor y de su calidad de creador o propiedad intelectual sobre la obra, sólo que en el *copyleft* el autor autoriza explícitamente a los usuarios a usar, agregar, modificar, mejorar, adaptar, etc., los trozos de código que consideren pertinentes. Permite además obtener copias de la obra y distribuirla libremente a discreción, pero lo más importante de todo, el autor agrega una restricción para los usuarios posteriores, la de agregar restricciones a los subsecuentes usuarios a él. Esto es de singular importancia, ya que esta restricción de restringir, paradójicamente, se vuelve la clave de una distribución libre. De esta forma, cada usuario más allá del autor es libre de distribuir la obra, pero no puede imponer restricciones adicionales a su distribución. Cada nuevo usuario puede modificarla, señalando inclusive dentro de la obra la parte que es de su autoría, pero no puede restringir la capacidad de otros a modificarla a su vez o distribuirla. De esta manera se garantiza que la obra no podrá ser llevada al régimen propietario más adelante por alguna persona o institución malintencionada.

Si bien el concepto del *copyleft* ha sido utilizado por décadas como un mecanismo de distribución de *software*, probó desde hace mucho sus bondades y ha ganado muchos adeptos. De este concepto surgieron nuevas ideas para publicar y distribuir otros tipos de obras, como las musicales, fotográficas o los textos, sobre todo aquellos que viajan a través de la red mundial. A partir de entonces, cada vez es más frecuente encontrarse con algún tipo de do-

cumento u obra de texto o artística que establece su esencia de *copyleft*, manteniendo así su capacidad de circular ampliamente y sin restricciones por la red. Además, el concepto de *copyleft* no atenta por sí mismo contra las intenciones de obtener ganancias monetarias de una obra. No es un concepto de creación con fines sólo altruistas. *Copyleft* no implica tácitamente obras gratuitas. El autor puede comercializarlas y pedir y obtener ganancias por su creación.

En la actualidad existen varias aplicaciones y variantes del principio del *copyleft*; entre ellas la más aceptada, difundida y utilizada para publicaciones de todo tipo es la conocida como *Creative Commons* o simplemente CC.

Creative Commons es una organización no gubernamental sin fines de lucro orientada a darle a un autor la capacidad de establecer los límites de uso y explotación de cada una de sus obras publicadas en Internet desde el momento que la publica y distribuye. CC fue fundada en 2001 en la Escuela de Leyes de la Universidad de Harvard por Lawrence Lessig, James Boyle, Michael Carroll, Hal Abelson, Eric Saltzman y Eric Eldred trasladándose al año siguiente a la Escuela de Leyes de la Universidad de Stanford; la organización fue presidida por el primero hasta 2008. Nació como un proyecto de licencias prototipo –o simplemente tipo– para trabajos artísticos que no limitase por completo el uso o reproducción de esas obras. En poco tiempo se convirtió en un proyecto internacional, dentro del cual las licencias tipo se fueron complementando y diversificando y se redactaron cada vez por un grupo técnico-legista de cada país de acuerdo a las disposiciones y estructura legal de cada uno de los países que quería integrarse al proyecto. Así en 2005 se pusieron a disposición del público las licencias de CC adaptadas a la legislación mexicana sobre derechos de autor.

En la práctica, CC es una herramienta que permite a cualquier persona que desea publicar una obra en Internet –texto, música, imagen, video, etc.– seleccionar al instante desde su sitio web y sin ayuda de asesoría legal una serie de diversas licencias tipo que ya han sido redactadas previamente por legistas expertos, y donde el autor selecciona ciertos derechos que él quiere otorgar a los que lo

Cómo preservar mi patrimonio digital personal

consulten así como ciertos derechos que él se reserva. Las combinaciones posibles de derechos resultantes son numerosas haciendo que el autor pueda hacer una cesión muy pormenorizada de lo que permite y lo que no. Esta licencia queda adosada al instante a su documento electrónico en tres formas: legible al público en general, en redacción formal legal y en formato de etiquetas de máquina legibles por computadora para que esas características legales sean halladas por un navegador web. De esta forma, otro usuario que encuentra el documento en la web y desea utilizarlo sabe exacta e instantáneamente lo que el autor le permite hacer y lo que no, sin necesidad de contactos o permisos adicionales.

Esta metodología ha sido adoptada ya en 28 países y está en vías de adopción en varias decenas más, ya que simplifica enormemente el otorgamiento y retención de derechos en publicaciones electrónicas sin sustituir las leyes de derechos de autor ni el dominio público, pero facilitando enormemente la distribución legal y organizada de documentos electrónicos. Ha probado ya ser un excelente complemento a las disposiciones legales en el mundo de las publicaciones electrónicas, ya que ha probado ser una herramienta que llena muy bien la brecha que existe entre el concepto del *copyright* –todos los derechos reservados–, donde nada es permitido sin autorización expresa del autor, y el concepto de dominio público –ningún derecho reservado–, donde no se requiere ningún permiso del autor. En suma: las licencias CC autorizan el uso de una obra digital bajo ciertas condiciones sin dejar de protegerla legalmente. Es en suma un esquema de ciertos derechos reservados.

Las licencias CC se basan en cuatro premisas que el autor puede combinar pudiendo lograr así varios tipos de licencias diferentes. Estas premisas son: 1. Reconocimiento –*Attribution*–, 2. Comercializable –*Comercial*–, 3. Obras derivadas –*Derivative Works*– y 4. Licenciamiento Recíproco –*Share alike*–.

Reconocimiento. El material creado por un autor puede ser copiado y distribuido por terceros, siempre y cuando el nombre del autor o autores originales y subsecuentes sea mostrado en los créditos del que lo usa subsecuentemente.

Comercializable o no. El autor permite o no que el material creado por él pueda ser distribuido, copiado y exhibido por terceros con fines comerciales, o pide una compensación económica en caso de comercialización.

Obras derivadas o no. El autor permite o no que puedan hacerse adaptaciones, cambios, traducciones, resúmenes, versiones alternativas, etc., derivadas de su obra. En este caso, se permite que el nuevo autor agregue su participación en los créditos de reconocimiento.

Licenciamiento Recíproco. El material creado por un autor puede ser distribuido, copiado y exhibido por terceros siempre y cuando al compartirlo de nuevo –sea original o derivado– no se agreguen nuevas restricciones para hacerlo. Las obras originales o derivadas de ella deben estar bajo los mismos términos de licencia que el trabajo original.

Cuando un autor desea publicar una obra en Internet, previamente entra al sitio web de *Creative Commons* del país que le corresponde, y con ayuda de una plantilla existente al efecto selecciona de entre las posibles opciones mencionadas las que desea para esta obra en particular; al final, tendrá a su disposición una página legible por el público, el texto completo legal de la licencia y los *tags* o etiquetas necesarios para búsqueda en los navegadores. Todo ello debe ser agregado a la página web donde se publica esa obra, para informar así a los potenciales usuarios qué es lo que el autor permite en el uso de ella y qué es lo que no. De esta manera la obra queda cubierta legalmente de usos no deseados y maliciosos dentro de la red pero instantáneamente a la vez se maximiza la oportunidad de ser utilizada, copiada y distribuida por otros usuarios de la red.

Cuando nosotros encontramos en la web algún documento que nos interesa copiar y redistribuir, debemos verificar que esto es legalmente posible de acuerdo con las declaraciones legales del propio documento. Las declaraciones *Creative Commons* facilitan mucho este aspecto, ya que al ver la licencia, instantáneamente podemos saber si estamos autorizados a hacer lo que deseamos con él

Cómo preservar mi patrimonio digital personal

y hasta dónde. Por ejemplo, una licencia expresada en una simple línea como “*Creative Commons Attribution-NonComercial-NoDerivs-Share Alike-United States 3.0*” en un sitio web nos informa que esa obra se publica bajo una licencia del tipo *Creative Commons* donde para ser reutilizada debe tener reconocimiento al autor, sin uso comercial, debe ser tal cual el original sin trabajos derivados del mismo y con licenciamiento recíproco al original, según la licencia CC para los EUA en su última versión 3.0. También es muy común encontrar la licencia simplemente por siglas: CC-BY-NC-ND-SA US 3.0. CC de *Creative Commons*, BY de “menciónese autoría”, NC de NonCommercial o sin fines comerciales, ND de NonDerivative o sin derivados, SA de Share alike o licenciamiento recíproco, US o USA que indica que fue generada en los EUA.

Una licencia “*Creative Commons Attribution-NonComercial-Share Alike-Mexico 2.5*” o simplemente CC-BY-NC-SA Mexico 2.5 nos indica que para utilizar y redistribuir ese documento se requiere dar atribución al autor, no se permite con fines comerciales, se autorizan obras derivadas y se tiene que redistribuir con licencia igual al original, según licencia tipo México en su última versión 2.5. Esta línea abreviada puede ser agregada a nuestra pregunta hecha en un buscador web. También es común ver esto expresado con íconos de CC. En el Anexo 5 puede verse como ejemplo el texto completo de cómo queda una licencia extensa en redacción de tipo legal para este caso anterior.



Al guardar nosotros un documento que hemos tomado de la web en estas circunstancias –sea texto, imagen, etc.– es sumamente recomendable agregar en los metadatos del mismo la dirección URL de donde se tomó el documento así como la línea resumen o las siglas de la licencia; de esta forma ese documento guardará

dentro de sí mismo los datos de la autorización legal con la que fue tomado y en caso de futuras consultas y aclaraciones tendremos disponible esa información evitando así posibles dolores de cabeza futuros en este aspecto.

Seguridad informática

La sabiduría consiste en poder reconocer diversos peligros y escoger de entre ellos el menos dañino.

Niccolò Machiavelli, *El Príncipe*

Este no es un libro sobre seguridad informática; es un libro acerca de preservación documental digital y –como veremos un poco más adelante– no son sinónimos. Empero, es imposible plantear ninguna estrategia de preservación documental digital en un ambiente que no tenga un mínimo de seguridad informática. Fallas graves en esta última darán al traste con cualquier estrategia de preservación documental, sin importar qué tan bien diseñada esté. Por lo mismo, previo a entrar en materia de las recomendaciones y estrategias de preservación, es necesario dedicar un pequeño espacio al estudio y entendimiento de los conceptos básicos de la seguridad informática, donde también se pueden hacer algunas recomendaciones al respecto.

Hasta principios de la década de 1990, previo a la globalización de las telecomunicaciones y las redes mundiales de teleproceso, los sistemas de información tenían como una de sus características más relevantes la de funcionar de manera aislada o en redes privadas muy pequeñas en las cuales la seguridad proporcionada por el acceso físico y algunas simples barreras informáticas bastaban para que la seguridad de la información dentro de las computadoras estuviese garantizada; esto incluía especialmente a las computadoras personales. Por lo mismo, no había mucha preocupación al respecto ni estrategias específicas en este sentido. En la actualidad, los

Cómo preservar mi patrimonio digital personal

sistemas de información han sido sustituidos casi en su totalidad por Tecnologías de Información y Comunicaciones –TIC– convergentes, por inmensas y cada vez más complejas redes institucionales locales y regionales, por servidores y computadoras personales que cada vez tienen mayor capacidad de proceso y de acceso a otras computadoras y cuya interconexión se extiende mundialmente. Obviamente, la red global forma ya parte de la infraestructura operativa de sectores estratégicos de todos los países como el comercial, energía, transportes, banca y finanzas –por mencionar algunos– y desempeña un papel fundamental en la forma en que los gobiernos proporcionan sus servicios e interactúan con organizaciones, empresas y ciudadanía, y es un factor cada vez más creciente de intercambio de información de manera individual por parte de los ciudadanos toda vez que se forman redes sociales cada vez más complejas. El número y variedad de equipos, dispositivos, servicios y modalidades que integran la infraestructura de acceso se ha multiplicado, e incluye ya variados elementos de tecnología fija, inalámbrica y móvil, así como una proporción creciente de accesos que están conectados de manera permanente. Como consecuencia de todos estos cambios el volumen, naturaleza, disponibilidad y sensibilidad de la información que se intercambia a través de esta infraestructura se ha modificado y ha aumentado de manera muy significativa.

Por lo mismo, hoy en día la información es un activo muy valioso para casi todas las organizaciones; para algunas de ellas es su activo más valioso y por ello se invierten considerables recursos en crearla, administrarla, mantenerla, distribuirla, etc. Algo muy semejante puede decirse de la información creada y en propiedad de las personas. Aparte de su valor intrínseco, por la misma convergencia de variados sistemas tecnológicos se produce, procesa, almacena y distribuye cada vez más y más información proveniente de procesos informáticos en formatos digitales, acumulándose ya a lo largo de los años cantidades muy considerables de este insumo. Aunado a la creciente abundancia y economía de medios de almacenamiento que describimos en el primer capítulo, como consecuencia vemos crecer día a día en los equipos de las personas información

depositada en estas formas digitales, la que casi sin excepción debe ser operada, distribuida y consultada tanto en forma local como en forma remota.

Pero precisamente por esas condiciones contemporáneas de ser digital, multiaccesible y necesariamente operada en red, la información se enfrenta a riesgos de daño o pérdida. Como resultado de esa creciente interconexión masiva y global, los sistemas y las redes de información se han vuelto más vulnerables ya que están expuestos a una cantidad creciente así como a una mayor variedad de amenazas remotas. Esto hace a su vez que surjan nuevos retos que deben abordarse en materia de seguridad. La seguridad informática pretende eliminar o contener estos daños o pérdidas desde un principio. La razón de ello es simple: pretender aplicar estas medidas *a posteriori* a una información que ya está almacenada es mucho más caro y difícil; y muchas veces inútil, puesto que ya está dañada o no es confiable.

Gene Spafford, experto en seguridad, hizo una analogía entre los esfuerzos de los especialistas en seguridad informática y los de los cardiólogos:

Nuestros pacientes están conscientes de que la falta de ejercicio, la dieta rica en grasas y el tabaco son dañinos para todos ellos; no obstante, continúan fumando, comiendo grasa y viviendo sedentariamente hasta que llega el infarto. Entonces todos ellos claman por una píldora mágica que los cure inmediatamente, sin ningún esfuerzo. Todos además afirman que su condición no es su culpa: es defecto genético heredado de sus padres, o es culpa de las compañías tabacaleras o de hamburguesas, etc. Los culpables siempre son terceros; nunca ellos. De pasada nos reprochan a los médicos el no haberlos cuidado mejor y evitado el infarto. ¿Esto les suena familiar? Pero no tiene por qué ser así; podemos hacerlo mejor. Necesitamos dejar de hacer las cosas como lo hacemos usualmente y comenzar a enfocarnos en una calidad de principio a fin. La seguridad informática debe ser construida desde el mismísimo principio, y no tratar de tapan el pozo después del accidente.⁵³

53 Gene Spafford (2000), *23rd National Information Systems Security Conference*.

Cómo preservar mi patrimonio digital personal

Por lo anterior, desde hace ya un buen tiempo que se han hecho esfuerzos a nivel mundial en materia de seguridad informática para poder contender contra los riesgos de la operación informática en red. Numerosas teorías, metodologías, técnicas, servicios, equipos y programas han sido creados al efecto buscando ofrecer a las empresas y personas seguridad en su información digital. En esencia, la gran parte de este esfuerzo se ha hecho para las organizaciones ya que ahí se manejan cantidades más grandes de información muy sensible y esto representa un gran volumen de negocio. Empero, dentro de esta obra y por su misma naturaleza omitiremos las consideraciones destinadas a las organizaciones limitándonos exclusivamente a aquéllas que son aplicables a los individuos, especialmente aquéllas relacionadas obviamente con la preservación de su material digital.

Definimos seguridad informática como: “el proceso de establecer y observar un conjunto de estrategias, políticas, técnicas, reglas, guías, prácticas y procedimientos tendientes a prevenir, proteger y resguardar de daño, alteración o sustracción a los recursos informáticos de una persona y que administren el riesgo al garantizar en la mayor medida posible el correcto funcionamiento ininterrumpido de esos recursos”.

CONCEPTOS FUNDAMENTALES DE LA SEGURIDAD INFORMÁTICA

Para poder comprender el concepto integral de la seguridad informática, es indispensable entender los diversos conceptos básicos que la componen, ya que de otra forma no es posible establecer una base de estudio; estos conceptos son:

Recursos informáticos

Consisten en el equipo de cómputo y telecomunicaciones así como sus dispositivos; los sistemas, programas y aplicaciones, también los datos e información de una persona. Se les llama asimismo activos informáticos. Obviamente todo tiene un valor: las compu-

tadoras y demás equipo, los programas, etc. Pero si reflexionamos a fondo, podemos llegar fácilmente a la conclusión de que lo más valioso que poseemos son nuestros documentos y datos: ¿cuánto valen nuestras fotografías y videos, nuestras grabaciones, nuestras notas, tareas, apuntes y tesis, nuestros correos electrónicos, los expedientes de pacientes, los datos o contabilidades de clientes, y toda esa información valiosa que hemos acumulado con tesón y paciencia por años? Al fin y al cabo, las computadoras, los sistemas y los programas son reemplazables, se pueden reponer en caso de pérdida, y en todo caso, la pérdida es sólo de dinero; la información valiosa no se puede reponer tan fácilmente, y a veces nunca. Siendo estimables todos los recursos informáticos, debemos poner especial atención en datos e información, ya que sin duda estos son los más valiosos por lo que representan para nuestro quehacer, además por el valor sentimental, intelectual o profesional que tienen para cada uno de nosotros. Y estos no se pueden reponer sólo con dinero.

Amenazas informáticas

*El problema del spam desaparecerá dentro de dos años.
Bill Gates, 24 enero 2004
Foro Económico Mundial en Davos, Suiza*

Las amenazas informáticas consisten en la fuente o causa potencial de eventos o incidentes no deseados que pueden resultar en daño a los insumos informáticos de la persona. Entre ellas, identificamos como las principales:

- ❖ El advenimiento y proliferación de *malware* o *malicious software*; esto es, código maligno o código mal intencionado: son programas cuyo objetivo es el de infiltrarse en los sistemas sin conocimiento de su dueño, con objeto de causar daño o perjuicio al comportamiento del sistema y por tanto a la información de una persona. Aquí se encuentran todos los virus, troyanos, espías, etcétera.
- ❖ La pérdida, destrucción, alteración o sustracción de información por parte del propio usuario u otra persona con acceso al equipo

Cómo preservar mi patrimonio digital personal

debidos a negligencia, dolo, mala capacitación, mal uso, ignorancia, apagado o elusión de dispositivos de seguridad y buenas prácticas.

- ❖ La pérdida, destrucción, alteración, sustracción, consulta y divulgación de información por parte de otras personas o grupos externos malintencionados.
- ❖ El acceso no autorizado a documentos e información de la persona.
- ❖ La pérdida, destrucción o sustracción de información debida a vandalismo.
- ❖ La intrusión a los equipos de cómputo personales por parte de cibercriminales: personas o grupos malintencionados con objeto de utilizarlos para acciones ilegales tales como enviar *spam* a otros equipos o lanzar ataques masivos para bloquear servidores de terceros.
- ❖ Los *phishers*, especializados en robo de identidades personales y otros ataques del tipo de ingeniería social.⁵⁴
- ❖ Los *spammers* y otros mercadotecnistas irresponsables y egoístas quienes saturan y desperdician el ancho de banda de las personas y organizaciones con mensajes inútiles.
- ❖ La pérdida o destrucción de información debida a accidentes y fallas del equipo: fallas de energía, fallas debidas a calentamiento, aterrizamiento, desmagnetización, rayadura o descompostura de discos duros y otros dispositivos de almacenamiento, etcétera.
- ❖ La pérdida o destrucción de información debida a catástrofes naturales: inundaciones, tormentas, incendios, sismos, etcétera.
- ❖ El advenimiento de tecnologías avanzadas tales como el cómputo *quantum*, mismas que pueden ser utilizadas para descifrar documentos, llaves, etc., al combinar complejos principios matemáticos y computacionales.

⁵⁴ Sarah Granger, *Social Engineering Fundamentals. Part I: Hacker Tactics* [en línea].

Riesgos informáticos

No se puede obtener un seguro para el auto después de haberlo chocado. Tampoco se pueden respaldar los datos una vez que se han perdido. Estos se deben proteger ahora anticipando desastres y desgracias, antes de que sea demasiado tarde.

T. E. Ronneberg

El riesgo informático consiste en la probabilidad de que un evento nocivo ocurra combinado con su impacto o efecto nocivo en la información de la persona. Se materializa cuando una amenaza actúa sobre una vulnerabilidad y causa un impacto. Los principales riesgos se agrupan como:

- ❖ Pérdida, destrucción o fuga de información valiosa o privada para la persona.
- ❖ Sustracción o alteración de datos personales para usos malintencionados.
- ❖ Introducción de programas maliciosos a los sistemas de la persona, que pueden ser utilizados para destruirlos u obstaculizarlos, usurpar recursos informáticos, extraer o alterar información sin autorización, ejecutar acciones ocultas, borrar actividades, robo y detentación de identidades, etcétera.
- ❖ Acciones de ingeniería social malintencionada: *phishing*, *spam*, espionaje, etcétera.
- ❖ Uso indebido de materiales sujetos a derechos de propiedad intelectual.
- ❖ Daño físico a equipos, programas, etcétera.

Vulnerabilidades informáticas

Una vulnerabilidad es alguna característica o circunstancia de debilidad de un recurso informático la cual es susceptible de ser explotada por una amenaza, intencional o accidentalmente. Las vulnerabilidades pueden provenir de muchas fuentes, desde el diseño o implementación de los sistemas, los procedimientos de seguridad, los controles internos, etc.; se trata en general de protecciones inadecuadas o insuficientes, tanto físicas como lógicas, procedimen-

Cómo preservar mi patrimonio digital personal

tales o legales de alguno de los recursos informáticos. Las vulnerabilidades al ser explotadas resultan en fisuras en la seguridad con potenciales impactos nocivos para la organización. Más detalladamente, provienen de:

- ❖ Fallas de origen desde el fabricante en el diseño o construcción de programas, sobre todo en aquellos que provienen de un mercado masivo; por ejemplo, sistemas operativos, aplicaciones, protocolos, etcétera.
- ❖ Confianza excesiva en algún único dispositivo o mecanismo de seguridad.
- ❖ Falta de seguimiento de los procedimientos de seguridad, producidos por haber disfrutado de un periodo de seguridad adecuado durante cierto lapso.
- ❖ No hacer caso al monitoreo o advertencias de seguridad.
- ❖ Pobre o nula administración de los activos informáticos, debida principalmente a un mal seguimiento de esos activos y sus contextos de seguridad asociados de forma integral.
- ❖ Cambio frecuente hacia nuevas plataformas informáticas.
- ❖ Planes de contingencia nulos o pobres, tanto para situaciones cotidianas como extremas.
- ❖ Ignorancia, negligencia o curiosidad por parte de los usuarios de los sistemas.
- ❖ Equipos, programas y redes provenientes de generaciones tecnológicas demasiado atrasadas.
- ❖ Falta de concientización de las personas en general acerca de la importancia de la seguridad.

Impactos

Los impactos son los efectos nocivos contra la información de una persona cuando una amenaza informática se materializa. Al suceder incidentes contra la seguridad informática pueden devenir en:

- ❖ Destrucción de patrimonio documental digital valioso e irremplazable para la persona.
- ❖ Disrupción en las rutinas y procesos de la persona con posibles consecuencias a su capacidad operativa.

- ❖ Pérdida de la privacidad en registros y documentos de personas.
- ❖ Incremento sensible y no programado en gastos emergentes de seguridad.

Cada uno de estos efectos nocivos puede cuantificarse de tal forma que se establezca el impacto de ellos en la información y consecuentemente en la persona.

Nótese que las amenazas son cosas que pueden suceder mientras que los riesgos son cosas que habiendo sucedido tienen un impacto. Pero hay diferencias: por ejemplo, la amenaza de que un dispositivo de almacenamiento –disco duro, memoria USB– de nuestro equipo falle algún día, física o funcionalmente, tiene una probabilidad relativamente alta; tarde o temprano fallará; es decir, es altamente probable que la amenaza se materialice en un riesgo, es sólo cuestión de tiempo. Pero si tenemos adecuadas protecciones de nuestra información, cuando eso suceda será una molestia menor, ya que podremos recuperar nuestra información de inmediato y ello no causaría mayor problema: es un ejemplo de amenaza muy alta con impacto mínimo. Por otra parte, la amenaza “probabilidad de que nuestro equipo se incendie” es relativamente baja, pero si esto llegase a suceder y nosotros no tuviéramos protección adecuada de los datos, el impacto de ese evento sería catastrófico. Este es un ejemplo de amenaza baja pero con impacto muy grave.

Por lo mismo deseamos siempre reducir en lo posible la probabilidad de que las amenazas se materialicen, pero debemos estar conscientes de que es imposible de llevar esto a probabilidad cero, por más cuidados que tengamos. Lo que sí podemos hacer es lograr que el impacto de esa amenaza que se materializa en un riesgo sea nulo o mínimo. En otras palabras: pon atención a minimizar las amenazas a tu información, pero sobre todo, pon atención a neutralizar o minimizar el impacto si el riesgo ocurriera.

De aquí se desprende que el mejor escenario para cualquier persona es tener amenazas con el mínimo posible de probabilidad de ocurrencia y con la máxima probabilidad de neutralizar su impacto. Obviamente el peor escenario se da cuando una persona tiene amenazas con alta probabilidad de que ocurran y mínima probabi-

Cómo preservar mi patrimonio digital personal

lidad de neutralizar o reducir el impacto. Este último escenario es sin duda una receta para el desastre y ocurre cuando se descuida la seguridad informática.

En resumen, la seguridad informática pretende identificar las amenazas y reducir los riesgos al detectar las vulnerabilidades nulificando o minimizando así el impacto o efecto nocivo sobre la información de las personas. Si analizamos y juntamos todo lo anterior estamos ya en posibilidad de comprender por qué la seguridad informática se definió entonces como “el proceso de establecer y observar un conjunto de estrategias, políticas, técnicas, reglas, guías, prácticas y procedimientos tendientes a prevenir, proteger y resguardar de daño, alteración o sustracción a los recursos informáticos de una persona y que administren el riesgo al garantizar en la mayor medida posible el correcto funcionamiento ininterrumpido de esos recursos”.

Principio básico de la seguridad informática

“La seguridad informática no es un producto que se compra u obtiene, es un proceso siempre continuo”.

El objetivo primario de la seguridad informática es el de mantener al mínimo los riesgos sobre los recursos informáticos –todos los recursos– y garantizar así la continuidad de las operaciones informáticas de la persona al tiempo que se administra ese riesgo informático a un cierto costo aceptable. Para ello utilizaremos métodos, servicios y productos técnicos. Nótese que en primer lugar mencioné métodos: esto es deliberado, ya que lo más importante en la seguridad personal no son los equipos y programas al efecto, sino la manera en que opera una persona. Se obtiene más seguridad de hábitos, métodos y procedimientos que de productos al efecto.

El objetivo secundario de la seguridad informática –algo de nuestro especial interés desde el punto de vista de la preservación documental– consiste en garantizar que los documentos y archivos informáticos de la persona existan y mantengan siempre su confiabilidad total. Este es un concepto que varía de acuerdo con el

entorno de organizaciones, archivos, contextos, autores, etc.⁵⁵ En nuestro contexto de esta obra recordemos que tratamos de construir un concepto de seguridad informática destinado a lograr la preservación documental digital para las personas. De acuerdo con esta premisa, podemos establecer esa confiabilidad como la unión de varias características esenciales: permanencia, accesibilidad, calidad, etc., las cuales en conjunto mantienen la esencia total de los materiales digitales personales y como consecuencia serán confiables para la persona. Ya hemos comentado qué significan estos conceptos en los incisos anteriores.

De aquí se desprende que la manera de lograr el objetivo secundario de la seguridad informática consiste entonces en que los documentos, registros y archivos informáticos que son propiedad de la persona cumplan ahora y en un futuro en la medida más completa posible con estas características esenciales, y como consecuencia mantendremos su esencia y por tanto su confiabilidad.

METODOLOGÍAS PARA LA SEGURIDAD INFORMÁTICA PERSONAL

La seguridad será siempre excesiva hasta el momento en que fue insuficiente.

Robbie Sinclair, Jefe de seguridad de la Oficina de Energía del Condado, NSW, Australia

Respecto a la computadora

- ❖ Asegúrate que tu sistema operativo esté programado para recibir de forma automática las actualizaciones del fabricante; éstas por lo general tienden a cubrir defectos de construcción y vulnerabilidades detectadas dentro del mismo.
- ❖ Siempre instala un sistema antivirus en tu computadora o móvil,

⁵⁵ Organización para la Cooperación y el Desarrollo Económico (OCDE), *Directrices de la OCDE para la Seguridad de Sistemas y Redes de Información: hacia una cultura de la Seguridad* [en línea].

Cómo preservar mi patrimonio digital personal

y asegúrate además de establecer los parámetros de seguridad que vienen en el cortafuegos *–firewall–* del mismo. En el sistema operativo Windows vienen en la opción “Inicio”-“Panel de Control”-“Centro de Seguridad”. Esto también es válido para los equipos Apple, los cuales no están exentos de *malware* como algunos afirman.

- ❖ Considera la instalación de una herramienta antiespía. *Spyware* es el nombre genérico dado a programas de espionaje que han sido diseñados para monitorear en secreto tus actividades dentro de tu computadora. Estos programas son instalados generalmente por el propio usuario sin saberlo, ya que vienen ocultos dentro de otro archivo que él mismo descarga, y obviamente su uso es casi siempre malintencionado. Así, alguien puede desde hacer cosas sencillas, como enviar estadísticas de uso a un fabricante, hasta capturar directorios de correos electrónicos, contraseñas, números de tarjetas y otra información sensible, y luego reenviarlas a defraudadores. Los programas antiespía ayudan a proteger tu equipo contra amenazas de espionaje. Estos programas vienen a menudo ya incluidos dentro de los paquetes de programas antivirus; si no es el caso, existen muchos productos que se pueden adquirir e instalar gratuitamente y por supuesto, algunos más con costo.

Respecto al uso

En Dios confiamos; a todos los demás los escanearémos en busca de malware.

Anónimo

- ❖ Nunca, jamás, por ningún motivo hagas caso de correos que aparentan venir de tu banco, de tu administrador de correo o web, de alguna entidad pública o de algún servicio que te pida que des tu contraseña o tu número de tarjeta o “pin”. Ningún servicio legítimo de esos en el mundo te pedirá que lo hagas. A esto se le llama *phishing* o suplantación de organizaciones. Son personas que están tratando falsamente de obtener tus datos. No importa cuán serias sean las advertencias de que tu cuenta será cancela-

da si no lo haces, ni que ostenten los logotipos de la supuesta institución. 100% de probabilidad que son falsos.⁵⁶

- ❖ Nunca sigas el vínculo o *link* de un correo que sea sospechoso; te llevarán a una página que aparenta en su presentación ser de la supuesta institución, pero seguramente es maliciosa. En caso de duda, debes teclear en tu navegador directamente la dirección deseada: <http://www.elsitioquedeseo.com>, nunca mediante el vínculo ofrecido. Además, te lo aseguro, no te ganaste un crucero de lujo por el Caribe sin comprar un boleto de una rifa.
- ❖ Cuando entras a un sitio web que sabes que legítimamente va a requerir los datos de contraseña, tarjetas, etc. –tales como *Paypal*, *Amazon*, una aerolínea, etc.– siempre verifica que en el momento que esto suceda, la página debe ostentar la leyenda “https” antes de su dirección URL. Esta “s” final garantiza que es un sitio seguro y legítimo. Algunos navegadores despliegan además la imagen de un pequeño candado cerrado en la esquina inferior derecha. Al dar “click” sobre el candado o en la barra del navegador podrá verse además el certificado de ese sitio. Si al momento que el sitio web pide nuestros datos cruciales no se observa en él alguna de estas características, es mejor abstenerse.

Respecto a las contraseñas o *passwords*

Las contraseñas son como la ropa interior: no debes andarla enseñando, no la compartas con nadie, no debe ser demasiado pequeña, es mejor mantener el misterio, no la dejes regada por ahí y cámbiala con frecuencia.

Office of Policy & Education, University of Michigan at Ann Arbor

⁵⁶ “Una de las maneras más fáciles de robar contraseñas es simplemente solicitarla al dueño. Un informe del 2007 del Inspector General del Tesoro para la Administración Tributaria de los EUA determinó en una prueba al efecto que el 60% de los empleados del Servicio de Rentas Internas compartían sus contraseñas con una persona que llamaba y decía ser del departamento de cómputo interno de la oficina. Después de eso, el IRS ha estado trabajando para mejorar su seguridad”. “Hacker Camps Train Network Defenders” [en línea].

Cómo preservar mi patrimonio digital personal

- ❖ Siempre usa contraseñas *–passwords–* dentro de los sistemas que manejan información sensible. Algunas personas deciden *–esto es muy fácil–* poner una contraseña desde el encendido del equipo. Por supuesto, deben usarse para correo electrónico, acceso a cuentas y servicios que tengas, etcétera.
- ❖ No es recomendable dejar que los sistemas “memoricen” las contraseñas, como lo proponen hacer ciertos manejadores de correo electrónico, navegadores de web, etc. Estas ofertas de “el sistema puede recordar tu contraseña para que no la teclees en un futuro” siempre son tentadoras, pero es mejor no tomarlas. En caso de hacerlo, debemos recordar que en un futuro, cualquiera que abra esa aplicación dentro de nuestro equipo, estará dentro de ella sin más trámite.
- ❖ A menos que sea forzoso, no selecciones esas contraseñas “seguras por complicadas” de forma aleatoria tal como sugieren muchos sistemas:⁵⁷ si tú seleccionas una contraseña de 8 o más dígitos, con combinaciones aleatorias de letras minúsculas y mayúsculas, caracteres especiales y números, tal como “F9x7*WýBv6&”, en efecto es segura; el problema es que 9 de cada 10 usuarios no la memorizan: la escriben en algún lado. Acaban pegadas en una nota, ya sea electrónica en la computadora o bien de papel dentro de un cajón del escritorio, abajo de la mesa o del teclado, o escritas en un archivo de texto dentro de la computadora que para colmo se llama “pw.txt” o algo parecido. Al final son más fáciles de descubrir y detectar que una contraseña sencilla. El problema se agrava cuando debes recordar no una sino varias contraseñas seguras de este tipo, ya que tienes una diferente para acceder a distintos sistemas. En este caso 99 de cada 100 usuarios acaban por escribirlas dando al traste con la seguridad. Cuando se manejan más de seis contraseñas de este tipo 999 de cada 1000 usuarios las escriben. Si alguien quiere detectar tu

57 “La gran mentira de la seguridad informática consiste en afirmar que el uso de contraseñas complejas por parte de los usuarios mejora la seguridad; en la vida real, las personas escriben lo que no pueden recordar. La seguridad mejora diseñando de la manera en que los seres humanos se comportan en realidad”.
Jakob Nielsen.

contraseña es mucho más rápido y práctico encontrar dónde la escribiste que descifrarla.⁵⁸

Ciertamente, no deben usarse contraseñas muy cortas o muy obvias, tales como “12345” o “abc123”; no deben usarse secuencias del teclado, como “querty” o “asdfg”; no debes usar tu nombre o sobrenombre, o del cónyuge, padres o hijos, novio, novia, mascota, etc.; son muy obvias, aun con las extensiones “paco97”, “susieQ”, “mary2012”, “lachiquis25” o escribiéndolos de atrás hacia adelante. No debes usar tu número de registro de Hacienda, de credencial de la escuela, licencia de manejo, etc. No deben usarse tampoco fechas obvias como el cumpleaños, aniversario, etc. Con frecuencia, estas fechas se obtienen con mucha facilidad de tu página de Facebook. Las fechas no obvias, tales como la fecha en la que entraste a la secundaria o que partiste de viaje por primera vez fuera del país –asumiendo que las recuerdas– son muy buenos “complementos” numéricos para añadir a una parte de texto en las contraseñas. Así, “susieQ23082004” se vuelve una contraseña no obvia pero a la vez nemotécnica, asumiendo que la fecha es lo suficientemente cercana a tu mente para no olvidarla, no así para alguien que está tratando de detectarla.

No deben usarse palabras solas existentes en el diccionario en ningún idioma; lo mejor es usar palabras fuera de diccionario, o combinaciones de palabras dentro del diccionario pero formando frases combinando a su vez con variantes ortográficas o fonéticas de las palabras, palabras con números o fechas, o pequeñas frases cortas, intercalando siempre por ahí algunas mayúsculas, pero tratando siempre de que el resultado sea nemotécnico –de otra forma no funcionará–. Así, contraseñas como: “Megoñcio890”, “MadonnaenelSuperBowIXLVI”, “kchunRaRa240512”, “krisZanches*99”, “el-tiempoNOpasa”, “8*50CADAmartes”, etc., son seguras y serán más útiles en la medida en que podamos memorizarlas sin escribirlas.

58 “La rigidez y la complejidad del esquema de contraseñas de una organización es la medida de la pereza e incompetencia técnica del administrador de seguridad informática de esa organización”. *Michael Schrage*.

Cómo preservar mi patrimonio digital personal

Es todavía mejor usar “trucos” para complicar las contraseñas sin perder la nemotecnia; hay quien vuelve mayúsculas sólo las letras de la segunda mitad del alfabeto –de la “n” a la “z”–; así, los ejemplos anteriores se convertirían en: “megONciO90”, “madON-NaeNelSUPeRbOWIXIVI”, “kchUNRaRa240512”, “kRiSZaNcheS*99”, “eLTieMPONOPaSa”, “8*50cadamaRTeS”. O podemos hacer mayúsculas sólo las vocales: “mEgOncIO90”, “mAdOnnAEnElsUpEr-bOwlxlvI”, “kchUnrArA240512”, “krIszAnchES*99”, “EltIEmpOnOPaSa”, “8*50cAdAmArtEs”. Si somos consistentes, nunca se nos olvidará cuál letra va mayúscula y cuál minúscula. No cambies de método cada vez que cambias contraseña: lo peor que te puede pasar es que poseas cinco o seis contraseñas, pero en algunas hiciste mayúsculas la primera mitad del alfabeto, en otras la segunda, y en otra las vocales: es una receta perfecta para el olvido y la confusión. Si vas a cambiar de método cambia todas las contraseñas a ese nuevo método.

Las frases largas pero memorizables son muy recomendables: “se-llegóalunaporprimeravezen1969” o el acrónimo que se construye con la primera letra de cada palabra de esa frase: “slallppve1969”; si a esta técnica le agregas lo de las mayúsculas del párrafo anterior quedará una contraseña muy sólida, de longitud adecuada, pero aún nemotécnica: “SlallPPVe1969”. Esta contraseña consta de sólo tres elementos: frase, números y mayúsculas n-z, pero el resultado es de longitud adecuada, bastante fuerte y para nada deducible por extraños; además, muy memorizable.

No seas predecible: si hoy seleccionaste la contraseña “Megoncio890”, por favor, que la próxima no sea “Megoncio891”; si seleccionaste “melchoR255” que las siguientes no sean “gaspaR256” ni “baltasaR257”. Cambia la técnica de selección cada vez. Lo importante es que te ayude a recordar la nueva contraseña.

Si no puedes memorizarlas, aun con estas sugerencias, una técnica adicional de suma utilidad consiste en escribir todas nuestras contraseñas juntas en un único archivo de texto en la computadora –así siempre estarán a la mano–, archivo que siempre encriptaremos después de leerlo. Si usas un programa robusto y confiable para tal propósito –por ejemplo, EncryptOnClick– la

seguridad del archivo resultante ya cifrado es muy alta, y deberemos recordar sólo una única contraseña con la que se encripta y desencripta el archivo. Aunque alguien acceda a nuestra computadora le será prácticamente imposible desencriptar ese archivo y será por tanto seguro en ese aspecto. Esta técnica impone que como usuario debes ser muy disciplinado en volver a encriptar siempre tu archivo después de haberlo desencriptado y leído para consultar una contraseña. Si no lo haces así estarás dejando un bonito archivo con todas tus contraseñas totalmente expuestas y a la vista en tu computadora y habrá resultado más nocivo el remedio que la enfermedad. Atención, si no recuerdas la contraseña de desencriptado de ese archivo no habrá poder humano que permita descifrarlo, dada su alta seguridad. Pero la ventaja enorme de este método es que sólo hay que construir y recordar una única contraseña.

Además, este tipo de programas no sólo sirven para encriptar archivos de contraseñas: pueden cifrar cualquier archivo que consideres que contiene información sensible dentro de tu computadora, sobre todo los que contengan información personal o confidencial.

Existe también el mito de que las contraseñas deben ser cambiadas cada cierto número fijo de días 30, 60, 90, etc.; algunas organizaciones inclusive obligan a los usuarios a hacerlo. Está demostrado⁵⁹ que esto no ayuda a la seguridad. Mientras la contraseña no se haya comprometido, no es necesario cambiarla cada periodo fijo. Por supuesto que es bueno cambiarlas de cuando en cuando, pero es mejor hacerlo aleatoriamente o cuando hay una justificación; por ejemplo, después de haberla utilizado en un lugar público, como un café Internet o en redes inalámbricas públicas de baja seguridad.

Bruce Schneier resume todo lo anterior en una breve pero cierta frase: “La noción actual de contraseñas se basa en un oxímoron. La idea es tener una cadena aleatoria que sea fácil de recordar. Por

59 Ari Juels, científico en jefe de RSA, la división de seguridad de EMC. “*Password Expiration: Like Margarine and Water?*” [en línea].

Cómo preservar mi patrimonio digital personal

desgracia, si es fácil de recordar, algo como ‘Susan’, no es aleatorio; y si es aleatorio, como ‘r7U2*QnP’, entonces no es fácil de recordar.”

Respecto a los correos electrónicos

- ❖ Recuerda que cuando comienzas a escribir el nombre del destinatario, la mayoría de manejadores de correo electrónico te sugerirá direcciones similares a la que estás tecleando que ya se hayan usado antes. Si has enviado correos electrónicos a varias personas cuyo nombre o dirección comienzan de la misma manera –por ejemplo, “Pedro”– la función de autocompletar puede traer varios Pedros a la pantalla. Es muy común dar “click” a media sugerencia seleccionando otro destinatario con lo cual enviarás ese correo a una persona para la que no estaba destinada. Asegúrate de elegir la dirección correcta antes de hacer la selección.
- ❖ En caso de que el contenido de tu correo sea altamente confidencial, considera si vale la pena entonces que ese contenido vaya encriptado o protegido con una contraseña. Haz esto sólo cuando sea realmente necesario. A nadie le gusta recibir correos cifrados avisando que el próximo lunes es feriado en la oficina.
- ❖ Cuando uses una lista de correo grupal asegúrate de quién está en esa lista. Comprueba que en efecto deseas mandar ese mensaje a todos los integrantes de esa lista.
- ❖ Si deseas enviar un correo electrónico a un destinatario sin revelar su dirección a otros destinatarios del mismo correo, asegúrate de utilizar la opción de copia oculta –Bcc– y no copia simple –Cc–. Cuando se utiliza esta última cada receptor del mensaje podrá ver la dirección de todos los destinatarios a los que fue enviado ese correo.
- ❖ No mandes mensajes de “desuscribir” –*unsubscribe*– a las direcciones que mandan *spam*; sólo les estarás confirmando que el tuyo es un buzón de correo válido. Mejor anota la dirección de la que provienen en el filtro *antispam* de tu sistema operativo o manejador de correo para bloquearla.
- ❖ Dependiendo del contexto, no siempre es bueno usar tu cuenta de correo electrónico del trabajo para asuntos personales y vice-

versa. A veces es bueno considerar tener una cuenta de correo diferente para ambos mundos.

Respecto a las redes

- ❖ Si te conectas por vía inalámbrica Wi-Fi verifica el tipo de conexión a la que accedes: existen las redes totalmente abiertas, y con acceso de seguridad WEP, WPA, IPSEC y WPA2. La primera –como su nombre lo indica– es totalmente insegura por ser totalmente abierta; la WEP tiene una seguridad mínima, la WPA seguridad regular, la IPSEC y la WPA2 algo mayores. Ninguna es de alta seguridad, por lo que debe evitarse en lo posible enviar información valiosa como contraseñas cruciales, números de tarjetas, *nips*, etc., a través de ellas. Las contraseñas cotidianas como la del correo electrónico, biblioteca de la escuela, etc., deberán ser cambiadas con frecuencia dependiendo del nivel de la red a la que has accedido; si entras en red abierta o WEP, cámbialas muy frecuentemente.
- ❖ En algunos sitios web solicitan tu correo electrónico para darte alguna información a cambio de una descarga. Si no te van a enviar alguna respuesta importante vía el correo electrónico o no te interesa conservar la relación con ese sitio, usa para ello una clave de correo desechable obtenida en sitios gratuitos que se obtienen buscando *anonymous email account* en el navegador, y no tu verdadera cuenta.
- ❖ Administra tus *cookies*; estos son pequeños archivos que los sitios web instalan en tu computadora para almacenar información con respecto a tu perfil, así cuando regreses ahí te pueden ofrecer cosas que se supone te interesarán. En su inmensa mayoría son inofensivos y útiles para el usuario, pero como siempre, puede haber personas maliciosas que los usen con fines dañinos. Para evitarlo pueden seguirse ciertas reglas:
 - Configura tu navegador para que siempre te avise cuando un sitio web pretende instalar una *cookie*. Así, siempre puedes decidir si confías o no en ese sitio para hacerlo. No es conveniente dejarlas bloqueadas automáticamente, ya que muchos

Cómo preservar mi patrimonio digital personal

sitios web legítimos que son de tu interés no funcionarán si las bloqueas por completo.

- La mayoría de los navegadores te permiten habilitar y deshabilitar las *cookies* sitio por sitio para que permitas su uso en los sitios que te inspiran confianza.
- Utiliza un programa antiespía el cual detecta y permite eliminar las *cookies* “rastreadoras”. Existen también programas de administración de *cookies* con los que se pueden eliminar las más antiguas e inútiles y ayudan a controlarlas.
- Usa de vez en cuando las opciones de tu navegador respecto a *cookies*; con ellas puedes borrar de tiempo en tiempo algunos de los archivos no deseados o hacer un “borrado general” de las *cookies* de vez en vez. Recuerda que al hacer esto último los sitios legítimos a los que acudes no recordarán tus perfiles anteriores de búsqueda. Si esto no es grave para ti, hazlo con cierta frecuencia.
- Lo mismo de las *cookies* puede aplicarse al historial de los sitios web que has visitado; gracias a esta historia, cuando empiezas a teclear la dirección URL de un sitio web en tu navegador éste te ofrece todas las posibles coincidencias para poder seleccionar con un “click” y no tener que teclear todo. Puedes conservar los elementos de la historia, borrar algunos o borrar todos con la opción *clear history* en tu navegador. Huelga decir que el borrado de toda la historia como el de todas las *cookies* debe hacerse siempre al salir de una sesión web en un café Internet.

Respecto a las redes sociales

- ❖ Revisa las políticas de privacidad de los sitios a los que proporcionas datos personales. La mayoría de los servicios de redes sociales del mundo tienen pautas específicas de privacidad que se publican en sus sitios web. En México fue promulgada desde abril del 2010 la *Ley Federal de Protección de Datos en Posesión de Particulares*⁶⁰ la cual obliga a las personas y organizaciones

60 *Ley Federal de Protección de Datos en Posesión de Particulares* [en línea].

privadas en México que coleccionan datos personales a hacer explícitas las políticas bajo las cuales serán transferidos a terceros. Asegúrate de leer y comprender las condiciones estipuladas en estos sitios, ya que entre ellas se incluye la información que se revelará, vendida o intercambiada, a terceros –*spammers* inclusive–. Si no te convencen los términos de privacidad, no utilices el servicio o, si decides hacerlo, hazlo con suma cautela y parsimonia.

Un ejemplo de ello –el cual muchísimas personas usan sin estar al tanto– es Facebook; este sitio no tiene una política de privacidad de datos tiene una política de uso de datos y, si bien ha mejorado en los años recientes, tiene una serie de defectos; entre ellos están: la afirmación de esta empresa de que las *apps* o aplicaciones de terceros sólo tendrían acceso a la información personal necesaria para hacer su función cuando en realidad tenían acceso a casi todos los datos personales de los usuarios; la aseveración de que los textos, fotos y videos de cuentas borradas no podrían ser ya accedidos, lo que no era cierto; la declaración de Facebook de que no compartiría información personal con asesores, lo que de todas formas hizo, y la afirmación de la empresa de que cumplía el *US-EU Safe Harbor Framework*, normatividad para la transferencia de datos entre Europa y Estados Unidos, cuando en realidad no era así. Y hasta la fecha, todo el contenido que un usuario suba al sitio –texto, foto, video– puede ser utilizado por esta empresa para lo que ellos quieran al igual que el usuario, ya que por el simple hecho de subirlo el usuario cede derechos de uso a esa empresa. Algo semejante pasa con Google Chrome.

Esta política de privacidad defectuosa no quiere decir que el usuario deba dejar de usar el sitio –no, en lo absoluto–; sólo significa que uno debe ser cauto con lo que suba para no perder privacidad o patrimonio documental que pueda tener valor económico, histórico, etcétera.

- ❖ Sé discreto con los datos que compartes en redes sociales. Nunca escribas nada en una página de perfil, tablero de mensajes, mensajes instantáneos u otro tipo de formulario electrónico en

Cómo preservar mi patrimonio digital personal

línea que se exhiba a visitantes no deseados o abra la posibilidad de robo de identidad o amenazas maliciosas. Esto incluye los nombres personales y de negocios, direcciones, números de teléfono, cargos, fechas de nacimiento, aniversario, etc.; detalles de horarios, rutinas diarias, información muy personal tuya o familiar e información restringida de negocios. Algún día puede ser usada en tu contra.

- ❖ Ten cuidado con el *malware* de los sitios de redes sociales. Según algunos estudios, un muy alto porcentaje de los comentarios generados por usuarios en blogs, tableros de mensajes, mensajes instantáneos, etc., son *spam* o contienen enlaces maliciosos. La mayoría de sitios web más populares del mundo contienen en algún lugar de sus miles de páginas algunas que tienen propósitos malignos, obviamente sin el conocimiento ni anuencia del sitio web central.⁶¹ Para este propósito es que son útiles los programas antivirus, antiespías, etcétera.

Respecto a los dispositivos móviles

En proporción con los servidores de cómputo, computadoras de escritorio, laptops, etc., los dispositivos móviles o portátiles –iPod, iPad, teléfono, Smartphone, Blackberry, tableta, etc.– no han sido objeto de una gran cantidad de ataques contra su seguridad –todavía– debido principalmente a tres razones:

- ❖ La información y procesos realizados a través de ellos no son explotables en términos económicos por los interesados en esas actividades ilícitas. Los móviles están llenos de contactos, direcciones de correos, etc., pero son datos difíciles de capitalizar fácilmente y muchos de ellos pueden ser obtenidos de otras formas más fáciles.
- ❖ Los lugares donde normalmente se descargan esas aplicaciones para móviles, tales como *Android Market*, *Google Play*, *Verizon* o *Blackberry World App* tienen especial cuidado en revisar y

⁶¹ Washington State Department of Information Services, *Security and Social Networks* [en línea].

filtrar previamente las aplicaciones que ahí se ofrecen para que estén exentas de cualquier *malware*.

- ❖ La enorme cantidad de ambientes, plataformas, sistemas operativos, etc., que ha existido hasta la fecha en el ambiente de los móviles implicaría a los atacantes tener que programar múltiples variantes de ese *malware* para que fuese efectivo lo cual implica mucho trabajo adicional para tan poco rédito.

Nótese que en el párrafo anterior dice “todavía”; eso significa que el hecho de que hasta hoy no haya habido muchos ataques sistemáticos contra este tipo de dispositivos no pueda verse incrementado significativamente en un futuro. Según recientes estudios, desde el año 2010 el número de dispositivos móviles vendidos a nivel mundial rebasa ya el número de computadoras de escritorio y laptops. De hecho, en la medida en que se incremente la cantidad de procesos realizados por personas por medio de sus móviles que impliquen cuestiones monetarias –como pagos, transferencias o depósitos electrónicos– más atractivo será interferir en esos dispositivos y por tanto seguramente se incrementará el número de ataques e intrusiones a los mismos. Por lo mismo, no hay que esperar a que eso suceda y pueden seguirse desde hoy algunas reglas básicas de seguridad en móviles:

- ❖ Descarga aplicaciones –*apps*– sólo de lugares institucionales conocidos y serios, nunca de sitios personales.
- ❖ Mantén siempre una copia o respaldo externo de la información valiosa que se encuentra en el móvil –agendas, direcciones de correo, documentos de texto, imágenes, etc.– de la misma forma que se harían de tus computadoras de escritorio o laptop, de tal forma que en caso de robo o daño del dispositivo, pérdida de la información contenida en él, etc., te encuentres en posibilidad de rescatarla. En esencia, las medidas de respaldo de información de móviles son prácticamente las mismas que para computadoras de escritorio o portátiles. Las opciones de “sincronización” de los móviles con las computadoras son muy útiles para este propósito. En los móviles que no tienen esta opción, tener guardada en casa una segunda tarjeta de memoria con el respaldo de lo importante es prioritario.

Cómo preservar mi patrimonio digital personal

- ❖ Obtén e instala las últimas versiones o “parches” de las aplicaciones que usas, ya que estas nuevas versiones por lo general detectan y corrigen fallas en la seguridad de la aplicación.
- ❖ Si guardas información sensible en tu móvil, como contraseñas de acceso a sistemas de pago, nips o contraseñas de tarjetas de crédito, etc., guárdalas encriptadas –hay aplicaciones para ello–; o si las guardas de manera explícita hazlo de manera tal que si alguien externo las roba no sea obvio inferir a qué proceso, banco, sitio o actividad pertenecen y por tanto no le sean útiles para obtener un beneficio económico.
- ❖ El dispositivo móvil, igual que la computadora, guarda una serie de datos acerca de su historia y uso. El simple hecho de borrarlos con las opciones básicas del móvil no los elimina realmente, simplemente los quita de la lista de lo que está disponible al usuario, pero físicamente siguen ahí, en la memoria, disponibles para cualquiera que tenga acceso al móvil. Por lo mismo, es recomendable el uso de aplicaciones especializadas, tales como *History eraser* para Android, la cual efectivamente borra historias de llamadas, de búsquedas, mensajes, memoria caché, etc. Existen otras aplicaciones, llamadas en general *Eraser* o “borrador” para otras plataformas, como Windows, iPod, para borrar datos en USB, etc. Si manejas procesos o información sensibles, son muy recomendables.

Como ha podido verse a lo largo de este capítulo, son muchos y muy variados los temas de la seguridad informática; podrían escribirse no tan sólo innumerables capítulos, sino libros al respecto. Empero, el propósito de presentar este tema dentro de esta obra es solamente sensibilizar e introducir al lector a los conceptos y principios básicos de la seguridad informática y brindar algunas recomendaciones fundamentales para trabajar dentro de ella, reiterando que ésta es un ingrediente indispensable de la preservación documental digital.

Como se ha establecido a través del principio básico de la seguridad informática, “La seguridad informática no es un producto que se compra u obtiene, es un proceso siempre continuo”. Esto debe de

tenerse siempre presente: como en muchos otros aspectos de tecnología, la seguridad es 20% productos y 80% procedimientos, método y sentido común. La mayor seguridad no se logra instalando una superbóveda reforzada, sino muchos pequeños recursos de seguridad que se traslapan, se apoyan y se complementan unos a otros: la lista de sugerencias y metodologías presentada anteriormente respecto de la computadora, uso, contraseñas, correos electrónicos, redes y móviles es un buen ejemplo de ello. Si uno la revisa, podrá notar que los recursos tecnológicos a adquirir enunciados en esa lista son mínimos, y en la mayoría de ellos se pueden encontrar buenas alternativas en *software* gratuito o muy económico. Las recomendaciones de método, procedimiento y lógica conforman la mayor parte de ella. Así sucede con el resto de los conceptos de seguridad: seguridad física del equipo, de las redes, etc. Cabe también resaltar que esto es válido para las personas, no así para las organizaciones, las cuales se rigen bajo otros parámetros de inversión, políticas, normatividades, etc.

Los conceptos de seguridad enunciados en este capítulo no son en absoluto una lista exhaustiva de todos los que una persona puede utilizar en materia de seguridad informática, pero sí cubren todo lo esencial acerca del tema pensando en las necesidades de esa persona, –reitero– no de una organización.

De acuerdo con el principio básico de la seguridad informática, la aplicación de estos conceptos no garantiza una seguridad absoluta y por completo impenetrable; si recordamos, el riesgo informático consiste en la probabilidad de que un evento nocivo ocurra combinado con su impacto o efecto nocivo en la información de la persona. Pero al combinar recomendaciones de seguridad informática con recomendaciones de preservación documental digital reduciremos enormemente –a cifras muy manejables– la probabilidad de que un evento desafortunado suceda y aún en caso de suceder lograr que el impacto se reduzca o se elimine al poder recuperar nuestra información íntegra y rápidamente.

La preservación de nuestros archivos digitales

Si inviertes más tiempo en buscar tu café que en respaldar tus archivos, algún día los perderás... es más, mereces perderlos.

Richard Clarke

Antes de entrar finalmente a las recomendaciones para la preservación documental digital conviene hacer un par de reflexiones previas. Ya hemos mencionado la cantidad de información digital que se produce en el mundo en nuestros días. Reflexionemos ahora por un momento en la enorme variedad de información digital que se produce en forma de diversos documentos y de entre la cual nosotros seguramente poseemos algunas de ellas como parte de nuestro patrimonio digital personal. Sin ser una lista exhaustiva, a este respecto podemos pensar en:

- ❖ Libros electrónicos, en todas sus variantes y modalidades: libros propiamente dichos, memorias, antologías, resúmenes, compendios, tesis, atlas, enciclopedias, diccionarios, etcétera.
- ❖ Diarios y revistas electrónicas, en todas sus variantes y modalidades: periódicos, semanarios, e-journals, e-magazines, blogs, etcétera.
- ❖ Otros materiales de texto, tales como: tareas y trabajos escolares, ensayos, reportajes, tesis, reportes, catálogos, listas de discusión, manuales, leyes y reglamentos, enseñanza de idiomas, agendas, directorios, notas, bibliografías, mensajes electrónicos, etcétera.
- ❖ Documentos administrativos y de trámite: oficios, memos, trámites y recibos de pagos, contribuciones o impuestos; correos electrónicos, documentos de seguros, etcétera.
- ❖ Archivos y expedientes de clientes, pacientes o proveedores.
- ❖ Imágenes digitales: como parte de colecciones de fotografías, mapas, pinturas, dibujos, diagramas, bocetos, iluminaciones, car-

Cómo preservar mi patrimonio digital personal

teles, manuscritos, planos, partituras, esculturas, viñetas, patentes, estampillas, etcétera.

- ❖ Audio digital: como parte de colecciones de música, entrevistas, conferencias, ponencias, mesas redondas, debates, enseñanza de lenguas, poesía, audiolibros, teatro, programas radiofónicos, etcétera.
- ❖ Video o cine digital: piezas de cine y video comercial o educativo; ponencias, mesas redondas, entrevistas, entretenimiento, periodismo, instalaciones o *performances*, etcétera.
- ❖ Bancos de datos, tablas y cartas de todo tipo: científicos, económicos, meteorológicos, geográficos, geológicos, sanitarios, estadísticos, astronómicos, jurídicos, antropológicos, educativos, etc. Datos espaciales y geo-espaciales.
- ❖ Documentos y materiales multimedios.
- ❖ Copias de páginas web: de entretenimiento, de comercialización, informativas, educativas, de organizaciones, gubernamentales, de servicios, etcétera.
- ❖ Piezas de *software*, juegos digitales, realidad virtual.

Dada la cantidad y variedad de documentación digital generada en los últimos años y dentro del ámbito institucional y en especial gubernamental se han producido una serie de normatividades, recomendaciones, estándares, etc., que buscan organizar y estructurar los esfuerzos que las organizaciones deben realizar a nivel mundial para tratar de garantizar el acceso actual y futuro a los documentos digitales relevantes. No existe hoy en día un esfuerzo único que establezca o defina los atributos y características ideales de todos los documentos digitales y sus procedimientos asociados con miras a su preservación a largo plazo. En los últimos 15 años se han ido estableciendo una serie de iniciativas, se han planteado modelos y marcos de referencia para tratar de definir o establecer los atributos de documentos, procedimientos, sistemas, organizaciones, etc., que se encuentren involucradas en la gestión y la preservación digital. Cada uno de ellos hace énfasis en elementos o atributos que considera importantes para el establecimiento del modelo en cuestión.

Estos modelos no pueden ser aplicados tal cual para la preservación documental personal por varias razones: en primer lugar porque cada variedad documental tiene sus propias reglas documentales de registro, recuperación y preservación. No son iguales las estructuras, organizaciones y normatividades que rigen por ejemplo a los libros, que a los documentos de archivo gubernamentales o a los videos; proceden de entidades y estructuras muy diferentes y sus modelos varían en muchos elementos. No existe una estructura documental universal que englobe a todos los tipos de documentos que hemos enunciado.

En segundo lugar, la inmensa mayoría de esos modelos y sus especificaciones están hechos para organizaciones y por lo mismo no pueden ser aplicados tal cual a la preservación documental personal, ya que contienen muchísimas especificaciones y estándares demasiado complejos para ser seguidos y observados a nivel de los individuos.

Claro que hay principios básicos, de experiencia y de sentido común que sí pueden ser extraídos de esos modelos y adaptados a las necesidades de las personas. De hecho éste es el objetivo principal de esta obra: extraer de entre todas esas variantes, modelos, iniciativas y estándares una resultante que pueda ser utilizada a nivel de los individuos para preservar adecuadamente su patrimonio documental digital. En caso de que a alguien le interese profundizar en esos modelos o estándares provenientes del campo de la archivística –obviamente para organizaciones– puede referirse al Anexo 6.

Por lo anterior, es necesario realizar en este punto una precisión: las recomendaciones y estrategias presentadas a continuación en esta obra no constituyen en su conjunto ningún estándar o recomendación internacional; no las hallará el lector como tal en ningún sitio o referencia. Pero todas y cada una de ellas provienen de algún estándar o modelo o conjunto de ellos relacionado con alguna variedad documental real –en especial la cadena de preservación– o provienen de un estándar *de facto* y todas han probado su eficacia en este propósito en su ambiente. No son simples recomendaciones teóricas basadas en la lógica: todas son medidas prácticas, útiles y probadas para el propósito que nos interesa a este respecto.

Cómo preservar mi patrimonio digital personal

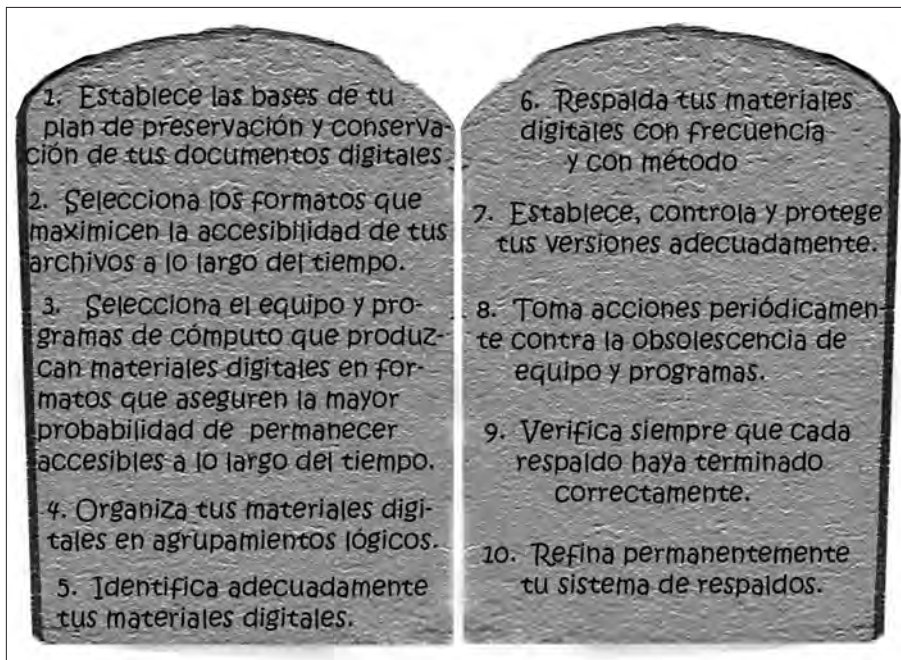
Para su mejor comprensión los temas han sido divididos en diez grandes recomendaciones, cada una de ellas abarcando un cierto rubro o aspecto. Están numeradas y marcadas con el símbolo



Dentro de ellas, hay un cierto número de recomendaciones más puntuales, o tips, cada una de ellas con el símbolo



RECOMENDACIONES PARA LA PRESERVACIÓN DOCUMENTAL DIGITAL





1. Establece las bases de tu plan de preservación y conservación de tus documentos digitales

Es necesario de entrada definir el propósito y alcances de nuestra conservación y preservación digitales para poder diseñar una estrategia al efecto. Hay que distinguir entre los materiales que deseamos mantener por corto o mediano plazo –esto es, los que conservamos– y los que deseamos guardar a largo plazo –los que preservamos–. Quien pretende preservar todo al final no preserva nada. Podemos respaldar todo nuestro material documental digital, podemos conservarlo todo, pero no es conveniente –dado que es imposible– pretender preservar todo para siempre. Por tanto, el primer paso es establecer un plan para la conservación y preservación de nuestros archivos digitales. Recuerda aquí las definiciones y diferencias ya presentadas entre conservación y preservación digital. Para acortar el nombre, denominémoslo simplemente como plan de archivos, conscientes de que en realidad el plan comprende conservación, respaldos y preservación. Para establecer estas bases de este plan podemos efectuar los siguientes pasos:



Realiza un inventario de “tipos” o variedad documentales que posees. Es necesario elaborar un inventario o relación de todos los tipos distintos de materiales que poseemos: libros, fotografías, mapas, partituras, artículos de revistas, textos, etc., partiendo de la lista presentada en el apartado anterior. Atención: debemos hacer una relación de todos los tipos de documentos, no una relación de todos y cada uno de los documentos existentes en nuestras computadoras, los cuales podrían ascender a varias decenas o cientos de miles. Este proceso consiste en establecer una especie de clasificación o tipificación de cuáles son los grupos de documentos más valiosos dentro de mi equipo, en orden decreciente. Por ejemplo, puedo clasificar mis documentos en: 1) expedientes de mis pacientes, 2) textos de trabajo, 3) mis fotografías, 4) mi co-

lección musical, 5) correos electrónicos, etc. Por supuesto para cada persona variará el tipo, importancia y prelación de sus documentos. La lista de tipos de materiales digitales que aparece al principio de este capítulo puede ayudarte en este propósito.



Establece alcance y objetivos. No todos los grupos documentales que posees tienen el mismo propósito. Es necesario definir de entrada los objetos de conservación o respaldo –de corto o mediano plazo– y los objetos de preservación a largo plazo. Por ejemplo las tareas escolares: si tenemos varias versiones de tareas y trabajos escolares, es conveniente conservar esas versiones durante un tiempo razonable, ya que podríamos necesitar alguna de las versiones intermedias en lapsos cercanos, pero no tiene ningún caso pretender preservar todas y cada una de esas versiones intermedias para la posteridad. A esas versiones intermedias las podemos denominar como objetos de respaldo de corto plazo. Si consultamos esos archivos, digamos, dentro de tres años, no recordaremos –ni nos serán útiles– las diferencias entre cada una de esas versiones. Sin duda será recomendable preservar la versión final de esa tarea por un plazo mediano, digamos, la duración de la carrera profesional. Este sería un objeto de conservación o respaldo de mediano plazo. Después de ello, es probable que sólo deseemos guardar algunas tareas o trabajos que fueron especialmente sobresalientes o representativos y, muy probablemente, la tesis. Estos son los objetos de preservación de largo plazo.

Si escribimos artículos o reportes científicos, debemos decidir de inicio si queremos preservar sólo los reportes o documentos finales de los proyectos, o también las versiones intermedias o incluir los datos primarios. En artes, puede desearse preservar sólo grabaciones o imágenes de las obras; o puede desearse guardar los elementos individuales que las conforman. Las bases de datos no pueden preservarse como tales dadas sus estructuras intrínsecas; es necesario definir previamente instantes periódicos en los que se toma una fotografía instantánea de una base de datos en una forma adecuada y pertinente y así lograr su preservación. ¿Cada cuando

se guarda esa instantánea? ¿Cada semana, mes, año? ¿Qué formato de exportación/importación debe tener?

Si tenemos expedientes de pacientes o clientes ¿cuánto tiempo debemos conservarlos? La respuesta puede ser: mientras sean nuestro pacientes o clientes más un cierto periodo; por un cierto número fijo de años, o para siempre. Esto debe pensarse y decidirse al momento de estar haciendo el plan de archivos.

Por lo anterior nos es útil esa lista o inventario de tipos de documentos que poseemos a partir de la primera recomendación, ya que con ayuda de ella, estamos en posibilidad de establecer para cada tipo o variedad de documentos cuáles de ellos se deben conservar por periodos finitos, ya sea cortos o medianos, y cuáles de ellos deben manejarse como documentos a preservar a largo plazo. O dentro de un mismo tipo de material, cuáles versiones o variedades nos interesa conservar por periodos cortos o medianos, y cuáles nos interesa preservar a largo plazo.

Debido a las dificultades técnicas propias de la preservación digital, la determinación de cuáles documentos deben ser sujetos de preservación debe hacerse lo más pronto posible, ya que esto incide en retroalimentación de procedimientos y a las estrategias de conservación y preservación, lo cual mejora las probabilidades de éxito del esfuerzo.



Define los recursos de conservación y preservación.

Como ya hemos mencionado, la componente tecnológica no debe ser el núcleo de nuestros proyectos de preservación documental digital, pero obviamente hay una cierta cantidad de tecnología que es indispensable. Todo proyecto de preservación requiere de ciertos recursos tecnológicos. Es necesario al inicio de un proyecto de preservación definir cuáles son los equipos y dispositivos que deben ser dedicados a la conservación y preservación documental –discos duros fijos o portátiles, memorias USB, tarjetas SM o SD, CD-R o DVD-R, etc. En este aspecto, es conveniente tener en mente que pueden conseguirse nuevos recursos, o reubicarse recursos ya existentes. Abundaré en este punto más adelante.



Diseña e implementa tus estrategias de seguridad informática.

La seguridad informática es como poner una cerradura en tu casa o en tu auto; finalmente no detendría a los ladrones, pero si es lo suficientemente buena los hará decidir moverse a otro blanco más fácil.

Paul Herbka

Como ya establecimos en el apartado de seguridad informática, para una adecuada preservación de los documentos digitales debemos tener un mínimo de seguridad establecida en nuestros sistemas e información. Muchas veces se invierte mucho esfuerzo en el desarrollo de las estrategias para la preservación, pero poco se hace con respecto a las estrategias para la seguridad informática de los documentos. Recordemos que preservación, conservación y seguridad informática no son sinónimos.⁶² Ya hemos establecido los conceptos y recomendaciones básicos a este respecto en el capítulo “Seguridad informática”.



Identifica todos los componentes digitales para cada tipo de documento. A diferencia del papel, el cual siempre es un componente identificable como un todo, los registros de documentos digitales pueden estar formados por varias capas, componentes, o elementos externos que deben ser integrados para reproducir el documento original. Para cada tipo de documento debe verificarse que se han identificado todos los elementos que componen un documento de archivo digital y que son transferidos integralmente a los respaldos. De poco sirve preservar un texto html si después descubrimos que las imágenes que correspondían a esa página estaban en otra carpeta que no fue protegida junto con el texto. Si usas fuentes tipográficas especializadas fuera de las comunes: Arial, Courier, Thorndale, etc., asegúrate de preservarlas junto con los documentos; de otra forma no existirán en un futuro y tus textos serán accesibles, pero con otro tipo de fuente. Estas fuentes no se encuentran en tus carpetas de datos; se encuentran

⁶² Juan Voutssas M., “Preservación documental digital y seguridad informática”.

por lo general en carpetas del sistema o system files y casi todo el mundo se olvida de preservarlas junto con los documentos.



Determina el valor de uso y la calidad suficiente de los diferentes tipos de documentos para preservación. Durante la definición de tipos o categorías de los documentos digitales en tu equipo es necesario establecer los valores de uso; es decir, límites y características de la usabilidad y aceptabilidad de los documentos de archivo para poder mantener a lo largo del tiempo las características de calidad suficiente, mínima o aceptable que consideras indispensable de lograr en cada tipo de documento para considerarlo aceptable y usable, tal como se estableció con detalle en el apartado “Calidad y valor de uso”. Esto permitirá a su vez definir o redefinir elementos tales como: resolución en escaneados, paletas de color, tasas de bits en archivos de sonido, compresión, etc., para cada tipo de documento. Recuerda que puedes tener más de un valor de uso para cierto documento lo cual lleva a tener diversas copias del mismo con diversas calidades: alta resolución para preservar, además de mediana o baja para compartir o publicar en la red. Esto aplica para textos en PDF, imágenes, archivos de sonido, video, etc. La recomendación adicional aquí es no revolver las copias de alta calidad con las otras; es conveniente separarlas en carpetas diferentes.



Establece los metadatos mínimos y los deseables para los diferentes tipos de documentos para preservación. Durante la definición de tipos o categorías de los documentos digitales existentes en tu equipo es deseable ir definiendo los metadatos mínimos y los deseables para cada tipo de documento. Atención; esto no significa que en este momento deban agregarse a todos y cada uno de los documentos esos metadatos, sino que debemos establecer solamente cuáles deben ir en cada tipo de documento y descubrir cómo se agregan los metadatos a ese tipo de archivo. Por ejemplo, puede establecerse que para los documentos de texto se consigne autor o autores, título del documento y versión. Para archivos de sonido se puede decidir que se desea consig-

nar como mínimo intérprete y nombre de la pieza. Por cierto, para agregar metadatos a este tipo de archivo se puede hacer de distintas formas, dependiendo del sistema operativo; por ejemplo, en Windows, si apuntamos en el directorio a un archivo de música mp3 y seleccionamos con el botón derecho del ratón la opción “Propiedades” en la ventana que se abre, después la pestaña “Resumen” y el botón “Opciones avanzadas” veremos que se abre una ventana de diálogo en la cual se pueden consignar los metadatos para esa pieza, tales como nombre del intérprete, nombre de la pieza, género, título del álbum, la letra de la pieza musical, y algún otro comentario deseado. Recuerda que en este punto estás decidiendo cuáles metadatos serán agregados en un futuro a cada tipo de material, y no que debas registrar ahora esos metadatos en todos y cada uno de los archivos.

En este punto es conveniente en ir pensando en algún método que te acomode para describir/catalogar/registrar tus documentos. No tiene que ser implementado de inmediato, pero conforme el orden vaya llegando a tus documentos digitales, se hará cada vez más patente la necesidad de crear un sistema o método para lograr esto.



Identifica aspectos legales de la preservación. Es necesario identificar y establecer previamente las implicaciones legales del almacenamiento y distribución de la información.

Debe quedar muy claro que para almacenar información en nuestro poder –sea de creación propia o de terceros– no requerimos permiso; para publicarla y redistribuirla puede ser que sí, de acuerdo a las leyes de propiedad intelectual. Recuerda que es muy útil guardar en los metadatos de un cierto documento que hemos tomado de la red su declaración de Creative Commons u otra de distribución compartida legal para tener ese dato embebido dentro del documento. De otra manera no recordaremos en un futuro de dónde lo tomamos, y cuál es su licencia de uso legítimo. Si manejamos datos de terceros tales como expedientes médicos de pacientes, listas de seguros, etc., que incluyan datos personales: nombre dirección, teléfono, etc., debemos cumplir en México con ciertas disposiciones de la Ley Federal de Protección de Datos en Posesión de

Particulares⁶³ las cuales implican ciertas reglas mínimas que deben observar las personas o empresas que manejan datos personales de terceros para garantizar un uso adecuado y discreto de esa información. La ley se basa en principios de licitud, consentimiento, calidad, información, proporcionalidad y responsabilidad.

Con todo lo anterior ya estamos listos para desarrollar procedimientos para conservación y preservación ¿Para qué necesitamos ir desarrollando procedimientos?



Contrario a lo que uno puede pensar de inicio, la preservación se logra, sobre todo, gracias a métodos y procedimientos para hacer las cosas y sólo en menor grado gracias a dispositivos o accesorios tecnológicos. Tienes que estar dispuesto a ir desarrollando procedimientos o rutinas que facilitarán y darán método a tus estrategias de seguridad informática y de preservación documental digital. Los procedimientos a desarrollar tienen bases generales, pero deben ser ajustados por cada persona; cada quien tiene su estilo, frecuencia, tipología, etc. Los siguientes puntos te ayudarán a ir creando y desarrollando tus propios procedimientos.



2. Selecciona los formatos que maximicen la accesibilidad de tus archivos a lo largo del tiempo

Siempre que sea posible, prefiere formatos de amplio uso y aceptación, independientes de la plataforma, con compresión sin pérdida, con especificaciones ampliamente disponibles y en lo posible no propietarios. A este tipo de formatos de archivo se les conoce también como formatos abiertos –open formats–, lo cual significa que sus especificaciones de construcción han sido publicadas y están ampliamente disponibles y por tanto se les considera no

63 *Ley Federal de Protección de Datos en Posesión de Particulares* [en línea].

Cómo preservar mi patrimonio digital personal

propietarios. En ciertos casos esto implica también que están libre de patentes y de pago de derechos por su uso y que están ampliamente difundidos y aceptados. Por lo general estos formatos están disponibles para ser usados desde varias plataformas de cómputo; es decir, son formatos multiplataforma. Nótese que los formatos abiertos no necesariamente tienen que ser elaborados por aquellas organizaciones que producen programas de fuente abierta o libre –en los que los códigos fuente de los programas están disponibles para hacer cambios–.

Por lo mismo, debe evitarse en lo posible el uso de formatos altamente propietarios y por ende poco interoperables. A la capacidad que tienen equipos y programas para intercambiar datos y utilizarlos con otros equipos y programas sin mayores ajustes o cambios especiales se le llama interoperabilidad. A mayor interoperabilidad, mayor probabilidad de que los materiales digitales permanezcan accesibles a lo largo del tiempo. Por ejemplo, algunas cámaras fotográficas tienen la opción de guardar las imágenes tomadas en un formato propio del fabricante. Estos formatos deben ser evitados a toda costa desde un principio, ya que no cumplen con casi ninguna de las condiciones de interoperabilidad y preservación, y muy probablemente desaparecerán en un plazo corto. Con esta salvedad, para el uso cotidiano de documentos efímeros que se pretenden conservar a corto plazo puedes usar casi cualquier tipo de formato que te sea conveniente, pero para documentos que pretendas conservar a largo plazo es altamente recomendable utilizar formatos estables y que están reconocidos para preservación.

Por ejemplo, tú puedes crear y utilizar cotidianamente documentos de texto en formato docx si así te conviene, y guardar esos documentos en ese formato por plazos cortos o medianos, pero por ninguna causa se recomienda guardar la copia para preservación a largo plazo en ese formato, ya que no cumple con las premisas estipuladas para este propósito. La copia al efecto debería ser simplemente en formato doc o mejor aún, en formato odt o Microsoft Open XML Document. Para los archivos tipo PDF para uso cotidiano y a corto/mediano plazo, el PDF normal o simple es suficiente; para archivos

PDF con miras a preservación a largo plazo, se recomienda el uso de la versión PDF/A y sus variantes PDF/A-1a, PDF/A-1b, los cuales han sido establecidos en el estándar internacional *ISO 19005-1. Document management. Electronic document file format for long-term preservation. Part 1: Use of PDF 1.4 (PDF/A)*.⁶⁴ De preferencia, no apliques técnicas de compresión a tus materiales digitales; si decides hacerlo, aplica técnicas de compresión avaladas internacionalmente, como las denominadas técnicas de compresión sin pérdida, que ya hemos explicado en el apartado “Compresión de imágenes”.

Para archivos de imágenes, la recomendación es similar: puede usarse cualquier formato para manejo de archivos cotidianos cuya existencia está destinada al corto o mediano plazo: jpg, gif, png, etc.; para preservación a largo plazo, se deben seleccionar formatos al efecto, tales como tiff, aiff, jpeg2000 o png.

Lo mismo se puede decir de archivos de video: para archivos de uso cotidiano que no se pretenda conservar por periodos largos, pueden seleccionarse múltiples formatos; pero para preservación a largo plazo hay ciertos formatos que cumplen con los principio ya establecidos: por ejemplo, mov, mpeg2, mjpeg2000, mxf, mpeg-1-2, MP4/H.264.

A continuación se presenta una pequeña tabla con los formatos más interoperables para los tipos de documentos más comunes. Los formatos de la columna derecha son considerados de los más interoperables y por tanto ideales para preservación a largo plazo. Los de la columna del centro pueden ser usados para conservación a corto/mediano plazo. En caso de desear preservar documentos a largo plazo es necesario producir los documentos en los formatos ideales o en su caso convertirlos a ellos.

Tipo de documento	Formatos para conservación a corto/ mediano plazo	Formatos interoperables ideales para preservación a largo plazo
Imagen rasterizada	TGA, RAW, BMP, PSD, FPX, PCD, PCT, GIF	TIFF, JPEG (pérdida moderada), JPEG2000 (sin pérdida), PNG

⁶⁴ Library of Congress, *Sustainability of Digital Formats* [en línea].

Cómo preservar mi patrimonio digital personal

Tipo de documento	Formatos para conservación a corto/ mediano plazo	Formatos interoperables ideales para preservación a largo plazo
Imagen vectorial	AI (Adobe Illustrator), WMF (Windows Metafile)	SVG (Scalable Vector Graphics File)
Texto	DOC, DOCX	ODT, PDF, PDF/A, TXT, RTF, XML, DOCX (versión Office Open XML)*
Texto	PS (Postscript), EPS (Encapsulated Postscript)	PDF
Audio	WMA, RA, RM, SND (Apple Sound File), AU	WAV, AIFF, MP3, MIDI, XMF, OGG, FLAC
Vídeo	SWF, FLV (Flash Video), WMV, RV / RM (Real Video)	MP4/H.264 , MPEG-1/2, MOV, MJ2, DV, MXF (sin compresión)
Hoja de cálculo	XLS	ODS, XLSX (versión Office Open XML)*
Presentación	PPT	ODP, PPTX (versión Office Open XML)*
Correo	EML, PST, EUDORA, etc.	MBOX (Mailbox File)

* La versión Office Open XML ideal para preservación es la establecida por las normas ISO/IEC 29500:2008, "Information technology-Office Open XML formats" y ECMA-376 "Office Open XML File formats" segunda edición. No debe confundirse con OpenOffice.org (formato obsoleto usado por versiones anteriores de OpenOffice.org).



3. Selecciona el equipo y programas de cómputo que puedan producir los materiales digitales en los formatos que aseguren la mayor probabilidad de permanecer accesibles a lo largo del tiempo

Para preservación a largo plazo, prefiere en lo posible equipos y programas que lean y produzcan materiales digitales en formatos estándares, que sean interoperables y multiplataforma y que permitan por tanto intercambiar documentos

con otros. Derivado del punto anterior, siempre que adquieras o cambies programas o equipos separa aquellos que te permitan mantener los formatos que ya has seleccionado previamente para uso cotidiano de aquellos para preservación. El acceso futuro y continuo a los documentos digitales dependerá en buena medida de contar con los equipos y programas de cómputo adecuados: aquellos equipos y programas que no son compatibles con versiones previas o futuras –compatibilidad retrospectiva– dificultarán a la larga el acceso de los documentos digitales. Como ya mencionamos, se llama interoperabilidad a la capacidad que tienen equipos y programas para intercambiar datos y utilizarlos con otros equipos y programas sin mayores ajustes o cambios especiales. A mayor interoperabilidad, mayor probabilidad de que los materiales digitales permanezcan accesibles a lo largo del tiempo.

Un punto muy importante a remarcar es que en muchos casos equipos y programas propietarios pueden producir documentos en formatos abiertos y no forzosamente los documentos producidos por ese equipo o programa estarán en formato propietario. Lo opuesto también ocurre: un programa o sistema abierto no va a producir en automático documentos en formato abierto; también puede producirlos en formatos propietarios. Es por ello que la selección de formatos que hicimos desde el punto anterior se vuelve pieza fundamental en el proceso de preservación. Si el formato seleccionado para preservar es estándar y abierto, no importa con qué tipo de sistema o equipo haya sido producido: propietario o abierto. Como ejemplo de esto tenemos a los procesadores de texto. Los de fuente abierta como OpenOffice pueden producir y guardar textos tanto en formato odt –totalmente abierto y no propietario– como en formato doc, propietario. El programa de edición gráfica Photoshop, siendo propietario, puede producir imágenes en formatos tif, jpg o png –estándares abiertos para preservación–, así como imágenes en formato gif propietario. Como puede verse, la naturaleza abierta o comercial del programa no forzosamente define la naturaleza del formato que producimos con él; es nuestra elección deliberada y por ello es importante estar consciente de ella. Lo que es importante es que el equipo y programas actuales y

Cómo preservar mi patrimonio digital personal

futuros puedan continuar produciendo y manejando los formatos que hemos establecido como ideales para largo plazo.

Hay ciertos elementos que nos permiten seleccionar estos equipos y programas de acuerdo con esta premisa. Nótese que es prácticamente imposible lograr siempre que todo nuestro equipo y todos nuestros programas sean abiertos y estándares. Recordemos que nuestros equipos y programas en la actualidad abarcan una variedad muy amplia de fabricantes, dispositivos, marcas, versiones y preferencias nuestras, y esto involucra muchas veces computadores de escritorio, portátiles, *netbooks*, tabletas *-pads-*, teléfonos inteligentes *-smartphones-*, agendas electrónicas, lectores de e-libros *-e-readers-*, etc. Conscientes de que es imposible lograr esta premisa en forma absoluta, en la medida en que nos acerquemos a ella estaremos facilitando el camino de la preservación. Si bien desde el punto de vista práctico es imposible que todos nuestros equipos y programas sean abiertos y estándares, tener una amplísima variedad de equipos y programas de todo tipo de fabricantes y estándares sin ningún concierto a la larga facilitará la confusión y pérdida de información.

Por ejemplo, si tienes un *e-reader* o dispositivo lector de libros electrónicos de la marca Kindle debes estar consciente de que es una preferencia tuya para leer esos documentos, ya sea porque te acomoda, porque te parece práctico, económico o por cualquier otra razón; eso está bien. Pero debes estar consciente que esos libros en formato Kindle no podrán ser preservados a largo plazo: dependen estrechamente del dispositivo el cual es totalmente propietario; sus programas para usarlo son totalmente propietarios. Lo más probable es que en unos pocos años el formato desaparezca, o las nuevas versiones del *e-reader* no puedan leer las versiones anteriores de esos textos, o muchos riesgos más. Son por tanto copias efímeras de ese libro, y la probabilidad de preservarlas a largo plazo es casi nula. Es un equipo con muy poca o nula interoperabilidad.

Si deseamos preservar un libro electrónico que poseemos en la actualidad es necesario producirlo y guardarlo en un formato que maximice esa probabilidad y que pueda ser leído por una variedad

más amplia de dispositivos, actuales y futuros. Por ejemplo, un formato PDF. Este formato, así como los equipos y programas que pueden leerlo nos brindan una perspectiva de mayor probabilidad para poder preservar nuestro libro electrónico a largo plazo. Si lo guardamos en un formato PDF/1-A, el especial para preservación a largo plazo, mejor todavía. Por el contrario, si todos tus libros digitales están guardados en formato Kindle dentro de un dispositivo de esta marca y de la generación actual –la cuarta– desde hoy y en un futuro tanto formato como equipo quedan indisolublemente ligados, y para poder accederlos será necesario tener un aparato de esa marca y de esa generación.

Por tanto, seleccionar un dispositivo Kindle para lectura de libros estando consciente de su efímera duración –simplemente porque el aparato es cómodo y por el placer de leerlos– es una decisión correcta y práctica, pero esperar que nuestra colección de libros en este formato pueda ser preservada a largo plazo no parece ya una decisión adecuada. Los mencionados elementos que pueden ayudarnos a la hora de la selección son:



Para preservación, prefiere equipo y programas de cómputo estándares. Esto es una de las mejores cosas que puedes hacer para que tus materiales digitales duren. Los estándares avalados por organismos nacionales o internacionales son los mejores; a estos se les conoce como estándares *de jure*. Por ejemplo, los equipos que afirmen que tienen comunicación inalámbrica Wi-Fi deben cumplir con el estándar mundial IEEE 802.11 en alguna de sus variantes “a”, “b”, “g”, o “n”.⁶⁵ El sello Wi-Fi en un equipo implica una certificación que otorga la organización Wireless Ethernet Compatibility Alliance o simplemente Wi-Fi Alliance cuando este equipo cumple con ese estándar *de jure*, independientemente de que el

65 *IEEE. Institute of Electrical and Electronics Engineers.* (Se pronuncia “I triple E” en Hispanoamérica e “I e-cubo” en España). Organismo mundial que emite innumerables estándares en ingeniería; entre ellos los de telecomunicaciones. En particular, el estándar 802.11 consiste en un estándar para telecomunicaciones inalámbricas que permite interoperar a equipos muy diversos entre sí en entornos públicos, de negocios y de usuarios residenciales.

equipo sea computadora, teléfono, consola de videojuegos o hasta un refrigerador que se conecta inalámbricamente a la red.⁶⁶ Si este tipo de estándares no existiese para tus materiales, puedes ayudar a asegurar la longevidad adoptando programas ampliamente reconocidos como interoperables y que por lo mismo son muy aceptados y utilizados. A estos estándares se les conoce como estándares *de facto*; por ejemplo, programas de procesamiento de texto que pueden producir documentos en formatos PDF para Adobe Reader o programas editores de imágenes que pueden guardarlas en tiff, jpeg, jpeg2000 o png. Los programas de fuente abierta –*open source*– que son generalmente no propietarios son altamente recomendados.



Selecciona aquellos equipos y programas que desplieguen los materiales tal como aparecieron originalmente.

Idealmente, los materiales digitales deben conservar su apariencia exacta a lo largo del tiempo para ser totalmente inteligibles y accesibles. Asegúrate de que los nuevos equipos y programas sean capaces de leer tus materiales anteriores en el formato en el que han sido conservados y de que desplieguen de manera correcta en la pantalla en la misma forma documental en la que se veían originalmente. Por ejemplo, en ciertos procesadores de texto sucede a veces que las nuevas versiones no despliegan adecuadamente archivos creados con versiones anteriores del mismo. En el caso de archivos de video o sonido estos deben ser reproducidos igual o mejor que con los programas y equipos anteriores. En otras palabras, los nuevos equipos y programas de cómputo deben ser compatibles con los anteriores.



Selecciona equipos y programas que te permitan compartir los materiales digitales con facilidad.

Los equipos y programas de cómputo deben ser capaces de poder recibir del exterior y dar salida a los materiales en diferentes formatos; esto es, además de interoperabilidad adecuada, se desea que el

⁶⁶ Pueden verificarse en la página de la “Wi-Fi Alliance” los equipos que cumplen con este estándar en: http://certifications.wi-fi.org/search_products.php

equipo tenga compatibilidad y comunicabilidad con otros. Ello te facilitará acceder a tus materiales además de transferirlos con facilidad hacia otros sistemas y recibir información procedente de otros sistemas. Un ejemplo de esto son los equipos que utilizan iTunes; a muchas personas no les gustan estos equipos ya que tienen muchas restricciones y poca interoperabilidad a diferencia de aquellos basados en mp3, altamente interoperables. Una de las consecuencias más evidentes de esta selección es la facilidad o no de transferir archivos de sonido entre plataformas y dispositivos diferentes.



Conserva todas las especificaciones de equipos y programas. Esta documentación, como los manuales de los equipos, sistemas, programas y otros materiales similares será esencial en un futuro para poder acceder a los materiales digitales o para migrarlos a nuevas plataformas según la tecnología vaya avanzando. Es particularmente importante documentar perfectamente los programas construidos por ti mismo, así como información detallada de todas las modificaciones y ajustes que les realices, y que describas claramente las características y propiedades de los materiales digitales que esos cambios introducen y los objetivos que persigues con esas modificaciones. Una buena práctica es documentar esos cambios como comentarios intercalados en las líneas del código de la programación. Así la información no se perderá ya que forma parte del archivo en sí, y esto será muy útil a aquellos que deban hacer cambios al programa en un futuro según la tecnología avance.

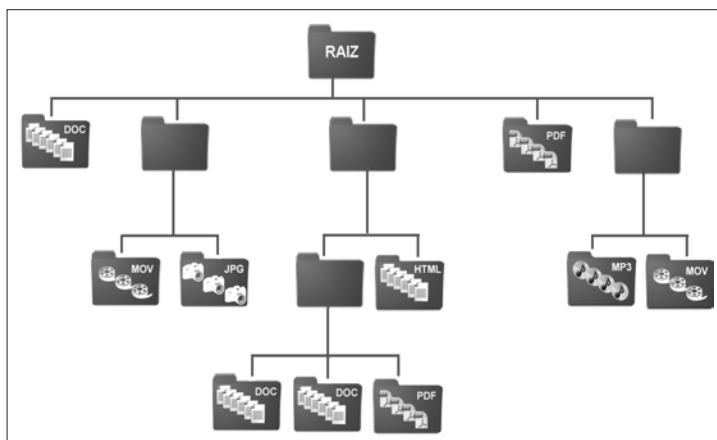


4. Organiza tus materiales digitales en agrupamientos lógicos

Organiza los materiales digitales en agrupamientos lógicos en todos los dispositivos electrónicos que manejes. La administración y recuperación de tus materiales digitales se optimiza notablemente cuando los organizas en conjuntos; de allí la importancia de

Cómo preservar mi patrimonio digital personal

que agrupes tus materiales digitales de alguna manera lógica. La categorización escogida puede ser el reflejo de la manera en que trabajas cotidianamente, o de tus actividades principales, procedimientos, áreas temáticas, tipo de materiales o alguna otra clase de organización estructural que a ti te acomode. Por supuesto también deben existir subdivisiones de la categorización documental. Tales agrupamientos pueden interrelacionarse entre ellos de forma simple o jerarquizada, según mejor convenga a tus necesidades. Como recomendación general, la estructura debería ser consistente con la que tú usarías en el caso de archivos en papel, colecciones de discos o de fotografías, u otros medios semejantes, de tal forma que los documentos del mismo tipo o relacionados con la misma actividad, tema o persona, puedan ser fácilmente identificados y recuperados como parte de un agrupamiento conceptual –un “expediente”– según se necesite. Generalmente, y siguiendo la tendencia que existe en equipos de cómputo desde hace muchos años, los agrupamientos se van construyendo en forma de folders, directorios o carpetas, los cuales se subdividen a su vez en otros subdirectorios o subcarpetas a distintos niveles de profundidad, y así sucesivamente en lo que se conoce como una estructura en forma de árbol, ya que su diagrama se asemeja a esa forma, sólo que invertida.



Estructura de carpetas y documentos en forma de árbol.

Independientemente del esquema de categorización o clasificación que elijas, es conveniente que separes en agrupamientos distintos los materiales de duración corta o mediana de los documentos que son para preservación a largo plazo. Si bien tus agrupamientos lógicos son iguales y como tú decidas, se recomienda que físicamente residan en distintas carpetas o subdirectorios, de tal forma que a la hora de tomar decisiones de copia, destrucción, transferencia, etc., eso te facilite su manejo. De hecho se recomienda que parte de la organización de tus documentos digitales esté basada –además de los diferentes tipos de ellos– en el periodo por el que ciertas clases de documentos requieran estar resguardados.

Asegúrate que los materiales “críticos”: documentos sumamente importantes, muy confidenciales o muy valiosos para tu persona o actividad queden más o menos contiguos en un cierto número manejable de carpetas, y evita por tanto que estos materiales se encuentren dispersos a lo largo de múltiples subdirectorios y carpetas, perdidos entre docenas o cientos de otros materiales de mucho menor relevancia. Muchos usuarios deciden guardar estos directorios sensibles en una “bóveda” encriptada dentro de su disco, la cual es un conjunto de archivos que son cifrados bajo una contraseña con la ayuda de un programa al efecto. Algunas memorias USB ya tienen esta característica de fábrica, la bóveda o zona cifrada y una zona libre para compartir.



Memoriza o escribe tu esquema de organización. Asegúrate de que memorizas perfectamente el esquema de tu organización documental, o mejor aún, escríbelo en un documento que describa todos los agrupamientos de materiales existentes, así como sus interrelaciones. En este documento, el cual es llamado esquema de organización, a cada grupo de documentos le es asignado un código o nombre que deberá ser asociado a su vez a cada documento de archivo del agrupamiento, sin importar su medio o ubicación. En consecuencia, todos los documentos de un mismo grupo compartirán ese código o nombre, seguido de otra parte o nombre que indique su secuencia dentro del grupo y sirva para diferenciarlos. Abundaré en esto más adelante.

Cómo preservar mi patrimonio digital personal

El identificar grandes agrupamientos de documentos de archivo facilitará su administración durante el tiempo que sean regularmente requeridos y ayudará a asegurar que los documentos que ameriten preservación a largo plazo sean marcados oportunamente y tengan la protección adecuada para asegurar su permanencia. A ti se te facilitará más y te resultará más eficiente asignar los periodos de retención –esto es, el tiempo que estableces para guardar ciertos materiales– a grupos de materiales que a documentos sueltos. El tratar de asegurar que ciertos materiales se conserven todo el tiempo que sea necesario deshaciéndose al mismo tiempo de lo que no es requerido puede volverse una tarea chocante e ineficiente si uno pretende hacerlo a nivel de documento por documento. Si tú estableces que dentro de un agrupamiento algunos documentos deben permanecer más que otros podrás ahorrar tiempo al examinarlos en conjunto como grupo. Además, siempre puedes crear subgrupos dentro de los agrupamientos que te faciliten aún más esta tarea durante el periodo de retención. A la destrucción o transferencia final de documentos a un repositorio histórico se le conoce en el medio archivístico como “disposición”.

Si te acomoda, puedes reflejar tus agrupamientos usando las estructuras que vienen ya integradas en muchos computadores, por ejemplo la de Windows. Puedes usar entonces la estructura existente de “Mis documentos”, “Mis Videos”, “Mi Música”, etc. A partir de esto puedes comenzar a agregar otras carpetas que no existan y desagregar en las subcarpetas que a ti te convengan hasta el nivel de desagregación que consideres pertinente. Si no lo deseas así, puedes partir del directorio raíz de tu disco duro “C”, “D”, etc., para construir de inicio tus carpetas y agrupamientos. Es válido también usar más de un disco duro con la única condición de que tengas en cuenta a la hora de hacer movimientos, protecciones, transferencias, destrucciones de archivos, etc., que estás operando en más de un disco duro.



No guardes duplicados ociosos. Aparte de los respaldos, siempre es bueno tener una copia extra de documentos muy importantes en otra carpeta, pero evita a toda costa

guardar duplicados ociosos de tus documentos. Por ejemplo, algunas personas guardan en su computadora copia de una fotografía digital en una carpeta “Del viaje”; otra copia de la misma fotografía en la carpeta “De la persona”; otra copia en una carpeta “De imágenes”, otra más en “Compartidas”, y así sucesivamente. Lo mismo sucede con algún documento de texto: guardan una copia del texto en el subdirectorío “Del mes”; otra copia del mismo documento en el subdirectorío “Del cliente”; otra copia en el subdirectorío del “Acumulado anual” y otra más en el subdirectorío “Completos”. Estas son copias ociosas y consumen a la larga mucho espacio y tiempo de respaldo, sobre todo en grandes colecciones de archivos voluminosos, como música, video, fotos, etcétera.

Si quieres tener referencia de un cierto documento en múltiples directorios por cuestiones de organización, ello es perfectamente válido, pero entonces usa la opción de “Crear acceso directo” que ofrecen prácticamente todos los sistemas operativos –Windows– por ejemplo, activa esta opción cuando posicionamos el ratón sobre el nombre de un archivo y oprimimos el botón derecho del mismo: aparecerá una lista con varias opciones entre las cuales se encuentra ésta. Los demás sistemas operativos: MAC-OS de Apple, Linux, etc., tienen una opción equivalente. Gracias a ella podemos crear un acceso directo, un ícono o referencia que puede ser arrastrado o llevado a otro subdirectorío o carpeta como si fuera el documento en sí, tantas veces como lo necesitemos, con la ventaja de que el documento original sólo existe físicamente en una carpeta y todos los demás son sólo accesos directos; esto es, apuntadores virtuales que nos permiten acceder al documento desde otro subdirectorío pero sin ocupar prácticamente nada de espacio ocioso adicional. Además, de esta forma, si haces un cambio en el documento original, se verá reflejado automáticamente en todos los otros accesos directos de todos los subdirectorios y cuando sea accedido desde cualquier carpeta tendrás siempre la última versión de ese documento. De otra forma, el efectuar un cambio en un documento del cual existen físicamente varias copias implica tener que actualizar todas y cada una de las copias de los otros subdirectorios; si omitieras actualizarlas todas, hay un riesgo muy grande al tener versiones distintas en distintos sub-

directorios, y esto introduce confusión y potenciales errores futuros. El acceso directo simplifica la tarea y evita de golpe este riesgo a la vez que reduce sensiblemente el espacio ocupado por copias ociosas y por ende el tiempo de respaldo. Atención, con esta técnica deben tomarse precauciones extras a la hora de borrar el archivo original; una vez que este es borrado o destruido, los accesos directos no contienen nada en realidad y sólo apuntan al limbo.



Extiende tu estructura a otros dispositivos. Esta estructura de agrupamiento debe irse extendiendo a todos los dispositivos que manejes: computadoras de escritorio, portátiles, netbooks, tabletas –pads–, teléfonos inteligentes –Smartphones–, agendas electrónicas, lectores de e-libros –e-readers–, etc. Por supuesto, la estructura no será idéntica a lo largo de todos ellos, pero debe conservarse lo más posible en forma homogénea. Si utilizas estructuras de organización completamente diferentes en tus distintos dispositivos, sin duda te crearás confusión futura.



Evita crear estructuras de carpetas o directorios demasiado grandes. Cuando tengas tantas subcarpetas dentro de una carpeta que ya no puedan ser visualizadas dentro de una sola pantalla, es tiempo de considerar crear nuevos niveles y subdivisiones de esa carpeta.



Mantén las carpetas con archivos de documentos, fotos, datos, música, etc., separados de las carpetas de los programas. Separar los programas en carpetas distintas de los demás materiales digitales reduce el riesgo de borrar esos materiales cuando instalas o actualizas programas. Por ejemplo, los archivos PDF que utilices en tu computadora deben estar en una o varias carpetas propias de ese tipo de documentos y no todos en la carpeta “Archivos de Programa/Adobe”.



Crema las subdivisiones más lógicas posibles de carpetas y subcarpetas. Si creas una carpeta llamada “Declaraciones de impuestos”, es lógico crear subcarpetas por cada año de

ellos y no sólo una gran carpeta con toda esta información acumulada. Si creas una carpeta de “Fotos de viajes” es lógico crear subcarpetas por cada viaje y no sólo una gran carpeta de fotos; si creas una carpeta de “Expedientes de pacientes” es lógico crear subcarpetas por cada paciente, y así sucesivamente.



Junta peras con peras y manzanas con manzanas. Si decides que tu estructura de carpetas debe ser por tipo de formatos de documento, agrúpalos así, y sé consistente: de esta forma, las presentaciones ppt de PowerPoint pertenecen a una carpeta, los artículos PDF a otra, los archivos mp3 a otra, las imágenes jpg a otra y así sucesivamente. La excepción a esto es cuando decides que el arreglo en carpetas es temático, por eventos, por persona, etc. Así, una carpeta del evento “Congreso literatura 2012” dada su naturaleza contendrá todos los archivos derivados de ese evento, ya que ése es el factor de agrupamiento, sin importar si sus contenidos son archivos doc, ppt, jpg, PDF, etc. Una carpeta “Ramos Juan” de un paciente contiene igualmente todo tipo de documentos derivados de ese paciente, sin importar su formato o tipo. Una carpeta “Mi tesis de licenciatura” contiene igualmente todos los documentos derivados de ese trabajo, sin importar su mezcla.

Algunas personas gustan mucho de usar un programa administrador de archivos especializado para organizar y manejar sus carpetas, archivos y documentos, como PowerDesk, FileQuest Gold o Universal Explorer. La mayoría nunca llega a acostumbrarse a su uso, y por tanto su utilización con miras a facilitar la administración de archivos y carpetas debe hacerse con toda reserva y queda a discreción del usuario. En todo caso y como sucede con muchas otras aplicaciones, sin duda la mejor recomendación es descargar las versiones de prueba, usarlas intensivamente por ese periodo, y al final decidir si vale la pena la inversión o no. El simple hecho de usar estos programas administradores de archivos no mejora automáticamente la administración de ellos. Como ya mencioné, depende mucho de qué tanto se acomode cada usuario a su uso.



5. Identifica adecuadamente tus materiales digitales



Identifica apropiadamente todos tus materiales digitales tanto en el nombre como en sus metadatos. La identificación plena sirve para diferenciar a un documento de archivo de otro, para distinguir entre distintas versiones del mismo documento y para mantener la identidad de un documento de archivo desde el momento de su creación hasta el de su preservación a largo plazo. Proporcionar un nombre con significado a un archivo dentro de la computadora ayuda a identificar su contenido además de facilitar su localización.



Crea tu propia nomenclatura de nombres y sé consistente con ella. Ve creando una nomenclatura uniforme de nombres de tipos de documentos o un código de identificación para ellos. Por ejemplo, si creas con frecuencia tareas escolares una nomenclatura posible para este tipo de documentos podría ser tarea + nombredelamateria + fecha; los nombres posibles de archivo de este tipo de documentos podrían ser “tarea-algebra1-18-02-12” o “tarea-geom-analitica-14-03-12” o “ensayo-etimologia-05-04-2012”. O puedes inventar un código para clasificarlos como “TA” para tareas, “TF” para trabajo final o “EN” para ensayo. Puedes optar también por suprimir los guiones en las fechas. Los nombres posibles entonces serían más cortos: “TAalgebra1-180211”, “TFg-analitica-140312” o “ENetimologia-05042012. Es necesario mantener los nombres lo más cortos posibles, de tal forma que puedan ser visualizados en la pantalla sin que se trunquen, y poder dictarlos o retectarlos sin error. Pero ¡atención!, los nombres no deben volverse tan cortos y ambiguos que no describan nada: “resumen mate” o “tarea bio” o “trabajo final 2012” o “ensayo1”, “ensayo2” y “ensayo3”, o “resumen ventas” son cortos, pero demasiado ambiguos en su descripción. Sé consistente con la nomenclatura una vez que la elijas: por ejemplo, escoge una manera de escribir las fechas y apégate a ella: debe evitarse en un momento dado escribir 140312, la si-

guiente vez 14032012, la siguiente 14-03-12, luego 14mzo2012, 03142012 y así sucesivamente.



En los nombres de archivos de piezas musicales, por ejemplo, se recomienda ponerlos siempre primero el intérprete, luego el nombre de la pieza y después alguna información adicional. Por ejemplo: “mozart-eine kleine nachtmusik-k 525” o “black eyed peas-i got a feeling-live at chicago”. De esta forma quedarán automáticamente ordenadas en el directorio todas las piezas musicales del mismo intérprete así como las distintas versiones de la misma. Si deseas escribir primero el nombre de la pieza, es válido; puedes agregar posteriormente en los metadatos los datos faltantes para ordenar. Lo que es indispensable es que una vez seleccionada una estructura te apegues a ella consistentemente. De otra forma, si algunos nombres de archivo están denominados primero por el intérprete, otros están denominados primero por el nombre de la pieza, otros por el álbum, el número de pista, etc., el orden será totalmente confuso –más bien será desorden–; favorecerá la existencia de copias duplicadas inútiles y se prestará poco a una administración y preservación futuras.



Aprovecha las características propias de la computadora para ordenar y buscar archivos. Si a ti te interesa que ciertos archivos se formen cronológicamente, entonces selecciona por ejemplo una nomenclatura que comience con la fecha: “2011-03-27-asuntoX” siempre será ordenada antes que “2011-04-14-asuntoX” al pedirle a la computadora ordenar los archivos por nombre. Atención, debe usarse una nomenclatura que se ordene correctamente: si seleccionas una nomenclatura del tipo: “27-03-2011”, “14-04-2011”, etc., se formarán primero todos los del día “01”, luego los del día “02”, luego los del día “03” y así sucesivamente, sin importar mes y año, lo cual haría inútil e inoperante la lógica del ordenamiento. También puedes auxiliarte con el sistema de la computadora pidiendo que ordene los archivos, no por nombre, sino por fecha de creación, si ese fuese el caso.



Facilitate la búsqueda. Una nomenclatura lógica construida con lo enunciado anteriormente ayuda también a buscar archivos más fácilmente por medio de las rutinas interconstruidas en el sistema. Por ejemplo, a un buscador de archivos le resultará muy fácil encontrar todos los archivos que comiencen con “tarea-algebra1*” sin importar la carpeta ni el orden en donde se encuentren. De esta forma podrás hallar rápida y fácilmente todas las tareas de la materia “álgebra I” con ayuda del buscador del sistema, o todos los trabajos finales buscando simplemente por “TF*” o todos los ensayos de la asignatura “etimología” buscando por “ENetimologia*”.⁶⁷



Mantén simples los caracteres de la nomenclatura. Se recomienda también que a la hora de crear nombres en una nomenclatura, se eviten las vocales con diacríticos, esto es, acentos y diéresis, ya que crean confusión a la hora de retectar o buscar nombres. Una excepción: decide si quieres utilizar la “ñ” y no la “n”. Deben evitarse también utilizar caracteres especiales dentro de los nombres de archivos, especialmente el punto “.”, la diagonal “/”, la coma “,”, el punto y coma “;”, los dos puntos “:”, el asterisco “*”, los paréntesis “(“)” y los corchetes “[“]” ya que en muchos sistemas forman parte de la estructura de direccionamiento de carpetas y archivos y pueden inducir confusión y errores de máquina a la hora de crear, transferir o buscar archivos. Se recomienda como una buena práctica el uso de sólo letras, números, espacios, guión y guión bajo para establecer la nomenclatura de los archivos. Por ello en los ejemplos anteriores los nombres de asignaturas decían: “algebra”, “analítica” y “etimología” en lugar de “álgebra”, “analítica” y “etimología”. Se recomienda uniformar el uso de mayúsculas y minúsculas a un solo tipo de letra cambiando sólo cuando es útil a la nomenclatura. Así, “TAalgebra” es más visible que “taalgebra” ya que las mayúsculas nos informan con claridad del código utilizado

⁶⁷ En casi todos los buscadores, escribir un signo asterisco * después de un cierto texto implica una especie de comodín; es decir: “lo que sea que siga”. Ejemplo: buscar mex* establece una búsqueda de la raíz “mex” seguida de cualquier texto.

para identificar, pero diferenciaciones en los nombres de archivos como “Poliza de Seguros 2011”, “Fotos Cancun Noviembre 2012” pueden inducir a confusión; mejor: “poliza de seguros 2011”, “fotos cancun noviembre 2012”. Nuevamente, la recomendación indispensable en este aspecto es que una vez seleccionada una manera de escribir la nomenclatura, te apegues consistentemente a ella.



Siempre que puedas, agrega metadatos a cada material.

La nomenclatura estructurada y lógica es muy útil en la administración de nuestros documentos digitales, pero es conveniente reflexionar en este punto acerca de que la identificación plena a nuestros materiales va más allá del simple nombrado adecuado de sus archivos. A la información acerca de materiales digitales que sostiene su identificación y recuperación se le conoce como metadatos. La mayoría de las aplicaciones de cómputo que manejan información automáticamente etiquetan a los materiales digitales con ciertos datos que permiten establecer su identidad ya que ese tipo de información es necesaria para poder localizarlos efectivamente. Sin los metadatos, sería prácticamente imposible encontrar un documento sin abrirlo previamente y leer a lo largo de carpetas y subdirectorios. Además de su identidad, los metadatos describen las propiedades y atributos de los materiales digitales, características que son necesarias para su uso y posterior preservación; de ahí la importancia de asegurarse de que todas las propiedades y atributos esenciales sean registrados correctamente. Por lo general, prácticamente todos los programas que crean documentos ofimáticos –textos, hojas de cálculo, presentaciones, etc.–, los programas que crean archivos de imágenes, los que crean o editan archivos de sonido, etc., cuentan con una zona o comando que permite introducir estos metadatos al documento, ya sea al momento de crearlo o en alguna etapa posterior de su uso. Debemos acostumbrarnos a usar esta opción para ir agregando estos metadatos a cada uno de nuestros documentos.

A las propiedades y atributos que manifiestan la identidad de los materiales digitales se les conoce como metadatos de identidad. Sin ser una lista exhaustiva, estos incluyen:

Cómo preservar mi patrimonio digital personal

- a) Nombres de las personas involucradas con la producción de los materiales digitales. Estos pueden ser:
 - 1. Autor. Persona física o moral que emite un documento o en cuyo nombre u orden el documento ha sido emitido.
 - 2. Destinatario. Persona física o moral a quien se dirige o para quien está previsto un documento de archivo, cuando aplique.
 - 3. Receptor. Persona física o jurídica a quien se le envía copia del material, incluyendo copia ciega, cuando aplique.
- b) Nombre del documento. El título del documento o material digital.
- c) Temas, materias o asunto. La materia, asignaturas o palabras clave de las que trata el material digital.
- d) Fecha(s) de creación, última modificación y transmisión, cuando apliquen.
- e) Código o clave de clasificación, o nombre del subdirectorío o subcarpeta a la que pertenece el material.
- f) Indicación de anexos al documento, cuando aplique.
- g) Indicación de los derechos de autor o de propiedad intelectual, cuando aplique. Aquí pueden agregarse licencias *Creative Commons* si es el caso. Véase el apartado “*Copyright*, derechos de autor y *Creative Commons*” al respecto.
- h) Indicación de borrador o del número de la versión del documento, cuando aplique.
- i) Datos propios del tipo de material: cada tipo de material tiene información muy propia e importante, por ejemplo: resolución para imágenes, ISSN para revistas y artículos, etcétera.

Por supuesto, no es posible desde el punto de vista práctico llenar todos y cada uno de los metadatos enunciados anteriormente para todos y cada uno de nuestros documentos, pero debemos notar que algunos de estos campos son sumamente útiles para ciertos tipos de documentos. Por ejemplo, el campo de “versión” en documentos que se van construyendo en forma de Wiki o colaboración colectiva, o en aquellos textos que vamos puliendo a lo largo de varias sesiones con distintas y sucesivas versiones. Cuando usamos en nuestro documento alguna imagen, texto, etc., que

proviene de una licencia *Creative Commons* de la red es muy útil consignar en estos metadatos como referencia la línea con el tipo de permiso así como el URL de dónde sacamos esa licencia, para evitarnos futuras molestias. Cada persona debe ponderar cuáles son los metadatos mínimos que debe registrar para cada uno de los tipos de sus documentos y cuáles son los datos adicionales deseables que pueden ser integrados a esos documentos. La recomendación principal en este aspecto es que no dejes a tus documentos sin ningún metadato. Obviamente, la decisión de cuáles metadatos para cuáles documentos forma parte de tu plan de archivos enunciado anteriormente.



6. Respalda tus materiales digitales con frecuencia y con método

Es imposible respaldar lo que ya se dañó. No esperes que el computador te diga: 'pienso fallar próximamente; sería bueno que respaldases'.

Anónimo

Realiza más de una copia de tus materiales digitales con frecuencia, periódicamente, y bajo un plan preestablecido. Partiendo del hecho de que ninguna computadora ni ninguno de sus discos duros para almacenamiento es a prueba de fallas, al usarlos existen innumerables factores que pueden provocar corrupción o pérdida accidental de datos o documentos de archivo o, mejor dicho, existen innumerables factores que indudablemente van a provocar corrupción o pérdida accidental de datos, tarde o temprano. Cuántas veces hemos dicho u oído decir: "...pero si ayer todo funcionaba perfectamente y ahí estaba la información; ¿por qué hoy ya no?". Ya hemos comentado acerca de los riesgos informáticos en la sección correspondiente. Si bien la mayoría de los discos duros y otros dispositivos de almacenamiento tienen un factor de

falla promedio bastante aceptable, conocido como MTBF,⁶⁸ la realidad es que la falla se puede presentar entre la primera hora y la última de la vida del equipo. Cuando un disco duro afirma que tiene 200,000 horas de MTBF quiere decir que el fabricante afirma que esa cifra es un promedio entre fallas, y no que el disco está garantizado por 200,000 horas sin fallar. De hecho, este número de horas –el cual puede verse en la publicidad actual– correspondería a casi 23 años funcionando 24 horas al día. Yo no espero ni creo ni deseo que mi disco dure esos años en funcionamiento ininterrumpido y no me confiaré. El número en realidad no dice nada.



Obtén copias de seguridad desde la creación. Mientras escribes y editas tus documentos, utiliza siempre las opciones de respaldo automático que vienen en esas aplicaciones. Todos los programas que crean documentos –de todo tipo, pero especialmente los de oficina– tienen opciones interconstruidas que permiten establecer automáticamente una copia de seguridad, la cual se creará cada vez que se abre un archivo con este programa. En caso de falla de energía, colapso del programa, etc., la copia se cierra instantáneamente y la siguiente vez que se restablezca el sistema puede ser recuperada, por lo general hasta el estado que guardaba hasta la última protección antes de la falla. Si además configuramos estos sistemas para que realicen automáticamente una actualización de esa copia con frecuencia, digamos cada cuatro o cinco minutos, en caso de una falla del programa o de la energía, lo más que podremos perder son esos pocos minutos entre actualizaciones del documento. Pero estas opciones deben ser activadas por el usuario. Por lo general no se encienden en automático.

Hasta hoy, la mejor forma de asegurarse contra estos accidentes es realizar copias de los documentos frecuente y periódicamente.

68 MTBF. “*Mean Time Between Failures*” o “*Tiempo Medio Entre Fallas*”. Media aritmética o promedio de ocurrencia medido en un conjunto de fallas de un cierto dispositivo.

Desarrolla tu propio “método” de respaldos.



*El buen juicio proviene de la experiencia;
la experiencia proviene del mal juicio.*

- ❖ Si eres de los que nunca respaldan su información, y hasta ahora nunca te ha sucedido un percance mayor, ¡felicidades!, has estado jugando a la ruleta rusa con tu información y hasta ahora has corrido con suerte; pero, ¿cuánto más va a durar? Por ley de probabilidades, cada vez estás más cerca de un accidente mayor de información.
- ❖ Si eres de los que respaldan su información después de haber sufrido un percance mayor, mis condolencias, tú eres de los que usan el método de “después de ahogado el niño, tapemos el pozo” y ya pagaste el precio de la falta de respaldos y has aprendido por la vía más dura.
- ❖ Si eres de los que realizan respaldos sólo cada vez que cambias de computadora, o cuando agregas un disco duro a la misma, o en algún otro evento que sucede cada vez que se alinean los ocho planetas, seguramente te has encontrado a la hora de un percance con que sí posees respaldos de una parte de los materiales, pero faltan muchos de los archivos, las versiones están muy atrasadas, los directorios se encuentran revueltos, etc., y por lo mismo, mucho del material es inútil por obsoleto, está incompleto, o hay que invertir mucho tiempo en tratar de arreglar las carpetas, habiendo perdido de todos modos información relevante.
- ❖ Si perdiste la información que era crucial para tu trabajo, o la invaluable versión final de ese documento que con tanto trabajo construiste, o las preciosas y únicas fotos de ese viaje o de esa fiesta memorables, y que apenas hace unos pocos días u horas estaban ahí y sientes la inclinación de cambiar tu alma por la posibilidad de recuperar esa información, mi más sentido pésame. Como dato estadístico, menos de la mitad de estas pérdidas de información son recuperables. Como dice la sabiduría popular: la experiencia es el más cruel de los maestros; primero te aplica el examen y luego te da la lección.

Cómo preservar mi patrimonio digital personal

Todas estas situaciones provienen, en esencia, de una raíz común: falta de método de respaldo. Respaldamos a veces, respaldamos parcialmente, respaldamos en varios dispositivos, respaldamos de una manera o de otra, etc. Ello debido a que respaldar todos nuestros materiales con mucha frecuencia es algo que aparentemente es engorroso, puede tomar mucho de nuestro tiempo y por ello no es práctico y tendemos a evitarlo. Esto es cierto sólo en parte: sucede cuando se hace sin método, y por lo mismo debemos desarrollar uno que nos ayude a respaldar minimizando el tiempo y el esfuerzo y maximizando la protección. ¿Cuáles son los factores que debemos tomar en cuenta para desarrollar nuestro método de respaldo?

Para contestar a esta pregunta debemos reflexionar antes en cuántas variantes existen de los respaldos:

- ❖ Respaldo total de la computadora a imagen del disco. Este tipo de respaldo obtiene una copia idéntica del disco duro de la computadora; incluye sistema operativo, programas instalados, datos y documentos, archivos ocultos, archivos borrados, todo. Este respaldo es recomendable recién instalada la computadora o cuando se ha usado la opción “Reinstalado de origen” por alguna causa, y se usa para reconstruir el sistema y sus datos originales después de una falla grave de la computadora, volviendo a dejarlo como se encontraba en un inicio. Dado que el sistema operativo, los programas, los datos y los documentos son actualizados y cambiados con frecuencia, no se debe usar para protecciones y recuperaciones parciales de datos o programas.
- ❖ Respaldo de todas las carpetas de datos. Este tipo de respaldo abarca todas las carpetas que contienen específicamente documentos y datos –textos, hojas de cálculo, presentaciones, imágenes, sonidos, videos, bases de datos, etc.– y omite todas las carpetas del sistema operativo y programas. Salvo el respaldo a imagen del disco mencionado en el párrafo anterior, el cual sí reconstruye todo hasta su último estado, no tiene caso proteger carpetas con programas o partes del sistema operativo, ya que en caso de daño, si simplemente se copian de regreso a la computadora con la intención de restaurarlos por lo general no funcionarán, ya que es muy común que ciertos parámetros,

direcciones, archivos ocultos, registros, etc. necesarios para su correcto funcionamiento no sean congruentes y esto ocasiona falla del sistema o programa y requiere forzosamente de su reinstalación. Este método requiere de un mínimo control por parte del usuario, ya que las protecciones se hacen sobre el total de las carpetas de datos, pero puede llegar a tomar mucho tiempo y espacio si el total de las carpetas es muy voluminoso.

- ❖ Respaldo de sólo las carpetas que han sufrido cambio desde el último respaldo anterior. Esto ahorrará tiempo, pero implica que lleves un buen control de las carpetas que han quedado estables y de las que han sufrido cambios. Si las carpetas son muy voluminosas, el tiempo de respaldo puede seguir siendo muy alto.
- ❖ Respaldo sólo de los documentos que han sufrido cambio desde el último respaldo anterior. Esto ahorrará todavía más tiempo que los tipos de respaldo anteriores, pero implica que lleves un riguroso control de los documentos que han quedado estables y los que han sufrido cambios. La probabilidad de cometer errores y omisiones graves de esta manera crece sensiblemente si no se lleva un estricto control.

Teniendo esto en mente podemos empezar a diseñar nuestro método personal de respaldo. Para ello debemos seguir los siguientes pasos:



Obtén respaldos de inicio. Siempre que adquieras una computadora o cualquier otro dispositivo –*netbook*, tableta, teléfono inteligente, etc.– obtén una copia con un respaldo a imagen del disco, es decir, una copia que incluya todo el sistema, sus programas y datos, tal como vienen de fábrica. En general casi todos estos dispositivos tienen en el sistema operativo esta opción de respaldo total a imagen: ubícala y úsala. Lo mismo debe hacerse cada que el equipo regrese de una restauración mayor como cambio de discos o memoria, reinstalación de fábrica, etc. En el sistema operativo Windows puede verse esta opción en la pantalla principal, en “Inicio”, “Todos los programas”, “Accesorios”, “Herramientas del sistema”, “Copia de seguridad”. No uses esta opción para respal-

dos frecuentes. Si no lo hiciste de inicio, obtén un respaldo total actual; sólo ten en cuenta y recuerda que tu respaldo guardará el estado en el que se encuentra actualmente tu equipo. Cuando hagas este respaldo de inicio asegúrate que los archivos que estás respaldando estén completos, íntegros y libres de virus. Este tipo de respaldo sirve para poder restablecer la máquina al estado en que se encontraba de inicio, en caso de pérdida de sistema operativo, programas principales, etcétera.



Separa programas de datos. Si ya seguiste o estás en proceso de organizar tus materiales digitales en agrupamientos lógicos (cuarta acción indicada en este capítulo) tienes o tendrás pronto carpetas con datos separados de los programas. Normalmente los programas o aplicaciones al instalarse te proponen una carpeta de instalación o te piden que tú definas el nombre de esa carpeta. Por ejemplo, la mayoría de aplicaciones proponen instalarse en una carpeta llamada “Archivos de programa” o *Program Files* en donde a su vez se creará una subcarpeta con el nombre de la aplicación que se instala: Adobe, McAfee, StarOffice 8, etc. Si el usuario ha aceptado estas propuestas, al cabo de un tiempo todas las aplicaciones instaladas estarán en esa carpeta y subdivididas a su vez en subcarpetas propias de cada aplicación. Esto facilita los respaldos, pues podemos hacer la protección de esta carpeta “Archivos de programa” solamente cuando hayamos instalado una nueva aplicación o una nueva versión de alguna de ellas. No es necesario respaldarla cada vez que hacemos una protección de los datos, y podemos hacer así respaldos independientes de datos y programas. Esto reduce mucho los tiempos de respaldo ya que por lo general estas carpetas de instalación de aplicaciones son muy voluminosas y si ya tienes una buena cantidad de ellas el tiempo de respaldo de esta carpeta puede tomar tiempos muy largos. Si no deseas aceptar ese nombre de carpeta tú puedes seleccionar el que quieras; solamente debes ser consistente y no instalar unas veces en una carpeta con un nombre, otra vez en otra carpeta, la siguiente en otra, ya que eso no permite tener las aplicaciones contiguas y por lo mismo hace su respaldo casi imposible. También debes tener cuidado si tienes más

de un disco duro. Muchos usuarios acaban con la mitad de sus aplicaciones instaladas en “C:\Archivos de programa” y la otra mitad instaladas en “D:\Archivos de programa” por no fijarse y no ser consistentes con las instalaciones. Esto dificultará enormemente en un futuro el respaldo de las aplicaciones. Atención, si esto ya sucedió, no se puede simplemente cortar y copiar estas carpetas de un disco duro al otro, ya que los sistemas quedaron apuntando hacia ciertas unidades de disco durante la instalación. La simple acción de cambiarlos de lugar los dejará casi seguramente inhabilitados para funcionar. En este caso, es mejor desinstalar la aplicación y volverla a instalar en el disco duro correcto.



Mide tus tiempos de respaldo de datos. Cuantifica el material digital que posees y el tiempo que usas para el respaldo total de todas las carpetas de datos. Si ya seguiste o estás en proceso de organizar lógicamente tus materiales digitales tienes o tendrás pronto carpetas con datos separados de los programas y por lo mismo ya no respaldas simultáneamente carpetas de programas con carpetas de datos. Si el conjunto de las carpetas de datos no es demasiado voluminoso y el tiempo para respaldarlo es razonable –menos de treinta minutos– protege todas las carpetas con datos cada vez. Por supuesto, omite las carpetas de sistema y programas, típicamente las que se encuentran en la carpeta “Archivos de programa”. Ten en cuenta que conforme vayas acumulando datos digitales con los años, este método irá tomando más y más tiempo y espacio por lo que se irá haciendo ineficiente y tendrás que considerar cambiar hacia otros más elaborados.



Si el tiempo de protección de todas las carpetas de datos es muy grande, no deseas estar realizando ese método de respaldo. Asumiendo que ya realizaste un respaldo total de sistema, programas y datos como respaldo de inicio, debes decidir si prefieres el método de respaldo de carpetas que cambiaron, o el método de respaldo de documentos que cambiaron que explicamos un poco antes. Para ayudarte en esta decisión, conviene tener en mente que estos tipos de respaldo pueden hacerse manualmen-

te; en éste debes dar la orden de cada carpeta o archivo que debe protegerse, y por tanto debes llevar manualmente el control de los archivos y carpetas cambiados y que deben copiarse.

Sin embargo, esta tarea también puede hacerse automáticamente con ayuda de un programa construido al efecto, con fines de respaldo y sincronización; esto es, sólo el copiado de archivos o carpetas que han cambiado entre dos dispositivos de origen y destino. En este aspecto, existe un gran número de productos disponibles, en una amplia gama de variedades: para sistema operativo Windows, para Apple, para Linux, etc. Los hay en tipo libre y de pago; por supuesto los de este último tipo son más completos, lo cual no significa que los de tipo libre no realicen la tarea adecuadamente. Hay una lista más o menos completa de este tipo de programas en Wikipedia⁶⁹ que puede serte de ayuda. No es exhaustiva, y debe considerarse que estos productos se actualizan constantemente y por ello puede estar desactualizada, pero sin duda es un buen punto de inicio en este aspecto si optas por estas herramientas. Es conveniente que consideres de inicio si quieres realizar los respaldos manualmente o usar alguno de estos programas para realizar la tarea del respaldo. No te precipites; te sugiero experimentar con varios de ellos y una vez que encuentres aquél que mejor te acomode, debes dominarlo, apegarte a su uso y respaldar con frecuencia. Todos los programas de pago de este tipo tienen una versión de evaluación temporal para experimentar con ellos. Una buena manera de lograr esto es comenzar desde hoy haciendo los respaldos manualmente por un tiempo –esto no cuesta nada y no posterga más tus respaldos– mientras haces la evaluación de los programas de respaldo y encuentras el adecuado.



Independientemente del conjunto de información que selecciones para respaldar, debes considerar si respaldas en un dispositivo propio tuyo –disco duro, memoria USB, etc.– o puedes considerar también la opción de hacer tus respaldos de in-

69 Wikipedia. “List of Backup Software”. http://en.wikipedia.org/wiki/List_of_backup_software

formación en un servidor de paga. Actualmente, éste es un servicio que se promueve más y más. Consiste en una empresa proveedora de servicios de alojamiento *-hosting-* la cual por una renta mensual o anual te asigna una cierta cantidad de bytes para almacenamiento *-Mega, Giga, Tera-*, así como una cantidad o cuota de transferencia de información en un cierto periodo; es decir, cuántos bytes puedes subir *-upload-* o descargar *-download-* de ese servidor en un cierto periodo por una cierta cuota. Por supuesto, la renta es proporcional a la magnitud de espacio requerido y a la cuota de transferencia. Contrario a lo que uno pueda pensar, esta no es una opción necesariamente cara, ya que hay proveedores que dan cantidades razonables para la mayoría de las necesidades personales a costos económicos. Pero debes tener en cuenta que para este tipo de respaldo requieres indispensablemente de un servicio de Internet de banda ancha *-mínimo un Megabit por segundo-*. De otra forma el tiempo que esto tarde puede ser realmente inoperante y muchísimo mayor que el de un respaldo local, el cual obviamente no requiere de ninguna conexión a la red. Las versiones más sofisticadas de estos sitios web incluyen acceso a aplicaciones y servicios más allá del simple alojamiento, y a esto se le conoce como “cómputo en la nube”. Obviamente, entre mayor sea el número de servicios y programas que se requieran más allá del simple alojamiento, el precio irá subiendo sensiblemente. No selecciones esos sitios web más sofisticados si sólo requieres alojamiento para respaldo. Si optas por este tipo de servicios, y aunque en general son muy confiables, haz de todas formas tu propia copia local de todas las carpetas de datos al menos una vez al mes, o más frecuentemente si has creado o modificado información importante para ti. No te confíes. No por el hecho de ser una empresa muy renombrada o importante es a prueba de todo. Existen muchas historias de información perdida, al menos por un tiempo, por ejemplo eventos de servidores incautados por el FBI debido a un usuario criminal, dejando inhabilitados a todos los que tenían información en esos servidores por un lapso, o la tristemente célebre historia del cierre total de la empresa *Megaupload*, dejando a muchos usuarios sin datos.



Tomadas las decisiones anteriores, y si has decidido realizar respaldos a nivel local, lo único que resta para hacer un plan de respaldos es seleccionar el dispositivo donde se va a respaldar; mejor dicho, los dispositivos. De la manera más básica, podrías hacer los respaldos de tus carpetas en otras carpetas de tu disco duro, o en otro disco duro de tu computadora. La enorme desventaja de esto es que si tu computadora sufre un daño estructural, es robada, se incendia o se moja, todos los respaldos desaparecerán al mismo tiempo, y te quedarás sin nada. Por lo mismo no se recomienda en lo absoluto. Se sugiere usar un dispositivo externo a la computadora. Algunos prefieren hacer respaldos usando el quemador de CD o DVD de la computadora, ya sea fijos o re-escribibles –CD-R y CD-RW, DVD-R y DVD-RW–.

Algunas personas, los que no tienen datos demasiado voluminosos, prefieren las memorias de estado sólido, conocidas como USB. La inmensa mayoría prefiere un disco duro externo con interfaz USB, dado lo práctico de este dispositivo y la relación costo-capacidad que es muy atractiva y rentable. Ésta es una de las mejores inversiones que puedes hacer en materia de protección de tu información. No compres un disco demasiado pequeño en capacidad, ya que no podrías proteger toda tu información –otro motivo por el cual cuantificaste el volumen de todas tus carpetas de datos desde el principio de este apartado–. No compres tampoco un disco con diez veces más capacidad de lo que necesitas: es dinero ocioso que no vas a usar –el disco estará obsoleto en tres o cuatro años y lo más probable es que desees remplazarlo–. Un disco que tenga entre dos y cuatro veces la capacidad actual requerida de respaldo será la mejor inversión. El precio varía con la capacidad, pero también de acuerdo con el tamaño. Si a ti no te importa tener un disco duro un poco más voluminoso, de portafolio, te ahorrarás una buena cantidad; si a ti te agrada tener un disco duro muy portátil, de bolsillo, deberás desembolsar algo más por esta comodidad. Aparte del tamaño, ambas variedades funcionan muy bien para el propósito de respaldo. Si eres fanático de los dispositivos ultra-avanzados la empresa Kingston acaba de anunciar sus pequeñas memorias

flash USB de 128 y 256 Gigabytes de capacidad. Eso sí, son bastante más caras que un disco duro externo de capacidad equivalente.



No pongas todos los huevos en una sola canasta.

Todos los discos duros vienen con garantía... de que colapsarán.

Anónimo

Al principio de este inciso mencioné que no era “el” sino “los” dispositivos. Esto es debido a que nunca, nunca, se recomienda respaldar en un solo dispositivo, sea local o remoto. Al igual que el disco duro de la computadora, el disco duro externo no tiene palabra de honor y fallará tarde o temprano, y si creemos en la Ley de Murphy,⁷⁰ fallará en el momento y circunstancia en que cause más daño, independientemente de cuán grande sea su MTBF. El alojamiento externo de paga, el de la nube, siendo muy seguro y confiable, tampoco es 100% a prueba de fallas, como ya mencionamos. Hay muchas personas que han decidido guardar su información –o al menos buena parte de ella– en sitios de la nube gratuitos. Esto no es malo; el que remuevan todas las copias de sus cámaras y computadoras sí lo es; por ejemplo, hay muchas personas que acostumbran guardar sus fotografías en Facebook, Flickr, Picasa, Photobucket, “23” o algún otro sitio semejante y luego las borran de sus propios dispositivos pensando que esa única copia queda segura ahí. Hay innumerables historias documentadas de pérdida total de esos malhadados materiales por diversas razones. La Ley de Murphy también aplica allí.

Por lo mismo, se recomienda encarecidamente tener al menos dos dispositivos para respaldo, los cuales se deben ir alternando en sus respaldos, de tal forma que en caso de falla de un dispositivo, el otro tendrá el respaldo anterior y el daño será menor. Obviamente cada uso del dispositivo reescribe nuevamente los datos ahí

⁷⁰ Edward Murphy, 1949: “Si algo puede salir mal, seguramente saldrá mal. De entre un conjunto de cosas que puedan fallar, aquélla que cause más perjuicio será la que falle”.

Cómo preservar mi patrimonio digital personal

respaldados destruyendo los anteriores y cambiándolos por una versión más nueva. Si decides tener más de dos respaldos, mejor; lo aseguro, nunca estará de más. Siempre es mejor pecar de exceso que de falta.

Se recomienda también que guardes los respaldos en ubicación física distinta del equipo: si sufres un siniestro y se inunda o quema tu oficina, o te roban tu equipo, de nada servirá tener varios respaldos si todos se encontraban junto al computador que sufrió el accidente y corrieron su misma suerte. Si has sido precavido y tienes más de un respaldo, no los guardes en el mismo lugar. Si viajas, nunca llesves el disco de respaldo en la misma maleta que la laptop; si algún percance sucede, les sucederá simultáneamente a la computadora y a su respaldo, y todo habrá resultado inútil. Si llevas información indispensable para tu viaje –ponencia, cotización, proyecto– asegúrate de llevar una copia extra en una memoria USB encima de ti, en el bolsillo de la camisa o el pantalón y no en una maleta de mano.

Muchas personas acostumbran obtener, –además de sus respaldos regulares en discos externos–, una copia eventual de emergencia de su información en DVD o algún dispositivo económico semejante, y guardarlas en lugares remotos a los que normalmente frecuentan. Obviamente, hay que refrescarlas o actualizarlas con cierta frecuencia. Si optaste por tener una protección en un servidor de la red, podrás descargar tu información desde ahí en caso de un percance.



Desarrolla una política personal rigurosa que asegure que tus materiales son respaldados con la debida frecuencia.

Decir que no tenemos tiempo para respaldar nuestros datos es como decir que no tenemos tiempo para ponernos el cinturón de seguridad al subirnos al auto.

T.E. Ronnenberg

Tus datos son tan buenos como tu último respaldo. Una vez que has hecho tu plan de respaldos, debes asegurarte de que sea ejecutado con frecuencia; esta última depende de cuán valiosa es la in-

formación que ha cambiado. Si estás haciendo cambios significativos a un documento muy valioso que debes presentar mañana en la escuela o la oficina, la frecuencia deberá ser de varias veces al día. Si acabas de descargar valiosas fotografías de tu último viaje debes hacer el respaldo ese día, sobre todo antes de borrar la memoria de la cámara. Si no has hecho cambios muy abundantes o delicados a tus documentos durante días entonces tal vez puedas programar la frecuencia de respaldo una vez a la semana. Asegúrate de que haces también respaldos de todos tus diferentes equipos: laptop, *netbook*, etc. A muchas personas les gusta sincronizar sus equipos; es decir, mantener las mismas versiones de documentos en diferentes equipos, de forma de poder trabajar y usar sus documentos indistintamente en cualquiera de ellos. Para ello, es necesario respaldar y se recomienda descargar y usar un programa sincronizado que se encarga de esta tarea rápida y eficientemente. Si no tienes esta opción, guarda los datos sensibles de tu teléfono en una memoria adicional en casa.



No hagas planes utópicos y ultra-ambiciosos, tales como proponerse respaldar toda tu información todos los días. Eso es igual que proponerse hacer una dieta superrestringida.

Cómo preservar mi patrimonio digital personal

Lo vas a lograr sólo unos cuantos días. En cambio, haz un plan de respaldos razonable y asequible que pueda ser llevado a cabo consistentemente y luego, apégate a él como religión, haz de él una costumbre. De nada servirá trazar un plan de respaldo eficiente, dinámico y accesible si luego olvidas ponerlo en práctica o le sucede lo mismo que a los propósitos de año nuevo, que son olvidados después de unas pocas semanas de realizarlos. Recuerda que de buenas intenciones está pavimentado el camino de los infiernos. Este no debe ser necesariamente un proceso largo y tedioso que tiene que ser repetido día a día como penitencia. Si te organizas bien de acuerdo con las recomendaciones presentadas, las protecciones largas y completas se hacen sólo muy de vez en cuando. Una vez organizadas las carpetas y la información adecuadamente, las protecciones frecuentes se hacen sólo de las carpetas y documentos cambiados y por lo mismo no toman demasiado tiempo y las protecciones cotidianas se hacen sólo de lo que es sumamente vital para nuestras actividades y por tanto toman todavía menos. De ahí la importancia de haber organizado la información adecuadamente. Por lo mismo, si haces un método de ello y te auxilias con un pequeño programa al efecto, descubrirás que el hacer buenos y frecuentes respaldos no implica estar horas y horas contemplando la pantalla de la computadora. Lo reitero, es simplemente una cuestión de método y organización.

Haz una lista de los tipos de información a respaldar y la frecuencia con que vas a hacerlo; asegúrate de que incluyes todo además de programas y datos: favoritos, *cookies*, correos, etcétera.



7. Establece, controla y protege tus versiones adecuadamente

Es necesario establecer una nomenclatura adecuada, consistente y uniforme de las distintas versiones de nuestros documentos existentes en nuestros equipos, así como llevar un control de ellas y realizar los respaldos adecuados. Como ya

hemos establecido en la definición de versión, pueden existir múltiples versiones de los distintos documentos alojados en nuestros equipos, por una serie de distintas razones:

- ❖ Documentos que vamos construyendo paulatinamente, y de los cuales tenemos distintas variantes que queremos conservar.
- ❖ Documentos que se van construyendo colaborativamente –Wikis– en sucesivas retroalimentaciones.
- ❖ Documentos que tenemos en distintos formatos o presentaciones: texto, html, PDF, doc, imagen jpg, texto en español, texto en inglés, antes y después del OCR, misma obra en distinta edición, misma pieza musical en distintas interpretaciones, etc., siendo en esencia la misma obra.
- ❖ Documentos que tenemos en versiones de alta calidad para preservación –por ejemplo imágenes tiff– y a la vez en baja calidad para distribución –imágenes jpg–. O documentos que tenemos en versión cotidiana –por ejemplo un texto en docx– y además en versión para preservación a largo plazo –por ejemplo xml u odt–.
- ❖ Documentos que tenemos en distintas versiones dependiendo del equipo o dispositivo en donde los usemos: por ejemplo, la versión doc en nuestra computadora y la versión para *documents to go* que tenemos en nuestra tableta electrónica. Aunque es la misma obra, no son iguales. Véase el apartado de versión en las definiciones.

Quitando los duplicados ociosos, los que ya hemos mencionado cómo evitar con ayuda del acceso directo, es totalmente válido y común tener distintas versiones de nuestros documentos. Sólo que es indispensable establecer ciertas características en su manejo ya que las diferentes versiones de nuestros documentos son una de las fuentes principales de errores y pérdidas documentales debido a su manejo deficiente.

Para contender con este problema, es necesario establecer tres ámbitos principales de acción:



Establecer una nomenclatura de versiones congruente, estructurada y consistente. Esto significa que debes establecer una nomenclatura de versiones a la brevedad posible. Partiendo del hecho de que ya has diseñado una nomenclatura de archivos acorde con lo recomendado en el apartado relativo a la identificación adecuada del material digital, es necesario agregar aquí que esa nomenclatura, además de ser estructurada, debe tener alguna manera de identificar versiones. Esto por lo general se hace agregando un sufijo o terminación al nombre del archivo. Algunas personas acostumbran poner al final del nombre un guión y algo así como V1, V2, V3, o ver1, ver2, ver3, o V1_0, V1_1, o V2_1, V2_2, etc. Recuerda que el punto, coma y dos puntos no son recomendados para ser parte de la nomenclatura. Algunas otras personas deciden poner la fecha de la versión al final del nombre en alguna forma: 240212, ver2012-06-14, etc.; o tal vez fecha y hora: 2012-06-14_08-14. En realidad, no hay ninguna estructura ideal: escoge la forma que mejor te acomode.

Los únicos puntos indispensables son que ella permita diferenciar siempre a una versión de la otra de manera constante y una vez seleccionada esta forma es vital apegarse a esta estructura, de otra forma no funcionará: si nombras a tu primer archivo “tarea 01 matematicas”, a la segunda versión “tarea matematicas 2”, a la tercera “tarea matematicas III”, luego “tarea matematicas 14 feb”, a la siguientes “tarea matematicas 18022012”, “tarea matematicas 2012-02-21”, “tarea matematicas temporal febrero”, “tarea matematicas ver 8”, “tarea matematicas final”, “tarea matematicas final-bis”, “tarea matematicas final-bis-2”, “tarea matematicas para archivar”, “tarea matematicas entregar febr 2012”, “tarea matematicas entregar pdf”, etc., después de un tiempo será imposible discernir cuál de estas versiones precede o sucede a cuál, cuál es la versión definitiva, cuál es la que conviene preservar en definitiva y otras cosas más. En esta situación, lo que acabará sucediendo en un tiempo es que estemos guardando múltiples copias ociosas de una misma tarea, consumiendo espacio y tiempo inútilmente, o acabaremos borrando algunas, con una alta probabilidad de borrar la que no era la versión definitiva. Si además conservamos varias copias de estos

archivos en distintos subdirectorios, la probabilidad de falla crece exponencialmente.



Llevar el control de tus versiones y archivos. Una vez resuelto el problema de la nomenclatura de las versiones, es necesario establecer un mecanismo de control de éstas. Debes decidir cuáles son versiones de trabajo temporales y cuál o cuáles serán versiones para respaldar. Esto es útil ya que puedes establecer que todas las versiones de trabajo pueden respaldarse en los discos correspondientes a protecciones temporales de un cierto día, pero que la versión que va siendo la definitiva para archivo puede ser depositada en otra carpeta destinada a ese propósito; por supuesto, esa versión definitiva en esa carpeta debe ser actualizada cada vez que haya una nueva versión. O puedes decidir ir destruyendo las versiones más antiguas y sólo conservar las más recientes, o también puedes decidir conservar todas las versiones de trabajo en la misma carpeta y una vez entregada la final destruir todas las previas. Lo importante es crear una política que pueda ser seguida y apegarse a ella. Lo peor es querer conservar todo por largos plazos por si las dudas. Al final acaba uno sin preservar nada. Sobre todo, es importante llevar a cabo las acciones cuando el recuerdo de las versiones y sus diferencias está más o menos fresco en la mente. Si se desea hacerlo seis meses después, lo más probable es que no se recuerde nada importante acerca de las distintas versiones y sea imposible valorarlas.



Realizar las protecciones de las versiones con la regularidad establecida en el plan de respaldo. Una vez que se ha establecido el control de las versiones, éste debe ser reflejado en los respaldos correspondientes. Recuerda que entre más numerosos y más trascendentales los cambios sufridos por las versiones, más frecuentes deben ser las protecciones.



8. Toma acciones periódicamente contra la obsolescencia de equipo y programas

La información digital dura para siempre, o por cinco años, lo que ocurra primero.

Jeff Rothenberg, científico senior en la Corporación Rand

Es necesario establecer una estrategia para contender con la velocidad a la que el equipo, dispositivos y programas de cómputo se vuelven obsoletos. El cambio constante y frecuente del ambiente tecnológico impone severos retos a la conservación y por tanto a la preservación a largo plazo de los materiales digitales. Dado que las personas actualizan regularmente sus equipos, sistemas operativos y programas de aplicación deben actualizar de tiempo en tiempo los documentos digitales en ellos contenidos, además de aquellos documentos que hayan sido transferidos a otros medios de almacenamiento tales como CD, DVD, memorias USB, tarjetas de memoria flash, etc.

Dicho de otra forma, cuando te percastes de que ciertas partes del entorno tecnológico se están volviendo obsoletas, es conveniente que las vayas migrando gradualmente a tecnologías actuales de acuerdo con tus requerimientos y posibilidades. Obviamente, los materiales digitales contenidos en ellos también deben ser migrados a soportes con tecnologías actuales. Cuando se reemplace cierto equipo, es importante verificar que el nuevo tenga capacidades que cubran perfectamente las del equipo anterior. Por ejemplo, un monitor nuevo debe poder desplegar un documento que tenga elementos gráficos exactamente de la misma forma en que el monitor anterior lo hacía en lo tocante a su forma documental. Un nuevo programa ofimático debe poder abrir y desplegar adecuadamente un documento de texto de su versión anterior. La planeación correcta de sustituciones de partes de la plataforma tecnológica en forma gradual te permitirá garantizar que ésta no se vuelva obsole-

ta de golpe al mismo tiempo y por tanto que no te imponga gastos muy fuertes al tener que sustituir todo de una sola vez.

Aquellos documentos digitales que hemos decidido preservar por periodos largos se enfrentarán al hecho de que es necesario conservarlos sobre soportes y bajo sistemas operativos y programas que se vuelven obsoletos con el tiempo. Ambos, soporte y entorno tecnológico de sistemas y programas, se vuelven una amenaza para la preservación de esos materiales a largo plazo. Repásense los apartados de permanencia y accesibilidad en el capítulo de conceptos básicos. En la actualidad no existe todavía una solución universal contra este problema, pero se han desarrollado algunas técnicas probadas que tienden a optimizar la preservación digital resolviendo tanto el deterioro de los soportes como la obsolescencia tecnológica. Ninguna de estas técnicas es total; cada una resuelve una parte del problema y sólo por un cierto lapso. El problema volverá a presentarse en un futuro. Empero, usadas en conjunto han probado ser una solución suficientemente buena hasta hoy. Dichas técnicas son: recopia, migración y emulación.

La recopia, también llamada renovación, rejuvenecimiento, refresco o refrescado *–refreshing–* consiste simplemente en copiar los materiales digitales de cuando en cuando hacia otros soportes más nuevos, más “frescos” y de mayor capacidad. En esta técnica se copia la imagen de los archivos tal cual está, sin modificación alguna. Se entiende por ello mismo que los formatos internos de los documentos no cambian ni las plataformas que los operaban. Es el simple traslado desde un soporte hacia otro considerado mejor, más moderno o simplemente de mayor capacidad: por ejemplo el traslado de información de un disquete a una memoria USB, de un CD-ROM a un DVD, etc. Su objetivo primordial es resolver la permanencia evitando que los soportes de los documentos lleguen a deteriorarse físicamente, como es el caso del óxido férrico en disquetes o cintas magnéticas, así como actualizar el entorno tecnológico que permita continuar leyéndolos usando tecnologías más nuevas y que estén disponibles, como sería el caso de actualizar archivos contenidos en disquetes hacia CD ante la inexistencia actual de los lectores de aquellos soportes.

Cómo preservar mi patrimonio digital personal

Si bien no es la solución total al problema, esta técnica ha sido desde hace décadas la manera más simple de lidiar contra la reducida duración de los soportes y la obsolescencia de la tecnología que los accede –subrayo, la tecnología que los accede y lee, y no la que los opera intrínsecamente–. Bajo esta técnica se recomienda hacer el recopiado de los registros cada cinco a siete años. En contra de lo que algunos pudieran pensar, esto no implica grandes costos. De hecho, la mayor parte del costo proviene del tiempo para ocuparse del proceso; los costos del equipo y soportes necesarios prácticamente no son considerables, debido al enorme incremento del costo-capacidad comentado en el primer capítulo de esta obra. Por ejemplo, copiar disquetes a un CD: la relación de capacidad es tal, que un CD-ROM puede guardar la misma cantidad de información que 650 disquetes. Un DVD típico puede guardar cerca de 3,000 disquetes. Un disco duro actual externo portátil de 15 x 9 x 2 cm de un Terabyte de capacidad puede guardar el contenido de 250 DVD llenos. Como puede observarse, el costo mayor no proviene del cambio de los soportes, sino del tiempo para hacer la copia de todos ellos. Es muy importante que te decidas a hacer la copia mientras aún tienes ambos dispositivos: si deseas hacer la copia de tus disquetes y ya te deshiciste de la última computadora que tenía una unidad lectora de disquetes, encontrarás esta tarea muy difícil. Si posees tarjetas de memoria SD o microSD asegúrate de que transfieras su contenido a otro dispositivo más moderno antes de deshacerte de tu computadora con lector de tarjetas de este tipo si la nueva computadora no las pudiese leer. Y tarde o temprano sucederá eso. Por lo mismo, no se recomienda que hoy las transfieras, sino que estés atento al momento oportuno cuando debas hacerlo. En eso consiste la estrategia.

La segunda técnica contra la obsolescencia tecnológica es la conocida como migración. Este método, a diferencia del anterior, no se queda en un simple copiado de medios, sino que va más allá. Implica el cambio de elementos de equipo, sistemas operativos, programas o cambios generacionales de la infraestructura de cómputo. Aquí sí se cambia la tecnología que los opera intrínsecamente. Por ejemplo, cambios de versiones de los documentos tipo doc del sis-

tema operativo MS-DOS al sistema operativo Windows; documentos doc de Office XP a docx de Office 2010 o de doc de Microsoft a odt de OpenOffice. Cambio de documentos del formato de Office, o de éste a Staroffice. Bases de datos en Base de OpenOffice migradas a Access de Microsoft. Fotografías guardadas en formatos propios del fabricante de la cámara que actualmente ya no son actuales hacia otros nuevos, como el formato Kodak Photo Image PCD a formato tiff o jpg. En estos ejemplos, la copia no fue hecha sólo hacia otro medio o soporte; implicó también la transformación de formatos de los documentos así como de las plataformas que los operaban.

El propósito primordial de la migración es el de preservar la integridad de los objetos digitales manteniendo la capacidad de los usuarios de acceder a ellos a lo largo de varias generaciones tecnológicas. La migración incluye al proceso de recopiado pero difiere de aquél en que no se queda en un simple cambio de medio; la mayoría de las ocasiones implica además cambio de formatos, versiones, sistemas operativos, estándares, etc. La imagen copiada podrá diferir entonces del original en lo tocante a su estructura interna, pero no debe diferir en cuanto a su accesibilidad, es decir, la disponibilidad y usabilidad de la información, en el sentido de la capacidad o facilidad futura de que habiendo tenido permanencia, esa información pueda ser consultada, reproducida y por tanto usada nuevamente tal como lo era antes. Este proceso por lo general consume mucho más tiempo y recursos que la recopia.

La migración exitosa está definida en mayor medida por el manejo adecuado de lo que se conoce como la paradoja de la preservación digital: en forma general, tenemos la sensación de que preservar significa mantener a las cosas sin cambio. Pero en las situaciones donde el documento conlleva cierto grado de tecnología –en especial la tecnología digital– si guardamos los documentos sin modificaciones, acceder a esta información se volverá cada vez más y más difícil si no es que imposible. Ambas situaciones combinadas crean esta paradoja. Por un lado, la intención de preservar el documento tan intacto como sea posible; por el otro, la intención de mantenerlo accesible permanentemente con las mejores herramientas disponibles en ese momento. Es de vital importancia que,

Cómo preservar mi patrimonio digital personal

a pesar de los cambios introducidos por el proceso de migración, el documento pueda seguir siendo considerado íntegro; esto es, que siga siendo auténtico, no en su aspecto o soporte, sino en su esencia, en su contenido.

Uno de los secretos de la migración exitosa en lo relativo a formatos consiste en apegarse en lo posible a formatos y sistemas no propietarios, lo más estándar posible: formatos documentales como txt, PDF, doc u odt, html, xml, jpg, cdda, mp3, mpeg4, etc. En este aspecto –más que nunca– el apegarse a estándares y normas paga a la larga.

La tercera y última de las técnicas es la emulación; en ella, se pretende replicar la funcionalidad de un sistema obsoleto que ya no tenemos o que ya no funciona. Esto puede entenderse mejor si imaginamos los antiguos juegos electrónicos de video, como los originales de Atari o Nintendo. Estos pueden ser fácilmente emulados en una computadora actual del tipo personal o en una consola de hechura actual fabricada ex profeso. No es exactamente el mismo programa antiguo el que vemos en la actual computadora o consola. Es un emulador que replica el funcionamiento del anterior para que funcione y se perciba igual. Otro ejemplo para entender este concepto existe en el ícono de MS-DOS que se observa en los sistemas operativos Windows. Cuando lo utilizamos y vemos operando en la pantalla ese sistema operativo anterior, en realidad no existe como tal en la computadora; Windows se encarga de emular o replicar su funcionamiento para que su uso y percepción por parte del usuario sean semejantes a aquél.

Hoy en día, se ha propuesto la emulación como una solución de largo plazo para el manejo de documentos. Incluye una variante que consiste en que todo trozo de información se encapsule junto con el programa o aplicación que permite explotarlo. Esto todavía es experimental aunque hay ya demostraciones prácticas de que es posible. En general, no es una opción que los usuarios personales construyan, ya que se requiere de muchas habilidades computacionales para construirlas, pero se menciona porque es algo de lo que los usuarios personales pueden aprovecharse cuando está disponible para ser usado y descargado.

Como puede observarse, el problema real no es la poca duración de los soportes digitales. En realidad, no están hechos para durar. Esto no significa que su duración no sea un problema; por supuesto que lo es. El punto es que no estriba en ello el problema principal, ya que esa situación puede solucionarse con cierta facilidad con alguna de las técnicas señaladas en este apartado. El verdadero problema de la obsolescencia tecnológica como el mayor obstáculo para el acceso a la información digital en un futuro proviene de la desidia, la ignorancia y la falta de atención a esa información. Si un día la copiamos y pensamos que estará ahí para siempre estamos equivocados. El mayor pecado en este aspecto siempre es el de procrastinación. Benjamin Franklin afirmaba: “Tú puedes demorarte; el tiempo no lo hará”. Debemos estar atentos a los signos del deterioro y a sus soluciones; si aplicamos correcta y oportunamente las medidas recomendadas, éstas funcionan adecuadamente y el problema se soluciona, aunque sea sólo por unos años. Si actuamos correctamente, siempre compramos más tiempo. Pero si dejamos pasar demasiado tiempo corremos el riesgo de enfrentarnos a una tragedia documental.



9. Verifica siempre que cada respaldo haya terminado correctamente

La excesiva seguridad es la madre del peligro y la abuela de la destrucción.
Thomas Fuller

Es indispensable verificar sin falta que cada respaldo ha concluido satisfactoriamente y que el contenido ahí guardado es íntegro, completo y confiable. El hecho de dar la orden de que se haga un respaldo no implica que éste se lleve a cabo sin fallas. Mil cosas pueden salir mal y causar que un respaldo esté incompleto o defectuoso. Entre estas causas están:

- ❖ Fallas de energía que interrumpen el proceso.

Cómo preservar mi patrimonio digital personal

- ❖ Fallas en el sistema que interrumpen el proceso; por ejemplo, algunos sistemas, cuando detectan que un archivo tiene como atributo sólo lectura interrumpen el proceso de copiado para pedir confirmación al usuario de esta acción; si el usuario no confirma, el proceso se quedará interrumpido permanentemente.
- ❖ El dispositivo de respaldo: memoria USB, disco duro, etc., se llena antes de acabar de respaldar; ello interrumpe el proceso.
- ❖ En procesos de transferencia automática a servidores web remotos, si la línea de transferencia se ve interrumpida, los sistemas no se recuperan de esa interrupción, dejando trunca su tarea, en este caso, el respaldo.

Esta lista, sin ser exhaustiva, nos ilustra acerca de posibles causas de falla al momento de hacer nuestros respaldos. Lo peor que nos puede suceder es que nosotros estemos muy confiados de que nuestra información está propiamente respaldada y nos percatemos de que no es así. Por lo general lo descubrimos cuando la información principal en la computadora se ha dañado o perdido y requerimos de nuestra preciosa protección sólo para descubrir que ella también se encuentra dañada o desaparecida.

Por esta razón –que es más frecuente de lo que suele pensarse– es que debemos verificar después de cada protección que ésta se ha llevado a cabo correctamente y todo está ahí como debiera estar.

Para ello existen distintas técnicas; pueden usarse aisladamente o en combinación:

- ❖ En protecciones ordenadas manualmente, verifica que no quedan en la pantalla mensajes de preguntas parciales tales como “El archivo X es de sólo lectura; ¿estás seguro que deseas sobrescribirlo?”. Eso significa un proceso que ha sido interrumpido. Si esto te sucede con frecuencia y tienes que dar la confirmación manualmente, significa que tienes archivos o carpetas marcados como sólo lectura. Entra a ellos con la opción “Propiedades” del archivo y desactiva esa condición.
- ❖ En protecciones automáticas hechas a través de programas al efecto, utiliza siempre las opciones de “Grabar bitácora” que se encuentran disponibles en ellos. Es indispensable revisar la bitá-

cora tan pronto como se pueda; en ella se podrá observar si el proceso terminó exitosamente o fue interrumpido.

- ❖ Los sistemas de respaldo automático y el sistema operativo Linux tienen como opciones la de *parity check* o verificación de paridad. Ésta es una cuenta de bits que se hace rápida y automáticamente entre un disco, carpeta o archivo origen y sus correspondientes destinos, verificando que las cantidades sean idénticas; obviamente, en caso de haber diferencias, esto indicará que la información respaldada no corresponde con la original.
- ❖ Cuenta el número de archivos y de bytes entre las carpetas origen y destino. Todos los sistemas operativos tienen una opción que permite contar el número de carpetas, de documentos dentro de las carpetas y el número de bytes de ellos. Por ejemplo, en Windows, si se coloca el ratón sobre el nombre de una carpeta o archivo, y se oprime el botón derecho, aparece un menú; seleccionando de entre sus opciones la denominada “Propiedades” el sistema nos informará del número de archivos dentro de una carpeta y el número total de bytes que ocupa esa carpeta o archivo. Así, se puede verificar que el número de carpetas o bytes de la carpeta origen corresponda al de la carpeta destino. Esta opción puede utilizarse con todo el disco de protección, obteniendo la cantidad total de bytes contenidos, la cual deberá ser congruente con lo respaldado.
- ❖ Es recomendable que se efectúen revisiones periódicas antivirus del disco de protección, al igual que se hace con el disco duro regular de la computadora. En todos los programas antivirus se puede seleccionar que el escaneado se haga sobre el disco de protección “D” o “E” o “F” o cualquiera que sea su nomenclatura en lugar del disco regular, o en su caso, sobre ambos.

En diversos estudios de respaldos de archivos se ha establecido que una de las causas de falla más frecuente dentro de los mismos es el relajamiento del procedimiento derivado de la confianza que surge tras periodos considerables sin fallas. Entre más tiempo pase sin que tengamos incidentes de seguridad nos iremos volviendo más y más confiados, de manera que empezamos a asumir y dar por hecho

lo que debíamos verificar. Esto es casi un hecho absoluto en lo tocante a verificaciones de los respaldos: en la medida que pasen semanas y meses en los que los respaldos se han realizado correctamente tenderemos a asumir que el siguiente será igual. Verifica siempre por alguno de los métodos enunciados anteriormente que en realidad así fue. En lo relativo a respaldos completos y exitosos nunca, jamás, por ningún motivo des por hecho lo que no hayas constatado.



10. Refina permanentemente tu sistema de respaldos

Ve refinando tu sistema de respaldos de una manera gradual, continua y permanente. El ingrediente principal en el refinamiento de todo sistema o método de esta naturaleza es la paciencia. Nadie logra el sistema perfecto de respaldos, conservación y preservación desde la primera vez, por más que lo planee y lo diseñe. Es necesario que cada uno lo vaya perfeccionando poco a poco, cada vez más, hasta que funcione como un mecanismo de relojería y al gusto de cada quién. Este debe ser un proceso gradual, en donde en cada ocasión notamos que nuestro sistema de nomenclatura y respaldos se va refinando cada vez más y por lo mismo funciona mejor y va siendo cada vez más fácil y más rápido de realizar.



Fíjate metas por actividad y apégate a ellas, no trates de hacer todo a la vez. Por ejemplo, si vas a renombrar archivos de acuerdo a una nomenclatura nueva, no trates de hacerlo de golpe para toda tu computadora. Te hartarás de ello pronto. Es mejor hacerlo con una carpeta a la vez, o con un cierto tipo de material, o seleccionar primero el material más reciente, o alguna estrategia que permita dividir, variar y entreverar las actividades.

- ❖ Renombra y rearregla tus carpetas cada vez que sientas que puedes acomodarlas mejor.
- ❖ Manda las carpetas y documentos inactivos a un archivo histórico que por lo general se guarda en dispositivos externos per-

manentes y económicos –CD o DVD– y borra los archivos de tu computadora para ahorrar espacio; ello te ahorrará tiempo de respaldos.

- ❖ Evita archivos ociosos duplicados; deshazte de la basura con frecuencia.
- ❖ No llenes el escritorio de documentos ociosos por largo plazo; regresa los documentos y carpetas no utilizados a su entorno en los discos duros.
- ❖ Nunca guardes CD o DVD de respaldos sin marcarlos; es muy fácil confundirlos después. Si es una protección permanente, debe escribirse en ellos la fecha y el tipo de contenido en ellos respaldado. Si son reescribibles, deben marcarse con una leyenda que permita diferenciarlos de los demás, tal como “disco de respaldo 1”, “disco de respaldo ventas”, etc. Jamás debes ponerles etiquetas de papel encima, ya que pueden despegarse con el calor interno de la unidad lecto-grabadora y dañarla. Es mejor usar un plumón marcador al efecto. La única excepción a esta recomendación son las etiquetas totales, es decir, aquellas redondas que cubren toda la superficie del disco. No obstante, debe verificarse cuidadosamente la garantía que otorgue el fabricante de las etiquetas en este sentido. En caso de duda, mejor no usarlas.
- ❖ A propósito de discos, recuerda además que los CD o DVD dejados dentro de un auto al sol se convertirán en unos minutos en un bello cenicero totalmente ondulado.



Revisa periódicamente la eficacia del plan de conservación y preservación. Cada cierto tiempo, digamos una o dos veces al año, revisa en forma panorámica cómo ha funcionado en lo general tu plan de respaldos, conservación y preservación; los problemas hallados, las soluciones, los avances y los pendientes, los tiempos, los resultados, etc., con el fin de replantear periódicamente la estrategia y refinarla gradualmente.

Conclusiones

Cada generación humana está convencida de que es más inteligente que aquella que la precedió, y más sabia que aquella que la sucederá.

George Orwell

Como ha podido observarse, el fenómeno de la preservación documental digital, aunque es un problema eminentemente derivado del uso de las tecnologías de almacenamiento electrónico, no se limita a ser un simple problema de tecnología y su obsolescencia; su solución por tanto tampoco radica en el simple almacenamiento o recopia de documentos digitales. Es toda una escuela de pensamiento cuyo objetivo central pretende llevar los materiales documentales existentes en archivos digitales hacia el largo plazo, muy largo plazo; idealmente a perpetuidad, además de mantenerlos usables, confiables y auténticos.

Como ha podido apreciarse también su eje central implica el desarrollo y observancia meticulosa de procedimientos que comienzan con una adecuada administración documental, continúan con un adecuado mantenimiento o conservación del material digital para desembocar finalmente en una preservación digital adecuada. Muchos de estos conocimientos y habilidades provienen del entorno de las organizaciones con acervos documentales digitales: bibliotecas y archivos digitales, gobierno y empresas, mediatecas, etc., pero pueden ser trasladados a la preservación personal.

Dada la cantidad y variedad de documentación digital generada en los últimos años dentro del ámbito institucional y en especial gubernamental se han producido una serie de normatividades, recomendaciones, estándares, etc., que buscan organizar y estructurar los esfuerzos que las organizaciones deben realizar a nivel mundial para tratar de garantizar el acceso actual y futuro a los documentos digitales relevantes. No existe hoy en día un esfuerzo único que es-

Cómo preservar mi patrimonio digital personal

tablezca o defina los atributos y características ideales de todos los documentos digitales y sus procedimientos asociados con miras a su preservación a largo plazo; en la práctica, existen múltiples modelos e iniciativas. Además, estos modelos no pueden ser aplicados directamente y tal como vienen para la preservación documental personal por varias razones: la primera porque cada variedad documental tiene sus propias reglas documentales de registro, recuperación y preservación. No son iguales las estructuras, organizaciones y normatividades que rigen por ejemplo a los libros, que a los documentos de archivo gubernamentales, a las fotografías o a los videos; proceden de entidades y estructuras muy diferentes y sus modelos varían en muchos elementos. No existe una estructura documental universal que englobe a todos los tipos de documentos que ya hemos enunciado.

Además, la inmensa mayoría de esos modelos y sus especificaciones al estar hechos para organizaciones contienen muchísimas especificaciones y estándares demasiado complejos para ser seguidos y observados a nivel de los individuos. Claro que hay principios básicos, de experiencia y de sentido común que sí pueden ser extraídos de esos modelos y adaptados a las necesidades de las personas. De hecho éste es el objetivo principal de esta obra: extraer de entre todas esas variantes, modelos, iniciativas y estándares una serie de resultantes que pueden ser utilizadas a nivel de los individuos para preservar adecuadamente su patrimonio documental digital.

En primer lugar, la base para la preservación es volverse consciente de que hay un problema con el manejo de información digital y de que puede contenderse con él: si no vale la pena guardarlo, no lo guardes; si vale la pena guardarlo, vale la pena preservarlo. Ninguna técnica funcionará si manejamos nuestra información digital en una feliz ignorancia y confiando en la suerte. La mejor manera de contrarrestar ese problema es estar al tanto de que existe y de que hay una solución. Espero haber logrado este primer propósito.

En segundo lugar, se requiere de una serie de conceptos y conocimientos básicos para entender cabalmente la naturaleza del problema y actuar adecuadamente en consecuencia. No es necesario

que cada quien se convierta en un informático profesional, pero sí es indispensable contar con una base conceptual mínima, sólida y real. Mucho de lo que aprende cotidianamente y sobre la marcha el usuario común al respecto es incompleto, está basado en rumores o es información errónea o parcial. Espero haber cubierto también este mínimo de conceptos básicos.

En tercer lugar, es importante estar al tanto de la seguridad informática y aplicarla. Aunque no son lo mismo, no puede darse una buena preservación sin un mínimo de seguridad.

Finalmente y para este último propósito se han presentado una serie de recomendaciones y estrategias puntuales tratando de compilar todo lo anteriormente presentado en una metodología práctica y realizable por cualquier persona. Como ya se comentó, las recomendaciones y estrategias presentadas a lo largo de esta obra no constituyen en su conjunto ningún estándar o recomendación internacional; no se hallarán como tal en ningún sitio o referencia, pero todas y cada una de ellas proviene de algún estándar o modelo relacionado con alguna variedad documental real o de un estándar *de facto*, y todas han probado su eficacia en este propósito dentro de su ambiente. No son simples recomendaciones teóricas basadas en los libros, los eruditos y la lógica: todas son medidas de buenas prácticas, útiles y probadas para el propósito que nos interesa a este respecto.

Espero haber podido resaltar que tanto seguridad como preservación dependen en mayor manera de método y procedimiento, y sólo en una mínima parte de productos y servicios que se adquieren en el mercado. Por supuesto que estos últimos son necesarios, pero su simple adquisición e instalación será inútil sin los primeros. Respecto a las recomendaciones y estrategias puntuales, no se espera que cada persona las aplique todas y al mismo tiempo; eso sería utópico e irrealizable. Pero definitivamente son un ideal al cual debemos aspirar para estar muy cerca de una preservación adecuada y satisfactoria, sin sobresaltos ni disgustos. Sin esperar su aplicación absoluta y total, debemos aspirar a lograr un mínimo aceptable de las mismas, y a ir incrementando poco a poco, pero inexorablemente, su número e intensidad. Si sólo ponemos en

Cómo preservar mi patrimonio digital personal

práctica unas cuantas cosas de inicio, sin lograr un todo aceptable y armónico, con un mínimo de completitud y funcionalidad, sin ir completando sus efectos y sin seguimiento, sólo estaremos empeorando las cosas, simulando que tenemos una estrategia de preservación sin estar ni siquiera cerca de implementarla: "...un poco de instrucción es peligrosa. Bebe copiosamente, o mejor no pruebes del manantial de la sabiduría".⁷¹

Como puede apreciarse también, la preservación documental digital no es un simple método para mantener cadenas de bits legibles y utilizables mañana; tiene una fuerte connotación cultural y de utilidad social. La preservación del patrimonio cultural de cada persona es un elemento crucial de las identidades personales, regionales y nacionales y es a su vez enlace de una persona con su comunidad y de ésta con su pasado. Las sociedades actuales, supuestamente de la información y el conocimiento, deben aprovechar y preservar su patrimonio cultural para el futuro. A pesar de la dificultad y los costos de preservar no es conveniente ni recomendable soslayar y posponer el problema. Todos los seres y conglomerados humanos, todas las regiones y países poseen riquezas documentales que necesariamente hay que preservar y distribuir. Forman en esencia su patrimonio cultural, y la información digital no es la excepción, sobre todo ahora con su enorme desarrollo. Nosotros, como parte de ese conglomerado humano tenemos una gran parte de la responsabilidad de hacerlo. Debe recordarse siempre que, a la larga –socialmente hablando– no preservar resultará más caro.

Una última reflexión: en el apartado de las definiciones se mencionó que esta obra trataría de la preservación de archivos bajo la acepción de computación: unidad lógica de información, generalmente indivisible, y no bajo el concepto de la archivística: la suma o conjunto integrado de documentos producidos o separados que participan en el mismo asunto o están relacionados con un mismo evento, persona, lugar, proyecto o materia, agregado de tal forma que pueda ser recuperado para una acción o como referencia. Son

71 Diálogo entre los bomberos Beatty y Montag en *Fahrenheit 451*, de Ray Bradbury.

elaborados y recibidos por una persona física o jurídica en el desarrollo de actividades y preservados.

Hemos tratado a lo largo de esta obra los archivos bajo la primer acepción, y por lo mismo los conceptos, recomendaciones, metodologías, etc., enunciados en esta obra son perfectamente aplicables para todos los lectores personales. Sin embargo, algunos pocos de ellos tendrán en sus manos archivos que se parecen a la segunda acepción, a la de la ciencia archivística: bien porque manejan expedientes de clientes o pacientes, o porque tienen conjuntos documentales digitales históricos importantes heredados de algún personaje o situación, o alguna otra situación parecida. En este caso, su manejo y preservación adecuada requiere además de otras técnicas archivísticas que no han sido tratadas aquí, pero que son indispensables para ese propósito, y deberán ser por tanto buscadas en otro lado y aplicadas por esos lectores: la forma documental y sus contextos, la forma fija y contenido estable, la clasificación y sus esquemas, la valoración, los requisitos de autenticidad, la fiabilidad, entre muchos otros elementos archivísticos deberán ser añadidos a las técnicas de preservación presentadas en esta obra. Si ése es el caso, para iniciar la búsqueda de esa información adicional el Anexo 6 será de particular ayuda por la información ahí presentada, en especial el modelo de la cadena de preservación.

Concluyo presentando nuevamente la pregunta introducida en el capítulo “La naturaleza de los documentos digitales”:

¿Cuánto valen nuestras fotografías y videos, nuestras grabaciones, nuestras notas, tareas, apuntes, tesis e investigaciones, nuestros correos electrónicos, los expedientes de nuestros pacientes, los datos o contabilidades de nuestros clientes, y toda esa información valiosa que hemos acumulado con tesón y paciencia por años?

Espero que el lector pueda dar ahora una respuesta. Yo me atrevo a dar una muy personal: ahora lo sé, valen mucho; y por lo mismo, vale la pena preservarlos.

*Forsan et hæc olim meminisse juvabit*⁷²

72 “Quizá algún día valga la pena recordar todo esto”.

GLOSARIO

ALA. *American Library Association* o Asociación Americana de Bibliotecas.

ancho de banda. La capacidad que tienen canales y bandas de transmisión de datos de recibir o transmitir información digital por unidad de tiempo. Por lo general se mide en MBPS o Megabits por segundo; es decir, cuántos millones de bits de información puede recibir o transmitir esa banda o canal en un segundo.

aplicación. Pieza de *software*, programa informático desarrollado para solucionar una necesidad específica de cierto tipo de usuario. Es distinto a los lenguajes de programación –que sirven para crear aplicaciones–, de los sistemas operativos –que controlan un cierto tipo de equipo– y de las utilerías –que realizan tareas específicas de mantenimiento o ambiente–. Existen aplicaciones simples o de primer nivel, las cuales realizan una tarea sencilla: por ejemplo, reproductor de sonido, compresor/descompresor de archivos, calculadora, editor de imágenes, navegador, etc. Existen también los conjuntos o *suites* de aplicaciones o de segundo nivel: por ejemplo los paquetes ofimáticos que contienen varias aplicaciones integradas tales como procesador de texto, de imágenes, hoja de cálculo, presentador, base de datos, etc., todo bajo un mismo conjunto o *suite*.

ASCII o American Standard Code for Information Interchange.

Código Estándar Americano para Intercambio de Información o ASCII. Código binario utilizado en la inmensa mayoría de las computadoras para representar los caracteres usados en todos los alfabetos occidentales descendientes del alfabeto latino: mayúsculas y minúsculas, dígitos, vocales con diacríticos, letras especiales de estos idiomas, tales como la “ñ”, la “ç”, la “ß”; caracteres especiales, tales como: “?”, “!”, “&”, “\$”, “%”, “@”, etc. El uso de este código permitió el uso estandarizado e interoperable de textos entre distintas marcas de computadoras, módems, impresoras, etc. Originalmente, cada carácter ASCII estaba compuesto por 7 bits, lo que daba 128 combinaciones distintas. En la actualidad se usa el ASCII extendido de 8 bits o UTF-8 lo cual permite

256 combinaciones. Se ha conformado como el estándar ISO-Latin 8 dentro del estándar mundial de caracteres Unicode.

cadena de bits. Datos digitales codificados en una secuencia no estructurada de bits o dígitos binarios y que son transmitidos, almacenados o recibidos como una unidad.

cadena de bytes –cadena de caracteres–. Secuencia de bits cuyos datos –bits o dígitos binarios– están agrupados en unidades estructuradas de longitud fija y predeterminada llamadas bytes o caracteres y tienen un valor específico de acuerdo a una tabla –letra, número, signo–.

CEN. Acrónimo de *Comité Européen de Normalisation* o *European Committee for Standardization* o Comité Europeo de Normalización.

compatibilidad. La capacidad de diferentes dispositivos o sistemas –programas, formatos de archivo, protocolos, lenguajes de programación, etc.– para trabajar juntos o intercambiar datos sin necesidad de modificación previa. *Véase también:* “multiplataforma”, “interoperabilidad”.

compatibilidad retrospectiva. La capacidad de un sistema o programa de compartir datos e instrucciones con versiones anteriores del mismo, o con otros sistemas y programas a los que ha sustituido o pretende sustituir. Algunas veces esta compatibilidad se limita a la capacidad de leer datos anteriores pero no se extiende a la capacidad de reescribir estos datos en formatos que puedan ser leídos por versiones anteriores.

compresión. Recodificación de datos digitales para reducir sus dimensiones y poder así ahorrar espacio de almacenamiento o tiempo de transmisión.

compresión con pérdida. Método de compresión de archivos de datos o imágenes usado para reducir su espacio; con esta técnica se logra un factor de compresión de archivos muy alto, pero el documento descomprimido ha perdido algo de calidad con respecto al original previo a su compresión, ya que reduce la cantidad de información dentro de los datos.

compresión sin pérdida. Método de compresión de archivos de datos o imágenes usado para reducir su espacio; con esta técnica

el documento descomprimido es idéntico al original previo a la compresión. Su factor de compresión no es muy alto.

disco compacto o CD. Medio de almacenamiento digital consistente en un disco plástico, grabado por un solo lado, capaz de almacenar cerca de 700 Megabytes de datos digitales sobre una pista espiral microscópica, continua, la cual es leída ópticamente gracias a un rayo láser. Sus especificaciones y formatos están establecidos en los estándares ISO 9660 y 13490.

disco duro. También conocido como “disco rígido”; es un dispositivo de almacenamiento de datos no volátil que usa un método de grabación magnética para almacenar datos digitales. Está formado por uno o varios platos o discos metálicos rígidos cubiertos con un material ferromagnético montados sobre un eje común que giran a gran velocidad dentro de una carcasa sellada. Sobre cada plato, y en cada una de sus caras, se sitúa un cabezal de lectoescritura para leer o grabar los datos.

disco óptico. Medio de almacenamiento digital consistente en un disco plástico, recubierto con una superficie brillante especial, en el cual los datos son grabados como bits: 1 y 0. Los hay del tipo no reescribible o fijo, en donde el grabado se hace en forma de pequeños orificios sobre la superficie del disco, sobre una pista microscópica; y del tipo reescribible, en donde el grabado se hace con reorientaciones opto-magnéticas de pequeñas partes de su superficie. Ambos son leídos ópticamente gracias a un rayo láser y sus señales son decodificadas por un microprocesador hacia impulsos digitales eléctricos procesables por una computadora.

disposición. Destrucción final de documentos de archivo o transferencia hacia otro archivo según se ha preestablecido en su “tabla de retención”.

estándar *de facto*. Un estándar que no ha sido emitido por ningún organismo oficial dedicado a ello, sino más bien impuesto por el uso y aceptación generalizados por parte de una comunidad.

estándar *de jure* –o *de iure*–. Un estándar emitido por algún organismo oficial dedicado y autorizado para la emisión de ellos. Pueden ser nacionales, como NOM, ANSI; multinacionales como CEN o internacionales, como ISO.

Cómo preservar mi patrimonio digital personal

HTML. Acrónimo de *Hypertext Markup Language* o Lenguaje de Marcado de Hipertextos.

IFLA. *International Federation of Library Associations and Institutions* o Federación Internacional de Asociaciones de Bibliotecarios e Instituciones.

ISO. Acrónimo de *International Organization for Standardization* u Oficina Internacional de Estándares.

interoperabilidad. La capacidad que tienen equipos, programas y sistemas para comunicarse y trabajar conjuntamente con otros sin ajustes o cambios especiales.

lenguaje de marcado. Un sistema de codificación legible por máquina así como sus reglas asociadas que son utilizados para describir la estructura lógica, distribución, forma de despliegue y estilo de un cierto documento digital. Existen varios lenguajes de marcado con diferentes reglas, propósitos y alcances; por ejemplo, el HTML, el SGML y el XML.

Lenguaje de Marcado de Hipertextos. *Hyper Text Markup Language*, o simplemente HTML. Es un lenguaje de marcado de textos muy simple que se utiliza para crear textos y demás elementos de una página web. Está compuesto por etiquetas –marcas o *tags*– que definen la estructura y el formato de cada uno de los elementos que componen el documento que verá el usuario a través de la página web. Esas etiquetas son leídas por todos los navegadores o *browser* web permitiendo que puedan ser visibles en cualquier computadora de forma homogénea además de permitir el hipertexto.

Lenguaje de Marcado Estándar Generalizado. *Standard Generalized Markup Language* o SGML. Lenguaje de marcado estándar internacional *ISO 8879:1986* utilizado para la definición formal de todo tipo de documentos de forma tal que los hace ser independientes del dispositivo, sistema y programa con el cual estos documentos fueron realizados.

Lenguaje de Marcado Extendido. *eXtensible Markup Language* o XML. Lenguaje de marcado estándar internacional desarrollado por el World Wide Web Consortium o W3C. XML es una versión de SGML, diseñado especialmente para los documentos de la

web. Permite que los diseñadores creen sus propias etiquetas, permitiendo la definición, transmisión, validación e interpretación de datos entre programas y entre organizaciones. XML.

malware o malicious software. “Código maligno” o “código mal intencionado”. Programas cuyo objetivo es el de infiltrarse en los sistemas de cómputo sin conocimiento de su dueño, con objeto de causar algún tipo de daño o perjuicio al comportamiento del sistema y por tanto a la información de una persona u organización. Aquí se encuentran todos los virus, “troyanos”, “espías”, etcétera.

mapa de bits. Representación digital de una imagen que consiste en una rejilla matricial o *raster* compuesta de puntos o celdas arreglados en líneas y columnas, cada uno de ellos representado por un conjunto de datos numéricos que determinan la posición y el valor del color de cada punto o pixel y que en combinación crean una impresión visual de líneas, sombras o matices específicos dentro de una imagen mono o policromática en la pantalla de una computadora o sobre una hoja impresa. También se le denomina imagen rasterizada, gráfico rasterizado o imagen matricial.

medio analógico. Soporte físico, tal como papel, pergamino, piedra, arcilla, película o los antiguos tipos de cintas de audio y video magnéticas, usadas para almacenamiento de datos en forma analógica.

medio digital. También conocido como soporte digital. Es el material físico, tal como un disco compacto, DVD, cinta o disco duro usado como soporte para almacenamiento de datos digitales.

metadato. Información que caracteriza o describe a otro recurso de información, especialmente con propósito de documentar, describir, preservar o administrar ese recurso. Cualquier archivo o base de datos que guarde información acerca de documentos, documentos de archivo, agregación de ellos, o cualquier otra de sus estructuras, así como sus atributos, procesos y cambios.

MBPS. Megabits por segundo. La capacidad que tienen equipos, dispositivos, canales y bandas de recibir o transmitir información digital por unidad de tiempo. Es decir, cuántos millones de bits de información puede recibir o transmitir en un segundo.

MTBF. *Mean Time Between Failures* o Tiempo Medio Entre Fallas.

Media aritmética o promedio de ocurrencia medido en un conjunto de fallas de un cierto dispositivo, medido en horas.

multiplataforma. La capacidad de sistemas y programas de cómputo de poder ejecutarse y percibirse en forma idéntica en distintas marcas y modelos de computadoras; ello se logra generalmente gracias a la utilización de normas, productos, formatos abiertos y estándares.

no propietario. Se llama así a las tecnologías de equipo, programas y aplicaciones de cómputo o formatos de archivos que no se encuentran protegidos por una patente o marca o que no son poseídos ni controlados por una sola compañía o institución o cuyo uso es permitido bajo esquemas de acceso abierto.

plataforma. La combinación específica de un cierto tipo de equipo de cómputo y su sistema operativo la cual le otorga características y comportamientos definidos y preestablecidos; suele ser similar entre modelos o familias de computadoras.

propietario. Régimen de uso de los programas, formatos, estructuras y otras herramientas que están protegidas bajo una patente u otro registro de propiedad industrial que pertenece a una empresa u organización y cuyo uso y licenciamiento está restringido, por lo general bajo pago de derechos, y cuya fuente o tecnología no está disponible al público y no puede ser modificada.

retención. Tiempo durante el cual un documento de archivo es conservado para su utilización dentro de la organización y que es preestablecido en una tabla de retención.

soporte. Material físico o substancia sobre la cual la información se registra o almacena. También se le conoce como medio o medio escriptorio.

Unicode. Código estándar mundial para representar con un código binario de 16 bits todos los caracteres de texto de todos los idiomas existentes y vivos en la actualidad, con miras a ser utilizados de forma estándar en las computadoras y dispositivos actuales.

XML. Acrónimo de *eXtensible Markup Language* o Lenguaje de Marcado Extendido. Es un lenguaje de marcado de textos diseñado para la web que ofrece un formato para la descripción estruc-

turada de datos. Esto facilita hacer declaraciones de contenido más precisas y obtener resultados de búsquedas más significativos en varias plataformas. Además habilita una nueva generación de aplicaciones para usar y manipular datos basadas en la web.

REFERENCIAS BIBLIOGRÁFICAS

Todas las referencias electrónicas han sido verificadas como existentes y exactas hasta el 1 de junio del 2012.

- Alpert, Jesse y Nissan Hajaj (2008), “We knew the web was big...”, en *Blog Oficial de Google* [en línea], <http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html> [Consulta: 25 de julio, 2008].
- American Library Association (ALA) (1988), *The ALA Glossary of Library and Information Science*, Chicago, American Library Association.
- Audio Engineering Society. Sitio web oficial: <http://www.aes.org>
- Bandić, Zvonimir, Dmitri Litvinov y M. Rooks (2008), “Nanostructured Materials in Information Storage”, en *MRS Bulletin* [en línea], vol. 33, septiembre, pp. 831-837. <http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=7961153>
- Comunidad Europea del Acero y el Carbón (CECA), Comunidad Económica Europea (CEE) y Comunidad Europea de la Energía Atómica (CEEA) (2001), *MoReq: Modelo de requisitos para la gestión de documentos electrónicos: especificación MoReq* [en línea], Bruselas, Luxemburgo. <http://www.mcu.es/archivos/docs/moreq.pdf> y tiene una actualización a MoReq2 en <http://www.moreq2.eu/moreq2>
- Consultative Committee for Space Data Systems (CCSDS) (2002), *Reference Model for an Open Archival Information System (OAIS)*, Washington, D.C., CCSDS Secretariat.
- Delgado Gómez, Alejandro (2011), *El Documento Electrónico en la Sociedad de la Información*, México, Archivo General de la Nación, 43 pp. (Cuadernos de Archivística, 1).
- Digital Curation Centre/DigitalPreservationEurope, *DRAMBORA. Digital Repository Audit Method Based on Risk Assessment*. Sitio web oficial: <http://www.repositoryaudit.eu/>
- Duranti, Luciana y Kenneth Thibodeau (2005), “The Concept of Record in Interactive, Experiential and Dynamic Environments: The view of InterPARES”, en *Archival Science*, Springer Netherlands,

Cómo preservar mi patrimonio digital personal

- vol. 5, núm. 2-4, diciembre, pp. 13-68. ISSN:1389-0166 (Print) 1573-7519 (Online). DOI 10.1007/BF02660804.
- Ganz, John F. (2007), *The Expanding Digital Universe. A Forecast of Worldwide Information Growth Through 2010* [en línea], International Data Corporation (IDC). <http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf>
- Glosario para la IME ICC (IFLA Meeting of Experts on an International Cataloguing Code/Reunión IFLA de Expertos sobre un Código Internacional de Catalogación)* (2004) [en línea], versión en español de Ageo García-Barbabosa, abril. http://deposit.d-nb.de/ep/netpub/76/22/36/991362276/_data_dyna/_snap_stand_2008_08_05/standardisierung/pdf/glossary_april2004_spanish_latinamerican_version.pdf
- Granger, Sarah (2009), *Social Engineering Fundamentals. Part I: Hacker Tactics* [en línea], Security Focus. <http://www.securityfocus.com/infocus/1527>
- Gulli, Antonio y Alessio Signorini (2005), *The Indexable web is more than 11.5 billion pages* [en línea], <http://www.cs.uiowa.edu/~assignori/web-size/>
- “Hacker Camps Train Network Defenders”, en *The Wall Street Journal* [en línea], http://online.wsj.com/article/SB120700735637678619.html?mod=yahoo_hs&ru=yahoo
- “How Big is the web?” (2008), en *The Raw Feed* [en línea], <http://www.therawfeed.com/2008/02/how-big-is-web-155583825-sites-report.html>
- International Council on Monuments and Sites (ICOMOS) (1931), *The Athens Charter for the Restoration of Historic Monuments* [en línea], http://www.icomos.org/athens_charter.html
- _____ (1964), *The Venice Charter. International Charter for the Conservation and Restoration of Monuments and Sites* [en línea], http://www.icomos.org/charters/venice_e.pdf
- International Data Corporation (IDC) (2010), *The 2011 IDC Digital Universe Study* [en línea], <http://www.emc.com/collateral/about/news/idc-emc-digital-universe-2011-infographic.pdf>
- International Federation of Library Associations (IFLA) (1998), *Principios para el cuidado y manejo de material de bibliotecas*

- [en línea], Edward P. Adcock, Marie-Thérèse Varlamoff y Virginie Kremp (eds.), Programa “Core” en Preservación y Conservación (PAC) de la Federación Internacional de Asociaciones e Instituciones de Bibliotecas (IFLA) y la Comisión de Preservación y Acceso, p. 7. Disponible en: <http://www.slideshare.net/Sandra-Haro/principios-para-cuidado-y-manejo-de-bibliotecas-ifla>
- International Federation of Library Associations and Institutions (IFLA) (2007), *Functional Requirements for Bibliographic Records. Final Report* [en línea], IFLANET Publications. http://www.ifla.org/VII/s13/frbr/frbr_current_toc.htm
- International Organization for Standardization (ISO) (1998) ISO/IEC Moving Picture Experts Group (MPEG) (ISO/IEC JTC1/SC29/WG11), formalmente estándar ISO/IEC 14496-Codificación de objetos audiovisuales.
- _____ (2004), ISO/IEC 15444-2:2004, Information technology. JPEG 2000 Image Coding System: Extensions [en línea], Final Committee Draft (FCD). <http://www.jpeg.org/public/fcd15444-2.pdf>
- _____ (2005), *ISO 19005-1:2005. Document management. Electronic document file format for long-term preservation. Part 1: Use of PDF 1.4 (PDF/A-1)* [en línea], http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=38920
- InterPARES Project (2004), *Business-Driven Recordkeeping (BDR) Model* [en línea], http://www.interpares.org/ip2/ip2_models.cfm
- _____ (2008), *InterPARES 2. Experiential, Interactive and Dynamic Records. Appendix 16. Overview of the Records Continuum Concept. Extracted and adapted from “Xiami An. An Integrated approach to Records Management”. Information Management Journal. July/August (2003): 24-30* [en línea], http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_appendix_16.pdf
- _____ (2008), *Preserver Guidelines. Preserving Digital Records: Guidelines for Organizations* [en línea], http://www.interpares.org/display_file.cfm?doc=ip2%28pub%29preserver_guidelines_booklet.pdf
- _____ (2010), *The International Research on Permanent Authentic Records in Electronic Systems/Guía del Creador Personal. Crea-*

- ción y Conservación de Materiales Digitales* [en línea], <http://www.interpares.org>, <http://www.iibi.unam.mx/archivistica>
- _____ (2010), *The International Research on Permanent Authentic Records in Electronic Systems. Glosario de Preservación de Archivos Digitales* [en línea], <http://www.interpares.org>, <http://www.iibi.unam.mx/archivistica>
- _____ (2012), *The International Research on Permanent Authentic Records in Electronic Systems. Glosario InterPARES de Archivística* [en línea] http://www.interpares.org/ip2/display_file.cfm?doc=ip2_dictionary.pdf&CFID=259655&CFTOKEN=35461332 y en <http://www.iibi.unam.mx/archivistica>
- Joint Photographic Experts Group (JPEG). Sitio web oficial: <http://www.jpeg.org>
- _____ (2002), ISO/IEC 15444-2:2004, Information Technology. JPEG 2000 Image Coding System: Code Coding System [en línea], Final Committee Draft (FCD). <http://www.jpeg.org/public/fcd15444-1.pdf>
- Juels, Ari (2008), “Password Expiration: Like Margarine and Water?”, en *Speaking of Security. The Official RSA Blog and Podcast* [en línea], <http://blogs.rsa.com/juels/password-expiration-like-margarine-and-water/>
- Komorowski, Matthew (2009), *A History of Storage Cost* [en línea], <http://www.mkomo.com/cost-per-gigabyte>
- Lessig, Lawrence (2001), *The Future of Ideas: The Fate of the Commons in a Connected World*, New York, Random House. ISBN 0-375-50578-4.
- Ley Federal de Protección de Datos en Posesión de Particulares* [en línea], <http://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>
- Library of Congress Preservation Website (1999), *Selection Criteria for Preservation Digital Reformatting* [en línea], <http://www.loc.gov/preservation/about/prd/presdig/pressselection.html>
- Library of Congress. *Sustainability of Digital Formats* [en línea], <http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml>
- Lyman, Peter y Hal Varian (2000), “How Much Information?”, en *Journal of Electronic Publishing* [en línea], vol. 6, núm. 2, di-

- ciembre. <http://www.press.umich.edu/jep/06-02/lyman.html>
ISSN 1080-2711
- _____ (2003), “How Much Information?”, en el sitio web oficial de la Universidad de California en Berkeley, EUA <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>
- Mandelbrot, Benoit (1977 [2003]), *La geometría fractal de la naturaleza*, Barcelona, Tusquets, 662 pp. ISBN: 978-84-8310-549-8
- Markoff, John (2005), “How Big is the web?”, en *Web Design & Technology News* [en línea], <http://www.webdesignsnow.com/news/081505a.html>
- Microsoft (s. a.), *XML para Principiantes* [en línea], <http://office.microsoft.com/es-mx/word-help/xml-para-principiantes-HA010034022.aspx>
- Monash University, *Records Continuum Research Group* [en línea], <http://www.infotech.monash.edu.au/research/groups/rcrg/>
- Morrison, Alan *et al.* (1998), “Creating and Documenting Electronic Texts. Chapter 3: Digitization. Scanning, OCR, and Re-keying”, en *AHDS Guides to Good Practice* [en línea], <http://ota.ahds.ac.uk/documents/creating/cdet/chap3.html> ISSN: 1463-5194
- National Archives of Australia (2007), *Functional Specifications for Electronic Records Management Systems Software (ERMS) y Guidelines for Implementing the Functional Specifications for ERMS* [en línea], <http://www.naa.gov.au/records-management/publications/ERMS-specs.aspx>
- Organización para la Cooperación y el Desarrollo Económico (OCDE) (2002), *Directrices de la OCDE para la Seguridad de Sistemas y Redes de Información: hacia una cultura de Seguridad* [en línea], París. <http://www.oecd.org/dataoecd/15/29/34912912.pdf>
- Organization for the Advancement of Structured Information Standards (OASIS) (2005), *OASIS Standards and Other Approved Work* [en línea], <http://www.oasis-open.org/specs/index.php>
- Portable Network Graphics (PNG) Specification (Second Edition)*. Information technology. Computer graphics and image processing. Portable Network Graphics (PNG): Functional specification. ISO/IEC 15948:2003 (E), (2002), David Duce (ed.) [en línea], World Wide Web Consortium. <http://www.w3.org/TR/PNG/>

Cómo preservar mi patrimonio digital personal

- RLG/OCLC Working Group on Digital Archive Attributes (2002), *Trusted Digital Repositories: Attributes and Responsibilities* [en línea], Mountain View, CA, Research Libraries Group (RLG). <http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf>
- RLG-NARA Digital Repository Certification Task Force (2007), *Trustworthy repositories audit & certification: Criteria and checklist* [en línea], <http://www.crl.edu/PDF/trac.pdf>
- Sayood, Khalid (2000), *Introduction to Data Compression*, 2a. ed., San Francisco, Morgan Kaufmann. ISBN: 1-55860-558-4.
- Shannon, Claude (1948), "A Mathematical Theory of Communication", en *The Bell System Technical Journal*, Bell Laboratories, vol. 27, pp. 379-636.
- University of Cornell (1938), "Timeline of Digital Preservation" [en línea], <http://www.library.cornell.edu/iris/tutorial/dpm/timeline/index.html>
- Voutssas M., Juan (2009), *Preservación del patrimonio documental digital en México*, México, UNAM, Centro Universitario de Investigaciones Bibliotecológicas, 207 pp. ISBN: 978-607-02-0583-5.
- _____ (2010), "Preservación documental digital y seguridad informática", en *Investigación Bibliotecológica*, México, UNAM, CUIB, vol. 24, núm. 50, enero-abril, pp. 127-155.
- Washington State Department of Information Services (2009), *Security and Social Networks* [en línea], <http://www.governor.wa.gov/media/SecurityandSocialNetworks.pdf>
- Webb, Colin *et al.* (2003), *Directrices para la preservación del patrimonio digital* [en línea], documento CI-2003/WS/3, UNESCO, Information Society Division, Biblioteca Nacional de Australia. <http://unesdoc.unesco.org/images/0013/001300/130071s.pdf>
- Wyner, Aaron D. (s. a.), *The Significance of Shannon's Work* [en línea], <http://cm.bell-labs.com/cm/ms/what/shannonday/work.html>
- Ziv, Jacob y Abraham Lempel (1977), "A Universal Algorithm for Sequential Data Compression", en *IEEE Transactions on Information Theory*, vol. IT-23, núm. 3, mayo, pp. 337-343.

ANEXOS

Anexo 1

Cómo contar múltiplos de bytes

Existen dos maneras de contar los bytes o caracteres con fines de almacenamiento. La primera es una manera simplificada y es la más aceptada en el medio, en donde se utilizan potencias de 10, para facilidad de las personas en general.

- 1 Kilobyte = 1,000 = 10^3 bytes o caracteres
- 1 Megabyte = 1'000,000 = 10^6 bytes o caracteres
- 1 Gigabyte = 1000 Megabytes = 1'000,000,000 = 10^9 bytes o caracteres
- 1 Terabyte = 1000 Gigabytes = 1'000,000 Megabytes = 1'000,000,000,000 = 10^{12} bytes o caracteres
- 1 Petabyte = 1000 Terabytes = 1'000,000 Gigabytes = 1'000,000,000,000,000 = 10^{15} bytes o caracteres
- 1 Exabyte = 1000 Petabytes = 1'000,000 Terabytes = 1'000,000,000,000,000,000 = 10^{18} bytes o caracteres
- 1 Zettabyte = 1000 Exabytes = 1'000,000 Petabytes = 1'000,000,000,000,000,000,000 = 10^{21} bytes o caracteres
- 1 Yottabyte = 1000 Zettabytes = 1'000,000 Exabytes = 1'000,000,000,000,000,000,000,000 = 10^{24} bytes o caracteres

Existe una segunda manera, menos utilizada, la cual surge al considerar que en realidad las potencias de los bytes no son potencias del número 10, sino potencias del número 2 ($2^{10} = 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 1024$), por tanto el número múltiplo es 1,024, el cual por simplificación fue “redondeado” a kilos de 1000 en vez de 1,024 que es el múltiplo exacto. El dato exacto al contar en esta forma es:

- 1 Kilobyte (KB) = $2^{10} = 1,024$ Bytes
- 1 MegaByte (MB) = $2^{20} = 1,024$ KB (KiloBytes)
- 1 GigaByte (GB) = $2^{30} = 1,024$ MB (Megabytes)
- 1 TeraByte (TB) = $2^{40} = 1,024$ GB (Gigabytes)
- 1 PetaByte (PB) = $2^{50} = 1,024$ TB (TeraBytes)
- 1 ExaByte (EB) = $2^{60} = 1,024$ PB (PetaBytes)
- 1 ZettaByte (ZB) = $2^{70} = 1,024$ EB (ExaBytes)
- 1 YottaByte (YB) = $2^{80} = 1,024$ ZB (ZettaBytes)

Anexo 2

Producción mundial estimada de contenidos originales, almacenados digitalmente, usando técnicas estándar de compresión, estimado en terabytes, para 2002 ⁷³			
Medio de almacenamiento	Tipo de contenido	Terabytes por año, estimado superior	Terabytes por año, estimado inferior
Papel	Libros	39	8
	Periódicos	138.4	27.7
	Revistas	6	1.3
	Documentos de oficina	1397.5	279.5
	Revistas masivas, etc.	52.9	10.2
	Subtotal	1633.8	326.7
Película	Fotografías	375,000	37,500
	Cine	6,078	12
	Rayos-X	20,000	20,000
	Video y televisión	19,176	19,175
	Subtotal	420,254	74,202
Óptico	CD musicales	58	6
	CD de datos	1.1	1.1
	DVD	43.8	43.8
	Subtotal	102.9	50.9
Magnético	Cintas audio/video	1,428,800	1,428,800
	Discos de PC	1,986,080	403,080
	Cinta digital	250,000	250,000
	Otros	1,284,430	1,284,430
	Subtotal	4,999,230	3,416,230
Total		5,421, 221	3,490,810

1 Terabyte = 1'000 Gigabytes = 1'000,000 Megabytes = 1'000,000,000 Kilobytes = 1'000,000,000,000 bytes

73 Peter Lyman y Hal Varian (2003), "How Much Information?" [en línea].

Anexo 3

El sistema binario

Hay 10 clases de personas en el mundo: las que entienden binario y las que no.

El usuario común y cotidiano de equipos e información no necesita entender el sistema binario para poder hacer uso de ellos, pero todos los que han tratado de entenderlo y dedican un rato a hacerlo por lo general acaban fascinados con el mismo, o al menos, aprenden a verlo de otra manera y a sentir respeto y admiración por todo el universo que puede construirse con simples y humildes “unos” y “ceros”. Precisamente para el usuario común y cotidiano ha sido escrito este anexo, si bien de una manera somera para hacerlo en unos pocos párrafos

Todo sistema numérico posicional permite un cierto conjunto de dígitos –es decir, números que pueden ser expresados con un sólo guarismo– así como reglas matemáticas para realizar las operaciones básicas con ellos. Al número de signos o guarismos permitidos en un sistema se le llama la “base” del sistema numérico. El sistema numérico decimal, con el cual estamos familiarizados la mayoría de los seres humanos, basa su estructura en el número diez, de ahí el nombre, debido a que los seres humanos tenemos diez dedos y con ellos se contaba originalmente; de hecho, *digitus* significa “dedo” en latín. De acuerdo con esto, el sistema decimal se basa en las siguientes premisas:

1. Dado que su base es el número diez, debe tener exactamente diez dígitos o guarismos que por sí mismos, sin combinaciones, representen un valor numérico contenido en ellos. Así, tenemos los guarismos 0, 1, 2, 3, 4, 5, 6, 7, 8, 9; diez signos que en sí mismos contienen el valor que representan: cero unidades, una unidad, dos unidades..., nueve unidades.

Cómo preservar mi patrimonio digital personal

2. Una vez que se agota la representación con un sólo guarismo, se recorre la unidad –el “uno”– una posición a la izquierda y se guarda la primera posición con ayuda del cero; esto es, del 9 sigue el 10, el cual ya es una combinación de dos guarismos, y decimos que hemos creado una posición de segundo orden con la ayuda del cero. Pero, si son un “uno” y un “cero” ¿por qué decimos que vale diez? Porque el “uno”, al estar recorrido una posición hacia la izquierda se multiplica por la base del sistema numérico, en este caso el número diez por ser sistema numérico decimal. En realidad los dígitos “10” nos dicen que tenemos la suma de un “uno” que vale diez por estar recorrido una posición a la izquierda y un “cero” que vale cero por estar en la posición original. En valores de posición esto sería:

Valor de la columna	10	1
Número	1	0

Un uno que se multiplica por diez más un cero que se multiplica por cero; total: diez.

Para continuar, iremos sustituyendo en la posición original todos y cada uno de los guarismos básicos, esto es, después del 1 con el 0, seguiremos con el 1 con el 1, el 1 con el 2, el 1 con el 3, y así sucesivamente hasta agotar toda la lista de guarismos originales, es decir, el 1 con el 9; sus valores son: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19. ¿Por qué decimos que dos “unos” contiguos, “11”, es el número once, si son dos “unos”? Porque por su posición, el uno de la izquierda vale diez, y el uno de la derecha vale uno; esto es:

Valor de la columna	10	1
Número	1	1

Un uno que se multiplica por diez más un uno que se multiplica por uno; total: once. Si seguimos con esta secuencia cambiando el guarismo de la extrema derecha veremos que el uno que vale diez

no cambia su valor; sólo los guarismos que están en la posición original. Así tendremos, diez más dos igual a doce, diez más tres igual a trece, y así sucesivamente hasta llegar a diecinueve.

Una vez agotadas estas combinaciones del número uno seguimos con el número 2, el cual sigue en la misma posición y se multiplica por diez; esto es:

Valor de la columna	10	1
Número	2	0

Un dos que se multiplica por diez más un cero que se multiplica por cero; total: veinte. Si seguimos con esta secuencia veremos que el dos que vale diez no cambia su valor; sólo los guarismos que están en la posición original. Así tendremos, veinte más uno, veintiuno; veinte más dos igual a veintidós, y así sucesivamente hasta llegar a veintinueve. Esta secuencia se repetirá con el tres, el cuatro, hasta llegar al nueve, el cual se combinará con todos los guarismos hasta formar el 99, noventa y nueve. Nótese que aunque son dos nueves un nueve se multiplica por diez y vale noventa y el otro por uno, por lo que vale nueve. Son dos nueves pero no valen lo mismo.

Agotadas las combinaciones de dos números lo que procede es recorrer el uno nuevamente un lugar a la izquierda guardando su lugar con dos ceros a la derecha, así, el número que sigue es 100, cien, combinación de tres dígitos. ¿Por qué cien si es un uno con dos ceros?

Valor de la columna	100	10	1
Número	1	0	0

La respuesta es conocida: porque ahora ese uno, al estar recorrido dos posiciones, se multiplica por cien, los dos ceros se multiplican por diez y por uno, pero porque son ceros el resultado es cero; resultado: cien.

Cómo preservar mi patrimonio digital personal

Seguiremos haciendo las combinaciones 101, 102, 103, etc. –primero las de la extrema derecha– y luego las de las posiciones siguientes sucesivamente, así, tendremos números como “111” y decimos “ciento once”. ¿Por qué ciento once si son tres unos? Porque por su posición, un uno que vale cien más un uno que vale diez más un uno que vale uno; total: ciento once. De esta forma tendremos combinaciones como 345 y decimos “trescientos cuarenta y cinco”: un tres que se multiplica por cien más un cuatro que se multiplica por diez más un cinco que se multiplica por uno: trescientos más cuarenta más cinco: trescientos cuarenta y cinco. Podemos seguir con esta secuencia infinitamente recorriendo el uno a la siguiente posición y multiplicando los valores subsecuentes; así, por ejemplo, 5874 sería:

Valor de la columna	1000	100	10	1
Número	5	8	7	4

$$5000 + 800 + 70 + 4 = 5874$$

Esto también puede ser expresado en potencias del número base, el diez; así:

Valor de la columna en potencias del 10	10^3	10^2	10^1	10^0
Número	5	8	7	4

O lo que es lo mismo: $(5 \times 10^3) + (8 \times 10^2) + (7 \times 10^1) + (4 \times 10^0)$

$$10^3 = 10 \times 10 \times 10 = 1000$$

$$10^2 = 10 \times 10 = 100$$

$$10^1 = 10 = 10$$

$$10^0 = 1 \text{ (cualquier número elevado a la potencia "0" es igual a 1)}$$

Las fracciones siguen exactamente la misma lógica: las cifras a la derecha del punto decimal se multiplican ahora por un décimo, un centésimo, un milésimo, un diezmilésimo y así sucesivamente hasta

el infinito. De esta forma, por ejemplo, .3271 se lee “tres mil doscientos setenta y un diezmilésimos” o lo que es lo mismo, la suma de tres décimos más doscientos centésimos, más setenta milésimos más un diezmilésimo. Lo que casi nadie se fija es que esto sigue siendo el corrimiento de potencias del 10, en este caso negativas:

Valor de la columna en potencias negativas del 10	Lugar del punto decimal	10^{-1}	10^{-2}	10^{-3}	10^{-4}
Número	.	3	2	7	1

O lo que es lo mismo: $(3 \times 10^{-1}) + (2 \times 10^{-2}) + (7 \times 10^{-3}) + (1 \times 10^{-4})$

$10^{-1} = 1/10^1 = 1/10 =$ un décimo

$10^{-2} = 1/10^2 = 1/100 =$ un centésimo

$10^{-3} = 1/10^3 = 1/1000 =$ un milésimo

$10^{-4} = 1/10^4 = 1/10000 =$ un diezmilésimo

Hasta aquí está claro cómo se construye el sistema decimal, el cual de todos modos ya conocíamos. Pero éste no es el único sistema numérico posible; de hecho, con la ayuda del cero, se puede construir un sistema numérico basado en el número ocho –octal–, que tendría sólo ocho guarismos, o un sistema basado en el número dieciséis –hexadecimal–, simplemente inventando dieciséis guarismos. O con casi cualquier otro número base, exceptuando el uno. Partiendo de aquí, luego entonces es posible poder construir un sistema numérico cuya base sea el número 2; esto es, un sistema binario.

La regla sería la misma: ya que la base es “2” deben existir exactamente dos guarismos que representan en sí mismos un valor, el “0” y el “1”. Una vez que se agotan los guarismos debemos empezar combinaciones recorriendo el “1” una posición a la izquierda y guardando este lugar con un “0”; esto es, la combinación “10”; sólo que aquí, dado que la base es 2 el número que se recorre no se multiplica por diez (la base ahora no es diez), sino por dos, que es la base actual.

En valores de posición esto sería:

Cómo preservar mi patrimonio digital personal

Valor de la columna	2	1
Número	1	0

Esto es: un uno que se multiplica por dos más un cero que se multiplica por uno, resultado: 2; y sí, éste es el número “dos” en binario. Seguirán las combinaciones de los guarismos; en este caso la combinación que sigue es “1” con “1”, “11”; son dos “unos” pero es un uno que vale dos más un uno que vale uno, total: tres. Agotadas las combinaciones de dos nuevamente recorreremos el “1” una posición a la izquierda, guardando su posición con “ceros”; así: 100 es: “cuatro”. ¿Por qué cuatro si es un “uno” con dos ceros?

Valor de la columna	4	2	1
Número	1	0	0

Porque es “uno” por cuatro más “cero” por dos más “cero” por uno.

Nótese que en el sistema decimal, el uno de la primera posición se multiplicaba por diez al pasar a la segunda = diez; éste por diez al pasar a la tercera = cien, y así sucesivamente –la base era diez–. En el sistema binario, el número “uno” original se multiplicará ahora por dos al pasar a la segunda columna = 2, éste por dos al pasar a la tercera = 4, éste por dos al pasar a la cuarta posición después = 8, y así sucesivamente –la base es dos–. De esta forma, la secuencia de las posiciones ya no es 1, 10, 100, 1000, 10000, etc., sino 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024 y así hasta el infinito. Si lo ponemos en potencias del 2, sería:

$$2^0 = 1 \text{ (cualquier número elevado a la potencia “0” es igual a 1)}$$

$$2^1 = 2$$

$$2^2 = 2 \times 2 = 4$$

$$2^3 = 2 \times 2 \times 2 = 8$$

Así, los números del uno al veinte en binario serían:

0, 1, 10, 11, 100, 101, 110, 111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111, 10000, 10001, 10010, 10011, 10100...

Nótese que en binario todos los números pares terminan en “cero” y los números nones terminan en “uno”.

Cualquier sistema numérico, sin importar su base, debe permitir las cuatro operaciones aritméticas básicas: suma, resta, multiplicación y división. En el sistema decimal, sabemos que esto es posible; analicemos la suma:

paso 1. columna de las unidades + $\begin{array}{r} 5682 \\ 3253 \\ \hline 5 \end{array}$

paso 2 columna de las decenas + $\begin{array}{r} 5682 \\ 3253 \\ \hline 35 \end{array}$ +Llevamos

paso 3 columna de las centenas + $\begin{array}{r} 5682 \\ 3253 \\ \hline 935 \end{array}$

paso 4 columna de las unidades de millar + $\begin{array}{r} 5682 \\ 3253 \\ \hline 8935 \end{array}$

Pasos de una suma, columna por columna, en sistema decimal

En este ejemplo en sistema decimal, en la segunda columna, la suma es 13; no podemos escribir trece en una sola columna, así que tomamos 10 unidades, las juntamos en una “decena” que “llevamos” a la siguiente posición de la izquierda, y escribimos el residuo en la presente columna; en este caso, el residuo es “tres”; escribimos tres y “llevamos una”; nótese que en realidad nos llevamos diez, sólo que agrupados en una decena; es decir, “llevamos una decena”.

Véase ahora un ejemplo de suma en sistema binario; en este caso sumaremos los números 85 + 116.

Cómo preservar mi patrimonio digital personal

paso 1.
columna 1

$$\begin{array}{r} 1010101 \\ + 1110100 \\ \hline 1 \end{array}$$

paso 2
columna 2

$$\begin{array}{r} 1010101 \\ + 1110100 \\ \hline 01 \end{array}$$

paso 3
columna 3

+Llevamos

$$\begin{array}{r} 1010101 \\ + 1110100 \\ \hline 001 \end{array}$$

paso 4
columna 4

$$\begin{array}{r} 1010101 \\ + 1110100 \\ \hline 1001 \end{array}$$

paso 5
columna 5

+Llevamos

$$\begin{array}{r} 1010101 \\ + 1110100 \\ \hline 01001 \end{array}$$

paso 6
columna 6

+Llevamos

$$\begin{array}{r} 1010101 \\ + 1110100 \\ \hline 001001 \end{array}$$

paso 7
columna 7

+Llevamos

$$\begin{array}{r} 01010101 \\ + 01110100 \\ \hline 1001001 \end{array}$$

paso 8
columna 8

$$\begin{array}{r} 01010101 \\ + 01110100 \\ \hline 11001001 \end{array}$$

Pasos de una suma, columna por columna, en sistema binario.
En este caso $85 + 16$; el resultado es, obviamente, 201

Nótese que en sistema binario, la lógica de la suma será equivalente a la decimal. En el sistema binario, cuando sumamos $0 + 0 = 0$; $1 + 0 = 1$; $0 + 1 = 1$, no hay problema. El inconveniente comienza al sumar $1 + 1$; la suma es dos, obviamente, pero no podemos escribir “dos” con un sólo guarismo en sistema binario; ¿qué se hace? Agrupamos las dos unidades en un par, el cual “llevaremos” a la siguiente posición a la izquierda; el residuo es “cero”, que escribimos en la presente columna; así, $1 + 1$ en binario es un par que me “llevo” a la columna de la izquierda (llevo uno) y el residuo es 0 que escribo en la presente columna. Nótese que me “llevo 1” que vale dos –o un par– y no “1” que vale una decena, ya que en este sistema al correrse la posición a la izquierda vale el doble y no diez veces dada la base. Cuando se suman tres “unos”: $1 + 1 + 1$ agrupamos un par –“llevamos uno”– y el residuo es 1, que escribimos en la presente columna.

No es el caso ponerse aquí a hacer demostraciones matemáticas, pero afirmamos que si hay suma, hay resta, ya que es el proceso inverso. Además, si hay suma, debe haber multiplicación, ya que esta es una serie de sumas sucesivas, y si hay multiplicación, habrá división, ya que éste es el proceso inverso de la multiplicación. En el sistema decimal sabemos que todo esto sí es posible, pues lo hemos hecho muchas veces. En el sistema binario también son posibles todas las demás operaciones; ya hemos ejemplificado la suma. No abundaré más al respecto.

¿Qué hay acerca de las fracciones? ¿Pueden construirse como en el sistema decimal? La respuesta es sí, desde luego; para la construcción de las fracciones se sigue exactamente la misma lógica. Veamos la construcción de ellas en decimal; para ello, usemos como ejemplo el número 3701:

Valor de la columna en potencias negativas del 10	Lugar del punto	10^{-1}	10^{-2}	10^{-3}	10^{-4}
Número	.	3	7	0	1

De acuerdo con el álgebra básica: $10^{-n} = \frac{1}{10^n}$

10^{-1} es igual que $1/10^1 = 1/10 = 0.1 =$ “un décimo”.

Y: 10^{-2} es igual que $1/10^2 = 1/100 = 0.01 =$ “un centésimo”.

Siguen “cero” milésimos y un “diezmilésimo”.

Y así sucesivamente hasta el infinito.

Por tanto “.3701” se descompone en:

$(3 \times 10^{-1}) + (7 \times 10^{-2}) + (0 \times 10^{-3}) + 1 \times 10^{-4} =$ tres por un décimo más siete por un centésimo más cero por un milésimo más uno por un diezmilésimo. O lo que es lo mismo:

$.3000 + .0700 + .0000 + .0001 = .3701$

En el sistema binario la lógica es exactamente igual, sólo que cada cifra a la derecha del punto decimal, en vez de ser un décimo

Cómo preservar mi patrimonio digital personal

de la anterior, esto es: un décimo, un centésimo, un milésimo, etc., es ahora la mitad de la anterior: un medio, un cuarto, un octavo, un dieciseisavo, un treintadosavo, etc., y así sucesivamente hasta el infinito. Es decir, ahora en vez de multiplicar cada guarismo por potencias negativas del número diez, se multiplica cada uno de ellos por potencias negativas del número dos. Usemos como ejemplo el número .1101 en binario.

Valor de la columna en potencias negativas del 2	Lugar del punto	2^{-1}	2^{-2}	2^{-3}	2^{-4}
Número	.	1	1	0	1

De acuerdo con el álgebra básica: $2^{-n} = \frac{1}{2^n}$

2^{-1} es igual que $1/2^1 = 1/2 =$ “un medio”.

Y: 2^{-2} es igual que $1/2^2 = 1/4 =$ “un cuarto”.

Y así sucesivamente hasta el infinito.

De esta forma, por ejemplo, “.1101” en binario sería la suma de un medio más un cuarto más cero octavos más un dieciseisavo o dicho en sistema decimal sería la suma de $0.5 + 0.25 + 0 + 0.0625$; por lo tanto .01101 binario es igual a 0.8125 decimal. En potencias negativas del 2:

Valor de la columna en potencias negativas del 2	Lugar del punto	2^{-1}	2^{-2}	2^{-3}	2^{-4}
Número	.	1	1	0	1

O lo que es lo mismo $(1 \times 2^{-1}) + (1 \times 2^{-2}) + (0 \times 2^{-3}) + (1 \times 2^{-4})$

$2^{-1} = 1/2^1 = 1/2 =$ un medio

$2^{-2} = 1/2^2 = 1/(2 \times 2) =$ un cuarto

$2^{-3} = 1/2^3 = 1/(2 \times 2 \times 2) =$ un octavo

$2^{-4} = 1/2^4 = 1/(2 \times 2 \times 2 \times 2) =$ un dieciseisavo

Y así, hasta el infinito. Pero ¿por qué hasta la fecha todo el mundo digital se basa en el sistema binario? ¿Por qué construir máquinas en este sistema si es más complicado para los seres humanos acostumbrados al decimal? La respuesta es sencilla y sobre todo práctica: entre los miles de sistemas numéricos posibles, el binario es el más pequeño en su base –no puede haber sistema numérico base 1 o base 0, ya que no pueden construirse con esta lógica–. En binario, sus dos guarismos tienen lo que se llama una componente tautológica, es decir, representan redundantemente en sí mismos a su naturaleza: un “1” implica “es”, “existe”; un “0” implica “no es”, “no existe”; pueden hacerse equivalencias perfectas: “positivo o negativo”, “prendido o apagado”, “ying o yang”, “sí o no”, absolutamente simples, sin ambigüedades. Por lo mismo, pueden hacerse equivalencias eléctricas igual de simples: hay corriente o no la hay; magnéticas: hay magnetismo o no lo hay; ópticas: hay luz o no la hay. Esto llevó al concepto del bit, *binary digit* o dígito binario, 1 o 0.

Esta simplicidad permite que los dispositivos tecnológicos que manejan estos números se simplifiquen enormemente en su construcción: si pensamos en una computadora por ejemplo, que debiese representar los objetos con números decimales, al ser una máquina que se basa en electricidad, requeriría de nueve generadores de voltaje diferentes para representar los números decimales, de tal forma que un milivoltio fuera un 1, dos milivoltios fueran un 2, etc.; requeriría de 9 generadores distintos de voltaje, más “no hay voltaje” para el cero: muy complicado. En cambio, una computadora –desde sus inicios hasta hoy– requiere de un solo generador eléctrico en toda la máquina para poder generar todos sus “unos”; los “ceros” serán la ausencia de corriente eléctrica; un reloj interno sincroniza todo.

Para representar la información dentro de un disco duro magnético, la cabeza grabadora va magnetizando de cierta forma puntos sobre la superficie del disco –sus “unos”– y dejará sin magnetizar donde vaya un “cero”: simple. Si esto se quisiera hacer con números decimales, requeriríamos un décimo de magnetismo para un uno, tres décimos de magnetismo para un tres, ocho décimos de magnetismo para un ocho, etc. Demasiado complicado para construirse en la práctica.

Cómo preservar mi patrimonio digital personal

Además, en caso de desajuste de un dispositivo interno –digamos el generador de voltaje– produciría serios errores. En caso de un desajuste de un 10% del generador de voltajes, por ejemplo, 3 milivoltios más 8 milivoltios se convertirían en $3.3 + 8.8$ milivoltios = 12.1 milivoltios. Es decir, la suma de $3 + 8$ daría 12, debido al desajuste del 10%. En el caso de un generador de voltaje basado en sistema binario, 1.1 milivoltios + 1.1 milivoltios, la suma de dos “unos” desajustados en 10%, serían 2.2 milivoltios; la máquina sólo sabe que le llegó un impulso eléctrico más otro impulso eléctrico, y seguiría siendo la suma de dos “unos”; el desajuste no genera error.

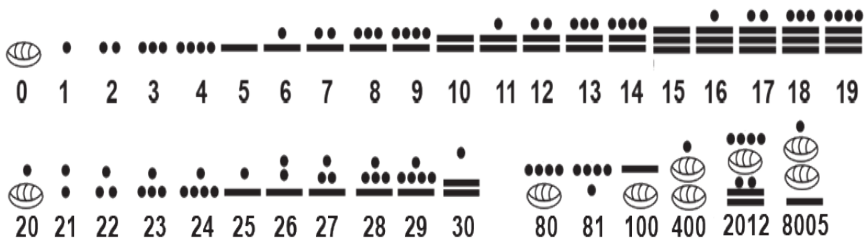
Si pensamos en la información grabada ópticamente sobre un CD o un DVD, como ya explicamos en el apartado “Dispositivos y documentos digitales”, el láser envía luz sobre la superficie del disco, y obtiene sus unos y ceros registrando si hay reflejo de la luz o no. Tratar de hacer esto con diez proporciones decimales de reflejos como “hay luz de intensidad uno”, “hay luz de intensidad cinco”, etc., sería tremendamente complicado.

Como puede verse, a pesar de su aparente complejidad, el sistema numérico binario en realidad ha simplificado enormemente la construcción de los dispositivos digitales modernos: computadoras, teléfonos, tabletas y miles más representando digitalmente todo tipo de datos e información.

Al principio de este apartado mencionamos que todo sistema numérico posicional permite un cierto conjunto de dígitos, es decir, números que pueden ser expresados con un sólo guarismo. Si el número de dígitos coincide con la base del sistema numérico y existe el “cero” toda la estructura posicional de conteo, las operaciones aritméticas, etc., serán posibles. Es decir, un sistema numérico base diez o decimal con diez guarismos, el sistema base dos o binario con dos guarismos, el sistema base ocho u octal con ocho guarismos, etc. Todos funcionarán sin importar la base seleccionada, como hemos establecido. No obstante, es posible construir y de hecho existen sistemas numéricos que tienen el “cero” pero el número de guarismos no coincide con su base numérica. Este es el caso del sistema numérico maya, el cual era base veinte, pero tenía sólo tres guarismos –“caracol”, “punto” y “raya”– para representar

todos los números; el caracol es el cero. Con este tipo de sistemas numéricos se puede construir y contar cualquier número hasta el infinito, pero no se pueden realizar las operaciones aritméticas básicas. Para los mayas, cuyo objetivo era establecer calendarios y cuentas cortas y largas de los días, esto fue suficiente.

Finalmente, si en un principio no entendiste la cita con la que comienza este anexo, es que no sabías el sistema binario, y pensaste que el número ahí consignado era un “diez”, lo cual quita todo contexto a la cita. Ahora que has leído este anexo, vuelve a leerla. Si notas que los guarismos “10” están escritos en binario, y que en realidad significan “dos” descubrirás instantáneamente el contexto la cita y ahora perteneces a esa clase de personas en el mundo que sí entienden el sistema numérico binario.



Ejemplo de numeración maya; posicional, pero con sólo tres guarismos.

Anexo 4

Tabla ASCII 0-63

Valor decimal	Número binario	Carácter	Valor decimal	Número binario	Carácter
0	00000000	Nulo	32	00100000	Espacio en blanco
1	00000001	Principio de encabezado	33	00100001	!
2	00000010	Principio de texto	34	00100010	"
3	00000011	Fin de texto	35	00100011	#
4	00000100	Fin de la transmisión	36	00100100	\$
5	00000101	¿Listo para envío?	37	00100101	%
6	00000110	Listo	38	00100110	&
7	00000111	Campana	39	00100111	'
8	00001000	Espacio hacia atrás	40	00101000	(
9	00001001	Tabulador horizontal	41	00101001)
10	00001010	Nueva línea	42	00101010	*
11	00001011	Tabulador vertical	43	00101011	+
12	00001100	Nueva página	44	00101100	,
13	00001101	Enter, retorno del carro	45	00101101	-
14	00001110	Sangría fuera	46	00101110	.
15	00001111	Sangría activa	47	00101111	/
16	00010000	Salir del enlace	48	00110000	0
17	00010001	Control dispositivo 1	49	00110001	1
18	00010010	Control dispositivo 2	50	00110010	2
19	00010011	Control dispositivo 3	51	00110011	3
20	00010100	Control dispositivo 4	52	00110100	4
21	00010101	No estoy listo	53	00110101	5
22	00010110	Sincronizar	54	00110110	6
23	00010111	Fin del bloque	55	00110111	7
24	00011000	Cancelar	56	00111000	8
25	00011001	Fin del medio	57	00111001	9
26	00011010	Sustituir	58	00111010	:
27	00011011	Escape	59	00111011	;
28	00011100	Separador de archivo	60	00111100	<
29	00011101	Separador de grupo	61	00111101	=
30	00011110	Separador de registro	62	00111110	>
31	00011111	Separador de campo	63	00111111	?

Tabla ASCII 64-159

Valor decimal	Número binario	Carácter	Valor decimal	Número binario	Carácter	Valor decimal	Número binario	Carácter
64	01000000	@	96	01100000	`	128	10000000	Ç
65	01000001	A	97	01100001	a	129	10000001	ü
66	01000010	B	98	01100010	b	130	10000010	é
67	01000011	C	99	01100011	c	131	10000011	â
68	01000100	D	100	01100100	d	132	10000100	ä
69	01000101	E	101	01100101	e	133	10000101	à
70	01000110	F	102	01100110	f	134	10000110	â
71	01000111	G	103	01100111	g	135	10000111	ç
72	01001000	H	104	01101000	h	136	10001000	ê
73	01001001	I	105	01101001	i	137	10001001	ë
74	01001010	J	106	01101010	j	138	10001010	è
75	01001011	K	107	01101011	k	139	10001011	ï
76	01001100	L	108	01101100	l	140	10001100	î
77	01001101	M	109	01101101	m	141	10001101	ì
78	01001110	N	110	01101110	n	142	10001110	Ä
79	01001111	O	111	01101111	o	143	10001111	Å
80	01010000	P	112	01110000	p	144	10010000	É
81	01010001	Q	113	01110001	q	145	10010001	æ
82	01010010	R	114	01110010	r	146	10010010	Æ
83	01010011	S	115	01110011	s	147	10010011	ô
84	01010100	T	116	01110100	t	148	10010100	ö
85	01010101	U	117	01110101	u	149	10010101	ò
86	01010110	V	118	01110110	v	150	10010110	û
87	01010111	W	119	01110111	w	151	10010111	ù
88	01011000	X	120	01111000	x	152	10011000	ÿ
89	01011001	Y	121	01111001	y	153	10011001	Ö
90	01011010	Z	122	01111010	z	154	10011010	Ü
91	01011011	[123	01111011	{	155	10011011	ø
92	01011100	\	124	01111100		156	10011100	£
93	01011101]	125	01111101	}	157	10011101	Ø
94	01011110	^	126	01111110	~	158	10011110	ž
95	01011111	_	127	01111111	△	159	10011111	ÿ

Tabla ASCII 160-255

Valor decimal	Número binario	Carácter	Valor decimal	Número binario	Carácter	Valor decimal	Número binario	Carácter
160	10100000	á	192	11000000	Ł	224	11100000	Ó
161	10100001	í	193	11000001	Á	225	11100001	ß
162	10100010	ó	194	11000010	⌈	226	11100010	Ô
163	10100011	ú	195	11000011	⌋	227	11100011	Ò
164	10100100	ñ	196	11000100	—	228	11100100	õ
165	10100101	Ñ	197	11000101	‡	229	11100101	Ö
166	10100110	ª	198	11000110	ã	230	11100110	µ
167	10100111	º	199	11000111	Ã	231	11100111	þ
168	10101000	¿	200	11001000	ℒ	232	11101000	ƒ
169	10101001	©	201	11001001	℞	233	11101001	Ú
170	10101010	¬	202	11001010	⊥	234	11101010	Û
171	10101011	½	203	11001011	⌞	235	11101011	Ü
172	10101100	¬	204	11001100	⌗	236	11101100	ý
173	10101101	¡	205	11001101	=	237	11101101	Ý
174	10101110	«	206	11001110	‡	238	11101110	-
175	10101111	»	207	11001111	⌘	239	11101111	´
176	10110000	⋮	208	11010000	ð	240	11110000	
177	10110001	⋮	209	11010001	Ð	241	11110001	±
178	10110010	⋮	210	11010010	Ê	242	11110010	=
179	10110011		211	11010011	Ë	243	11110011	¾
180	10110100	‡	212	11010100	È	244	11111000	¶
181	10110101	Á	213	11010101	ı	245	11110101	§
182	10110110	Â	214	11010110	í	246	11110110	÷
183	10110111	Ã	215	11010111	î	247	11110111	¸
184	10111000	©	216	11011000	ï	248	11111000	°
185	10111001	‡	217	11011001	Ĵ	249	11111001	ˆ
186	10111010		218	11011010	ƒ	250	11111010	·
187	10111011	¶	219	11011011	■	251	11111011	¹
188	10111100	¶	220	11011100	■	252	11111100	º
189	10111101	¢	221	11011101	ı	253	11111101	²
190	10111110	¥	222	11011110	ì	254	11111110	³
191	10111111	ŕ	223	11011111	■	255	11111111	

Anexo 5

Ejemplo del Texto legal de una licencia *Creative Commons* con Atribución-No Comercial-Licenciamiento Recíproco 2.5 (México) o CC-BY-NC-SA México 2.5

Código Legal⁷⁴

Atribución-No Comercial-Licenciamiento Recíproco 2.5 (México)

Creative Commons no es un despacho de abogados ni proporciona ningún tipo de servicio legal. La distribución de la presente licencia no crea ninguna relación de tipo cliente-abogado. *Creative Commons* proporciona la presente información “tal cual está.” *Creative Commons* no garantiza la información aquí proporcionada y se deslinda de cualquier responsabilidad por cualquier daño que resulte del uso de la misma.

Licencia

La obra (como se define posteriormente) se distribuye bajo los términos y condiciones de la presente licencia pública de *Creative Commons* (“CCPL” o “licencia”). La obra está protegida por la *Ley Federal del Derecho de Autor* y por cualquier otra ley que resulte aplicable. Cualquier uso distinto del autorizado por la presente licencia o por la ley del derecho de autor está prohibido.

Se entiende que por el mero ejercicio de cualquiera de los derechos aquí previstos sobre la obra, usted acepta y se obliga bajo

⁷⁴ La versión electrónica de este código está disponible en creativecommons.org/licenses/by-nc-sa/2.5/mx/legalcode

los términos y condiciones de la presente licencia. El licenciante le otorga los derechos aquí descritos considerando la aceptación por su parte de dichos términos y condiciones.

1. Definiciones

Obra Colectiva. Sin perjuicio de lo dispuesto por la *Ley Federal del Derecho de Autor* (LFDA), una obra colectiva bajo esta Licencia, es toda obra, como lo sería una publicación periódica, antología o enciclopedia, en la cual la Obra, en su totalidad y sin modificación alguna, y que junto con otras contribuciones que de igual manera constituyen obras independientes en sí mismas, integran una obra por sí sola. Para efectos de esta Licencia, una Obra Colectiva constituida de la manera descrita anteriormente no será considerada como una Obra Derivada (como se define posteriormente).

Obra Derivada. Sin perjuicio de lo dispuesto por la LFDA, una obra derivada, bajo esta Licencia, es toda obra que se basa en la Obra o en la Obra y otras obras pre-existentes, tales como una traducción, arreglo musical, dramatización, adaptación ficticia o cinematográfica, grabación de audio o producción de fonograma, reproducción artística, compendio o cualquier otra forma en la cual la Obra sea modificada, transformada o adaptada, a menos que constituya una Obra Colectiva de acuerdo a los términos antes descritos, en cuyo caso no será considerada, para efectos de esta Licencia, como una Obra Derivada. A fin de evitar cualquier duda, para efectos de esta Licencia, cuando la Obra en cuestión se trate de una composición musical o fonograma, la sincronización de dicha Obra con imágenes en movimiento regidas por una relación de tiempo será considerada como Obra Derivada.

Licenciante es el individuo o entidad que ofrece la Obra de conformidad con los términos y condiciones de la presente Licencia.

Autor Original. es la persona física que ha creado la obra.

Obra. Obra es toda aquella creación intelectual, susceptible de protección bajo la LFDA, ofrecida bajo los términos y condiciones de la presente Licencia.

Usted. “Usted” se refiere a cualquier persona física o moral que, sin haber infringido previamente ninguno de los términos y condiciones de la presente Licencia en relación a la Obra, ejerza los derechos concedidos en la misma o que, aun cuando previamente haya existido alguna violación de derechos, haya obtenido autorización expresa por parte del Licenciante para ejercerlos.

Elementos de la Licencia. Se refiere a los siguientes atributos o características de alto nivel que el Licenciante elija e indique en el título de esta Licencia: Atribución, No Comercial y Licenciamiento Recíproco.

2. Derechos de uso legítimo

Nada de lo dispuesto en la presente Licencia tiene por objeto reducir, limitar o restringir ninguno de los derechos previstos como de uso legítimo, surgidos de la doctrina norteamericana de primer venta o de cualquier otra limitación o excepción a los derechos patrimoniales concedidos al usuario de una obra por la LFDA o por cualquier otra ley aplicable.

3. Otorgamiento de la Licencia

De conformidad con los términos y condiciones previstos en la presente Licencia, el Licenciante le otorga a Usted una licencia universal, gratuita, no exclusiva y perpetua (entendiéndose que este término es determinado por la duración prevista en la LFDA para el otorgamiento de la protección intelectual específica) para ejercer los derechos descritos a continuación:

- a. Reproducir la Obra, incorporarla a una o más Obras Colectivas y reproducirla como parte de una Obra Colectiva;
- b. Crear y reproducir Obras Derivadas;
- c. Distribuir copias o fonogramas de la Obra, representar y exhibir públicamente la Obra y representarla públicamente por medios digitales de transmisión de audio, incluyendo el caso en el que lo antes descrito sea incorporado a una Obra Colectiva.

Cómo preservar mi patrimonio digital personal

- d. Distribuir copias o fonogramas de Obras Derivadas, representar y exhibir públicamente Obras Derivadas y representarlas públicamente por medios digitales de transmisión de audio.

Los derechos antes descritos podrán ser ejercidos en cualquier formato o medio conocido o por conocer. Los derechos descritos incluyen el derecho a realizar las modificaciones tecnológicamente pertinentes para el ejercicio de los mismos en cualquier formato o medio. Cualquier otro derecho no expresamente mencionado en la presente Licencia se entiende reservado por el autor, incluyendo, de manera enunciativa mas no limitativa, los derechos establecidos en las cláusulas 4(e) y 4 (f).

4. Restricciones

Los derechos conferidos en la sección 3 de la presente Licencia están sujetos a las siguientes restricciones:

- a. Usted puede distribuir, exhibir públicamente, representar o ejecutar públicamente o representar o ejecutar públicamente de manera digital la Obra únicamente bajo los términos de la presente Licencia, y siempre y cuando incluya una copia de la presente Licencia o se haga referencia a la dirección electrónica del sitio en donde se encuentre la misma en cada una de las copias o fonogramas de la Obra que Usted distribuya, exhiba públicamente, represente o ejecute públicamente o represente o ejecute públicamente de manera digital. Le queda prohibido establecer o imponer cualquier término o condición que altere o restrinja los términos de la presente Licencia o los derechos aquí conferidos. De igual manera, le queda prohibido sub-licenciar la Obra. Usted deberá mantener intactas todas las disposiciones relativas a la presente Licencia, incluyendo la referente a la exclusión de garantía. Le queda prohibido distribuir, exhibir públicamente, representar o ejecutar públicamente o representar o ejecutar públicamente de manera digital la Obra incluyendo medidas tecnológicas que controlen el acceso a la Obra o el uso de la misma de forma incompatible con los términos de la presente Licencia. Lo descrito anteriormente aplica de

igual manera al caso en el que la Obra se incorpore a una Obra Colectiva, sin que esto requiera que la Obra Colectiva en sí misma se sujete a los términos de la presente Licencia. Si Usted crea una Obra Colectiva, deberá, en la medida de lo posible, remover de dicha Obra Colectiva cualquier mención a que se hace referencia en la cláusula 4(d) de la presente Licencia en cuanto cualquiera de los Licenciantes le haga la notificación pertinente. Si Usted crea una Obra Derivada, deberá, en la medida de lo posible, remover de dicha Obra Derivada cualquier mención a que se hace referencia en la cláusula 4(d) de la presente Licencia en cuanto cualquiera de los Licenciantes le haga la notificación pertinente.

- b. Usted puede distribuir, exhibir públicamente, representar o ejecutar públicamente o representar o ejecutar públicamente de manera digital una Obra Derivada únicamente bajo los términos de la presente Licencia, bajo los términos de una versión posterior de la misma que contenga los mismos Elementos de la Licencia o bajo los términos de una Licencia *Creative Commons iCommons* que contenga los mismos Elementos de la Licencia (i.e. Attribution-Non Commercial- Share Alike 2.5 de Japón). Usted debe incluir una copia de la presente Licencia o hacer referencia a la dirección electrónica del sitio en donde la misma se encuentre o de cualquier otra licencia especificada según la oración anterior, en cada una de las copias o fonogramas de cada Obra Derivada que Usted distribuya, exhiba públicamente, represente o ejecute públicamente o represente o ejecute públicamente de manera digital. Usted no puede ofrecer ni imponer términos ni condiciones sobre las Obras Derivadas que de alguna manera alteren o restrinjan los términos y condiciones descritos en la presente Licencia o los derechos otorgados a los usuarios por la misma. De igual manera, Usted debe mantener intactas todas las disposiciones relativas a la presente Licencia, incluyendo las referentes a la exclusión de garantía. Le queda prohibido distribuir, exhibir públicamente, representar o ejecutar públicamente o representar o ejecutar públicamente de manera digital la Obra incluyendo medidas tecnológicas que controlen el acceso a o el uso de la Obra de forma incompatible con los términos de la presente Li-

Cómo preservar mi patrimonio digital personal

encia. Lo descrito anteriormente aplica de igual manera al caso en el que la Obra Derivada se incorpore a una Obra Colectiva, sin que esto requiera que la Obra Colectiva en sí misma, distinta de la Obra Derivada incorporada a la Colectiva, se sujete a los términos de la presente Licencia.

- c. Le queda prohibido ejercer los derechos conferidos en la cláusula 3 descrita anteriormente si el propósito principal de dicho ejercicio es la obtención de una ventaja comercial o compensación monetaria personal o lucro. El intercambio de la Obra por otras obras protegidas por los derechos de autor por medios digitales de *file-sharing* (compartidas como archivos digitales) u otros similares no se considerará que se hace con el propósito de generar ventajas comerciales ni compensaciones monetarias personales, siempre y cuando no haya ningún pago o compensación monetaria relacionada con dicho intercambio.
- d. Si Usted distribuye, exhibe públicamente, representa o ejecuta públicamente o representa o ejecuta públicamente de manera digital la Obra, cualquier Obra Derivada u Obra Colectiva deberá mantener intactas todas y cada una de las menciones al Derecho de Autor de la Obra y mencionar, en la medida de lo razonablemente posible en relación al medio o medios utilizados: (i) el nombre del Autor Original (o el pseudónimo si ese es el caso) si éste es conocido o proporcionado, y (ii) el nombre de la(s) otra(s) persona(s) a quienes el Autor Original o el Licenciante (tal como un instituto patrocinador, casa editorial o revista de investigación) le atribuyan cierto crédito; el título de la Obra si éste se provee; la dirección electrónica que el Licenciante mencione como asociada a la Obra, en el caso en el que ésta exista y en la medida de lo razonablemente posible, a menos que dicha dirección electrónica no haga referencia a la información del Derecho de Autor pertinente o de la Licencia que regule la distribución de la Obra; y en el caso de una Obra Derivada, la mención que especifique el uso de la Obra dentro de la Obra Derivada (tal como “Traducción al Francés de la Obra... del Autor Original...” o “Guión cinematográfico basado en la Obra original... del Autor Original...”). En el caso de que se trate de una

Obra Derivada o de una Obra Colectiva, dicho crédito o reconocimiento será otorgado de manera razonable, donde los demás reconocimientos de autoría aparezcan y en la misma manera en la que estos aparezcan.

- e. A fin de evitar dudas, cuando la Obra se trate de una composición musical:
- Regalías por la representación de Obras bajo Licencias Colectivas. El Licenciante se reserva el derecho exclusivo a recaudar, ya sea de manera individual o a través de sociedades de gestión colectiva, regalías por concepto de la representación o ejecución pública o de la representación o ejecución pública de tipo digital (tal como la transmisión vía Internet) de la Obra si tal representación o ejecución tiene fundamentalmente el propósito de generar alguna ventaja comercial, lucro o compensación monetaria.
 - Regalías por la creación y distribución de versiones. El Licenciante se reserva el derecho exclusivo a recaudar, ya sea de manera individual o a través de sociedades de gestión colectiva o representante especialmente designado, regalías por concepto de la creación, grabación y distribución de fonogramas que Usted lleve a cabo de la Obra (adaptaciones, versiones o interpretaciones de la misma Obra) si la distribución de tal versión, adaptación o interpretación tiene fundamentalmente el propósito de generar alguna ventaja comercial, lucro o compensación monetaria.
 - Derechos de Transmisión por vía Internet y Regalías Estatutorias (las previstas en ley). A fin de evitar dudas, en el caso en el que la Obra sea una grabación de audio o fonograma, el Licenciante se reserva el derecho exclusivo a recaudar, ya sea de manera individual o a través de sociedades de gestión colectiva, regalías por concepto de la representación o ejecución pública de tipo digital (tal como la transmisión vía Internet) de la Obra si tal representación o ejecución tiene fundamentalmente el propósito de generar alguna ventaja comercial, lucro o compensación monetaria.

5. Declaraciones, Garantías y Exclusión de Responsabilidad

Salvo que las partes acuerden por escrito algo distinto, el Licenciante ofrece la obra “tal como está” y no hace declaraciones con respecto a la obra ni ofrece ningún tipo de garantía, ya sea expresa, implícita, legal o de otro tipo, excluyéndose de manera enunciativa mas no limitativa, de las garantías de titularidad, de que sea adecuada o apropiada para el uso específico para el que fue diseñada o para el uso específico para el que se notificó iba o podía ser utilizada, de no violación de otros derechos, de la ausencia de vicios ocultos u otro tipo de defectos o errores o de precisión, ya sea que sean evidentes o no. En virtud de que ciertos países no aceptan estas exclusiones de responsabilidad, lo dispuesto en esta cláusula puede no ser aplicable a su caso.

6. Limitación de Responsabilidad

Excepto por lo dispuesto en la ley aplicable, en ningún caso y por ningún motivo será el Licenciante responsable frente a usted por daños especiales, incidentales, consecuenciales, punitivos u otros similares que surjan de esta licencia o del uso de la obra, aún cuando el Licenciante advirtiera que dichos daños podían surgir.

7. Terminación

- a. La presente Licencia y los derechos en esta conferidos se rescindirán automáticamente en cuanto Usted infrinja o incumpla los términos y condiciones de la misma. No obstante, lo antes dispuesto no afectará a aquellas personas físicas o morales que hayan recibido Obras Derivadas u Obras Colectivas de Usted bajo la presente Licencia, siempre y cuando éstas se mantengan en pleno cumplimiento de la presente Licencia. Las secciones 1, 2, 5, 6, 7 y 8 continuarán siendo válidas aún después de que esta Licencia se extinga.
- b. Sujeto a los términos y condiciones antes mencionados, la presente Licencia es perpetua (por todo el término de protección

conferido en la LFDA a la Obra). No obstante lo anterior, el Licenciante se reserva el derecho a distribuir la Obra bajo los términos y condiciones de una licencia distinta o a dejar de distribuirla en cualquier momento, siempre y cuando dicha decisión no cumpla el propósito de revocar la presente Licencia (o cualquier otra licencia que haya sido o requiera ser otorgada bajo los términos y condiciones de la presente Licencia) y los términos y condiciones de la misma continúen surtiendo efectos y se mantengan en vigor, a menos que la Licencia se extinga de la manera antes descrita.

8. Misceláneos

- a. Cada vez que Usted distribuya o represente o ejecute públicamente de manera digital la Obra o una Obra Colectiva, el Licenciante le otorga al usuario una licencia sobre la Obra en los mismos términos y condiciones aquí descritos.
- b. Cada vez que Usted distribuya o represente o ejecute públicamente de manera digital una Obra Derivada, el Licenciante le otorga al usuario una licencia sobre la Obra en los mismos términos y condiciones aquí descritos.
- c. El que alguna de las disposiciones descritas en la presente Licencia sea declarada inválida o incoercible de conformidad con la legislación aplicable no afectará la validez y coercibilidad del resto de las disposiciones contenidas en la presente Licencia. Las disposiciones que resulten inválidas o incoercibles serán modificadas para cumplir los requisitos mínimos indispensables para lograr esa validez y coercibilidad sin que se requiera para ello ninguna acción de las partes de la presente Licencia.
- d. Ninguna de las disposiciones aquí mencionadas se entenderá renunciada ni se presumirá consentimiento otorgado a violación alguna a menos que dicha renuncia o consentimiento se haga por escrito y se firme por la persona con las facultades pertinentes.
- e. La presente Licencia constituye la totalidad del acuerdo entre las partes en relación a la Obra. No existen acuerdos, convenios ni entendimientos con respecto a la Obra fuera de los aquí espe-

Cómo preservar mi patrimonio digital personal

cificados. El Licenciante no se considerará obligado por disposiciones adicionales que aparezcan en cualquier comunicación con Usted. La presente Licencia no podrá ser modificada sin el consentimiento mutuo que conste por escrito entre el Licenciante y Usted.

- f. No será necesario el envío o recepción de la aceptación de las condiciones de la presente licencia para que se considere perfeccionada, por lo que se entiende que cualquier uso de la obra en los términos de la presente licencia constituirá la aceptación de los términos y condiciones contenidos en la misma.

Creative Commons no es parte en esta licencia y no da ninguna garantía ni asume responsabilidad alguna en relación a la Obra. *Creative Commons* no será responsable frente a usted por ninguno de los daños surgidos de la presente licencia, incluyendo, de manera enunciativa mas no limitativa, los daños generales, especiales, incidentales o consecuenciales. No obstante lo anterior, si *Creative Commons* se ha manifestado expresamente como Licenciante, adquirirá los derechos y obligaciones propios del Licenciante dispuestos en la presente Licencia.

Exceptuando el propósito específico de mencionarle al público en general que la Obra se rige por la CCPL, ninguna de las partes podrá usar la marca "*Creative Commons*" o cualquier otra marca o logo relacionados con *Creative Commons* sin autorización previa y por escrito. Cualquier uso permitido deberá cumplir con la política de uso de la marca establecida por *Creative Commons* en el momento determinado, la cual podrá ser publicada en su sitio electrónico o puesta a disposición de otra forma si así se solicita.

Fin del texto legal

El sitio mundial de *Creative Commons* puede ser contactado en:
creativecommons.org/

El sitio *Creative Commons* México en:
creativecommons.mx

Anexo 6

Principales modelos e iniciativas en preservación archivística para documentos de archivo de organizaciones a nivel mundial para aquellos que deseen incursionar en la lectura de la preservación de archivos bajo la óptica institucional.

Estas lecturas no son para documentos personales

Dada la cantidad y variedad de documentación digital generada en los últimos años y dentro del ámbito institucional y en especial gubernamental se han producido una serie de normatividades, recomendaciones, estándares, etc., que buscan organizar y estructurar los esfuerzos que las organizaciones deben realizar a nivel mundial para tratar de garantizar el acceso actual y futuro a los documentos digitales relevantes. No existe hoy en día un esfuerzo único que establezca o defina los atributos y características ideales de todos los documentos digitales y sus procedimientos asociados con miras a su preservación a largo plazo. En los últimos 15 años se han ido estableciendo una serie de iniciativas, se han planteado modelos y marcos de referencia para tratar de definir o establecer los atributos de documentos, procedimientos, sistemas, organizaciones, etc., que se encuentren involucradas en la gestión y la preservación digital. Cada uno de ellos hace énfasis en elementos o atributos que considera importantes para el establecimiento del modelo en cuestión. Como ya se mencionó, estos modelos están hechos por y para organizaciones, y no pueden ser aplicados tal cual para la preservación documental personal por diversas razones, ya discutidas dentro del texto. Pero para conocer un poco más de estas propuestas, he aquí un resumen:

Cómo preservar mi patrimonio digital personal

La base de todas las iniciativas de preservación de documentos de archivo digitales es la norma *ISO 15489:2001 Information and Documentation: Records Management: Part 1. General*. Consiste en una guía para la gestión de documentos de archivo de una organización, sea cual sea su soporte. Esta norma establece que “[...] los documentos de archivo deben ser auténticos, confiables, completos, sin alteración, y deben permitir su uso y acceso. Asimismo, deben poseer metadatos que definan el contexto, contenido y estructura y deben reflejar con precisión la comunicación, acción o decisión”.

Entre las principales iniciativas podemos destacar los siguientes: El modelo de referencia conocido como OAIS –*The Open Archival Information System*– el cual provee un modelo de alto nivel acerca de la identificación de atributos de archivos y cuyo propósito es asegurar la permanencia de cierta información digital generada por una comunidad dada.⁷⁵ Observamos también el reporte acerca de “Repositorios Digitales Confiables” de RLG-OCLC en el cual se establecieron los atributos y recomendaciones para los mismos, entre las que destacan el desarrollo de marcos de trabajo y procesos tendientes a apoyar la certificación de repositorios digitales.⁷⁶ Derivado de estos esfuerzos, surgió la lista de verificación del *Digital Repository Certification Task Force*⁷⁷ la cual se enfoca a establecer los criterios necesarios para construir la confianza en los documentos de archivo. La Administración Nacional de Archivos y Registros de la Unión Americana –*National Archives and Records Administration*– adoptó como estándar para el manejo de archivos gubernamentales el denominado DoD.5015.2 del Departamento de la Defensa de ese país, el cual a su vez se deriva de estándares creados en la Universidad de la Columbia Británica en Canadá.

En el continente europeo, distinguimos las especificaciones establecidas por el proyecto de los Archivos Nacionales –*National Ar-*

75 Consultative Committee for Space Data Systems (CCSDS), *Reference model for an open archival information system (OAIS)*.

76 RLG-OCLC Working Group on Digital Archive Attributes, *Trusted digital repositories: Attributes and Responsibilities* [en línea].

77 RLG-NARA Digital Repository Certification Task Force, *Trustworthy repositories audit & certification: Criteria and checklist* [en línea].

chives– del Reino Unido, uno de los más ambiciosos y avanzados al momento, el cual consiste principalmente en un conjunto estandarizado de requerimientos funcionales para archivos electrónicos. Son notables también las especificaciones denominadas MoReq –Modelo de requisitos para la gestión de documentos de archivo– muy difundidas en ese continente. Este no es un modelo de preservación en sí; consiste en un modelo de requisitos funcionales para la gestión de documentos electrónicos de archivo elaborado por la Comisión Europea a través de su programa IDABC –*Interoperable Delivery of European e-Government Services to Public Administrations, Business and Citizens*– con objeto de ir estandarizando la gestión de archivos digitales en todos los países miembros de la Unión Europea así como por todos los interesados en el desarrollo y aplicación de sistemas de gestión de documentos electrónicos de archivo –archivistas, informáticos, proveedores de servicios, instituciones académicas, entre otros–. Sin ser un modelo de preservación, es interesante ya que abarca aspectos que son útiles para la misma.

En ese continente existen también las especificaciones del *Digital Curation Centre, Digital Preservation Europe*, etc. Son notables también las especificaciones establecidas por DRAMBORA⁷⁸ –*Digital Repository Audit Method Based on Risk Assessment*–. Esta es una herramienta de auto-diagnóstico para una aproximación sistemática para evaluación de riesgos inherentes a un repositorio digital.

En Australia se creó un modelo conocido como el *Records Continuum* –Continuo de los documentos de archivo– el cual ha sido trabajado por el *Records Continuum Research Group*⁷⁹ de ese país afiliado con el *Center for Information as Evidence at UCLA*, California. Este modelo se opone al más utilizado basado en el concepto del “ciclo de vida” de los documentos de archivo. El modelo del *records continuum* está construido sobre la afirmación de que la gestión del documento es un proceso continuo desde el momento

78 Digital Curation Centre/DigitalPreservationEurope, DRAMBORA. *Digital Repository Audit Method Based on Risk Assessment*. Sitio web oficial: <http://www.repositoryaudit.eu/>

79 Monash University, *Records Continuum Research Group* [en línea].

de su creación; los conceptos relativos a dicha gestión pertenecen a cuatro dimensiones o puntos de vista del observador.⁸⁰ Este modelo también es conocido como Modelo de gestión de documentos de archivo orientado a organizaciones –*Business-Driven Recordkeeping Model*– o simplemente *BDR Model*. En una temática relacionada con el tema, los Archivos Nacionales de ese país publicaron las “Especificaciones funcionales para *software* de sistemas de administración de documentos de archivo electrónicos” y las “Guías asociadas para su implementación”.⁸¹

El ciclo de vida de un documento es un modelo o concepto clásico de la archivística moderna. Se refiere a las distintas fases o etapas por las que pasan los documentos y a las transformaciones en sus valores legales, informativos, etc., desde su creación como co-productos de la actividad humana hasta a su eliminación o selección para su custodia permanente, generalmente por su valor histórico. Estas etapas son ocho: producción, clasificación, mantenimiento para el uso, disposición, descripción de las ayudas o auxiliares de consulta, preservación, referencia y consulta.

En la Universidad de la Columbia Británica de Canadá se creó desde 1999 un proyecto denominado InterPARES –*The International Research on Permanent Authentic Records in Electronic Systems*–. Este es un proyecto de investigación teórica y desarrollo de tecnología alrededor de la preservación de documentos de archivo electrónicos. Este proyecto perfeccionó un modelo de aproximación a la preservación digital denominado Cadena de preservación –*Chain of Preservation*– o simplemente COP.⁸² Este modelo estable-

80 InterPARES Project (2008), *InterPARES 2. Experiential, Interactive and Dynamic Records. Appendix 16. Overview of the Records Continuum Concept. Extracted and adapted from “Xiami An. An Integrated approach to Records Management”*. *Information Management Journal*. July/August (2003): 24-30 [en línea]

81 National Archives of Australia (2007), *Functional Specifications for Electronic Records Management Systems Software (ERMS) y Guidelines for Implementing the Functional Specifications for ERMS* [en línea]

82 InterPARES Project (2004), *Business-Driven Recordkeeping (BDR) Model* [en línea].

ce que los documentos de archivo digitales deben ser cuidadosamente manejados a lo largo de toda su existencia para asegurar que sean accesibles y legibles a lo largo del tiempo dejando su forma, contenido y relaciones intactas hasta el punto necesario para lograr su continua confianza como documentos de archivo. El manejo de los documentos de archivo digitales debe emanar de un entendimiento integral de todas las fases o etapas de la existencia de los documentos de archivo, desde el momento en que son generados, a través de su conservación por parte de su productor, durante la valoración, disposición y preservación a largo plazo como comprobantes auténticos de las acciones y asuntos de los cuales son parte. Desde la perspectiva de la preservación a largo plazo, todas las actividades tendientes a administrar documentos de archivo a lo largo de su existencia están interrelacionadas, como en una cadena, y son interdependientes entre sí. Si un eslabón de esa cadena llega a fallar, la cadena no puede cumplir su cometido. Si ciertas actividades y acciones no son realizadas sobre esos documentos de archivo, su confianza –esto es, su fiabilidad, autenticidad o exactitud– se vuelve cuestionable.

El modelo de la Cadena de preservación es ampliamente aceptado en Norteamérica –México inclusive–, en la mayor parte de Europa y mayormente en muchas otras partes del mundo: Sudamérica, África, lejano oriente. Es más utilizado y aceptado que el modelo del continuo de los documentos de archivo. En particular, la terminología y conceptos utilizados en esta obra se apegan al modelo de la cadena de preservación.⁸³

EL MODELO DE LA CADENA DE PRESERVACIÓN

¿En qué consiste el modelo de la cadena de preservación? En términos generales, es una secuencia de actividades y enfoques agrupados en varias etapas que conforman los “eslabones” de la cadena.

83 InterPARES Project (2008), *Preserver Guidelines. Preserving Digital Records: Guidelines for Organizations* [en línea].

Cómo preservar mi patrimonio digital personal

De todas y cada una de estas actividades y elementos se han ido haciendo investigaciones y especificaciones detalladas y su conocimiento profundo implica un estudio a fondo de ellas. Haciendo un resumen de las mismas, la cadena de preservación consta de cinco etapas:

1. Establecer el marco de referencia del fondo archivístico. No todos los fondos archivísticos a preservar tienen el mismo propósito. Es necesario definir de entrada el objeto de preservación.
 - 1.1 Establecer alcance y objetivos; definir el propósito y objetos del fondo documental a preservar.
 - 1.2 Allegarse los recursos económicos, tecnológicos y humanos. Tiempos de adquisición, de capacitación, etcétera.
 - 1.3 Enfocarse en los documentos de archivo digitales, y sólo en segundo plano en aquellos de los que se guarda copia en soporte tradicional.
 - 1.4 Ofrecer asesoría acerca de procedimientos, especificaciones y características de los buenos acervos digitales.
 - 1.5 Desarrollar procedimientos de control sobre la transferencia, conservación y reproducción de los documentos de archivo, en especial en procedimientos y sistemas usados para la gestión documental al interior de la organización.
 - 1.6 Implementar estrategias de conservación documental. En general se invierte mucho esfuerzo en el desarrollo de las estrategias para la preservación, pero poco se hace con respecto a las estrategias para la conservación, salvaguarda y seguridad informática de los documentos. Preservación, conservación y seguridad informática no son sinónimos.⁸⁴ Preservación documental digital se define como “la totalidad de principios, políticas y estrategias que controlan las actividades destinadas a asegurar la estabilización física y tecnológica, así como la protección del contenido intelectual de materiales –datos, documentos o archivos– y cuyo fin ulterior y a largo plazo, es el de asegurar la permanencia y acceso del contenido de documentos digitales a lo largo del

84 Juan Voutssas M., “Preservación documental digital y seguridad informática”.

tiempo y las tecnologías, independientemente de su soporte, formato o sistema”. Seguridad informática es “el proceso de establecer y observar un conjunto de estrategias, políticas, técnicas, reglas, guías, prácticas y procedimientos tendientes a prevenir, proteger y resguardar de daño, alteración o sustracción a los recursos informáticos de una organización y que administren el riesgo al garantizar en la mayor medida posible el correcto funcionamiento ininterrumpido de esos recursos”. Conservación documental digital: “acciones tomadas para anticipar, prevenir, detener o retardar el deterioro del soporte de obras digitales con objeto de tenerlas permanentemente en condiciones de usabilidad, así como la estabilización tecnológica, la reconversión a nuevos soportes, sistemas y formatos digitales para garantizar la trascendencia de los contenidos”.

2. Valorar los documentos de archivo para preservación permanente.
 - 2.1 Valorar lo más pronto posible cuáles documentos deben ser sujetos de preservación ya que esto incide en retroalimentación de procedimientos y características al creador de los registros, a la transferencia y a las estrategias de conservación y preservación, lo cual mejora las probabilidades de éxito del proyecto.
 - 2.2 Localizar múltiples propietarios, cuando los contenidos o usuarios de un fondo documental pertenecen a diversas organizaciones, inclusive de distintos subsistemas.
 - 2.3 Valorar la autenticidad y documentarla. Como parte del proceso inicial de valoración, es necesario establecer la autenticidad de los documentos de archivo ya existentes. Luciana Duranti establece: “El original de un documento –esto es, la primera instanciación completa de una entidad documental que alcanzó sus propósitos– desaparece en el ambiente digital la primera vez que es salvado. Lo que nosotros recuperamos siempre es una copia. No podemos preservar documentos –entidades– digitales: sólo podemos preservar la capacidad de reproducirlos una y otra vez. En este contexto,

una entidad digital preservada es considerada auténtica si puede considerarse que es –o aún mejor– declararse que es una copia auténtica por el custodio quien de fe de su identidad y de su integridad a lo largo del tiempo partiendo desde el momento en que lo ingresó a su acervo y ese custodio puede documentar además correctamente el proceso de conservación –inclusive cualquier migración posterior y sus consecuencias tanto en forma como en contenido–. Ello significa que, en lo relativo a documentos en medios tradicionales, la autenticidad fue establecida siempre a través del objeto mismo, del documento, así que el custodio sólo necesitó preocuparse de que el usuario analizase el objeto y sacara sus propias conclusiones acerca de su autenticidad. Con medios digitales, lo que el usuario necesita para analizar y concluir es conocer la autoridad y la capacidad –competencia– del custodio, así como la calidad de la documentación del proceso de conservación”.⁸⁵

Es necesario documentar los elementos que permitieron declarar la autenticidad, documentación que por lo general contiene los siguientes elementos:

- a) Expresión de los atributos de los documentos de archivo y vínculos a otros documentos –metadatos de integridad e identidad–.
- b) Privilegios de acceso.
- c) Procedimientos de protección contra pérdida o corrupción de documentos.
- d) Procedimientos de protección contra deterioro de medios y cambios tecnológicos.
- e) Establecimiento de las formas documentales; es decir, las reglas de representación según las cuales el contenido de un documento de archivo, su contexto administrativo, documental y su autoridad son comunicados.
- f) Autenticación de los documentos de archivo.

85 Luciana Duranti y Kenneth Thibodeau (2005), “The Concept of Record in Interactive, Experiential and Dynamic Environments: The view of InterPARES”.

- g) Identificación del documento de archivo autorizado u oficial.
 - h) Remoción y transferencia de documentación relevante.
- 2.4 Monitorear documentos de archivo seleccionados para preservación a largo plazo.
 - 2.5 Actualizar la valoración y sus criterios de tiempo en tiempo.
 - 2.6 Identificar todos los componentes digitales, las “capas”, componentes, o elementos que deben ser integrados para reproducir el documento original.
 - 2.7 Determinar la viabilidad de la preservación, establecer los límites y características de la usabilidad y aceptabilidad de los documentos de archivo para poder mantener a lo largo del tiempo las características de calidad intrínseca. Deben contemplarse también los procedimientos y costos recurrentes derivados de migración, transformación, emulación, reformato, etcétera.
3. Ganar, obtener o conseguir posesión o control sobre los fondos archivísticos seleccionados o agregar a los fondos ya poseídos con fines de preservación permanente.
 - 3.1 Desarrollar plan para transferencias documentales desde su productor o sucesor hacia el preservador de largo plazo con un plan previamente establecido.
 - 3.2 Aplicar procedimientos estandarizados para la transferencia de documentos de archivo de un productor hacia el preservador.
 - 3.3 Mantener el formato más antiguo que sea utilizable en el cual los documentos fueron producidos o mantenidos y sólo cambiarse cuando su uso ponga en riesgo su accesibilidad.
 - 3.4 Evitar duplicados ociosos de los documentos.
 - 3.5 Documentar todos los procesos que son realizados de manera rutinaria o eventual sobre los documentos: procesos de antivirus, integridad de la base de datos, compactación, transferencia, cambio de versión, etcétera.
 4. Preservar los documentos de archivo cuya autenticidad y viabilidad de preservación han sido confirmados y han sido recibidos e ingresados a un archivo histórico por el preservador.
 - 4.1 Describir/catalogar/registrar los documentos de archivo.

Cómo preservar mi patrimonio digital personal

- 4.2 Identificar aspectos legales de la preservación, transparencia, privacidad, seguridad, etcétera.
 - 4.3 Confirmar la eficacia de la estrategia de preservación elegida y en su caso, replantearla.
 - 4.4 Almacenar adecuadamente en cuanto a seguridad física, temperatura, humedad, respaldos, múltiples copias, renovaciones, etcétera.
5. Distribuir los documentos de archivo.
 - 5.1 Documentar cómo se hicieron las copias preservadas.
 - 5.2 Documentar los requerimientos para acceso a los documentos, tanto tecnológicos, como de autorización, tiempos, alcances, etcétera.

Como ha podido apreciarse, el modelo de la Cadena de preservación implica el desarrollo y observancia meticulosa de procedimientos que comienzan con la gestión documental, involucran a todo el ciclo de vida del documento y desembocan finalmente en una preservación adecuada. Y como ha podido verse en este resumen de los modelos e iniciativas alrededor del manejo y preservación de documentos de archivo digitales existen ya toda una serie de conocimientos y habilidades alrededor de la preservación que las organizaciones con acervos archivísticos deben ir cultivando y desarrollando para poder preservar sus fondos digitales adecuadamente y a largo plazo.

Cómo preservar mi patrimonio digital personal. La edición consta de 50 ejemplares. Coordinación editorial: Carlos Ceballos Sosa. Revisión especializada: Aurea Gabriela Mondragón Pérez. Revisión de pruebas: Francisco Xavier González y Ortiz. Formación editorial: José Luis Maldonado López. Instituto de Investigaciones Bibliotecológicas y de la Información/UNAM. Fue impreso en papel cultural de 90 g en los talleres de Cactus Display, S. A. de C. V., ubicados en Corregidora número 398, Col. Miguel Hidalgo, Tlalpan, México, D.F., C. P. 14260. Se terminó de imprimir en el mes de febrero de 2013.