

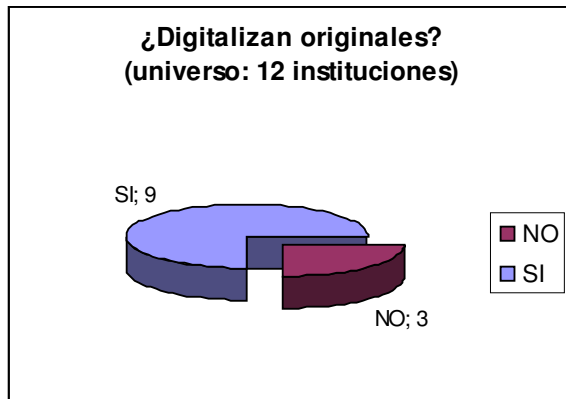


“Preservación del patrimonio digital en bibliotecas argentinas: estudio exploratorio y experiencia piloto”

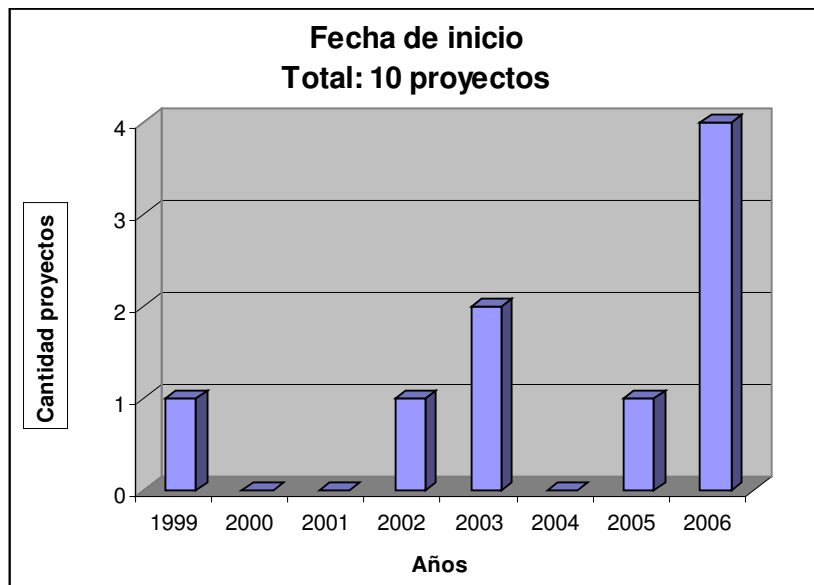
## Resultados encuesta autoevaluación institucional

### Primera Parte: Objetos Digitalizados

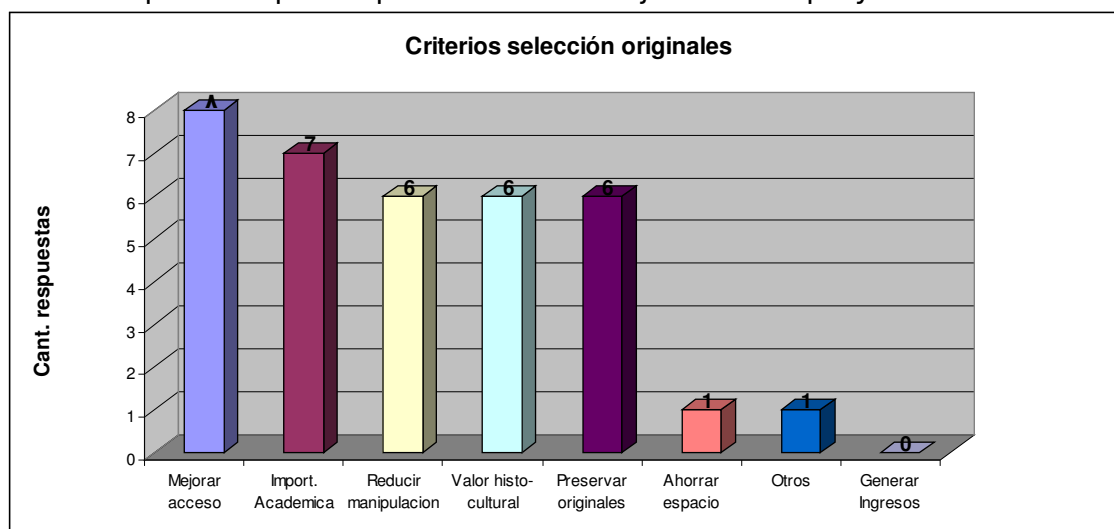
Sobre un total de 14 encuestas institucionales recibidas, se pudieron utilizar 12, por presentar datos de actividades ya iniciadas. De ellas, como se aprecia en el cuadro siguiente, 9 tienen proyectos de digitalización iniciados, y una más mantiene una colección digitalizada, aunque esa institución no realizó el procesamiento digital, y se limita a darle el hosting para Internet.



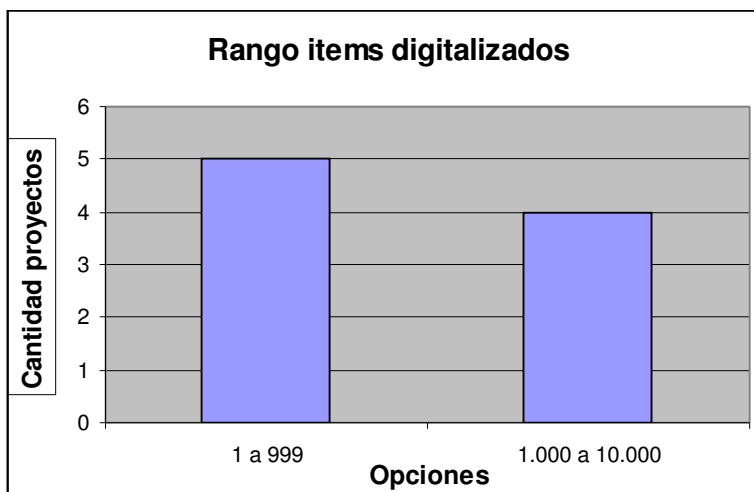
De los 9 proyectos/instituciones que iniciaron actividades de digitalización, los 4 más antiguos empezaron entre los años 1999 y 2003, mientras que los 4 más recientes se iniciaron todos durante el año 2006. Si estos datos pudieran generalizarse -hay que tener mucha prudencia con universos tan reducidos numéricamente-, hablaríamos de una tendencia hacia el crecimiento de este tipo de proyectos en las instituciones culturales de Argentina. De todos modos no es una conclusión muy apresurada, si tenemos en cuenta que existe una tendencia similar en muchos países del mundo.



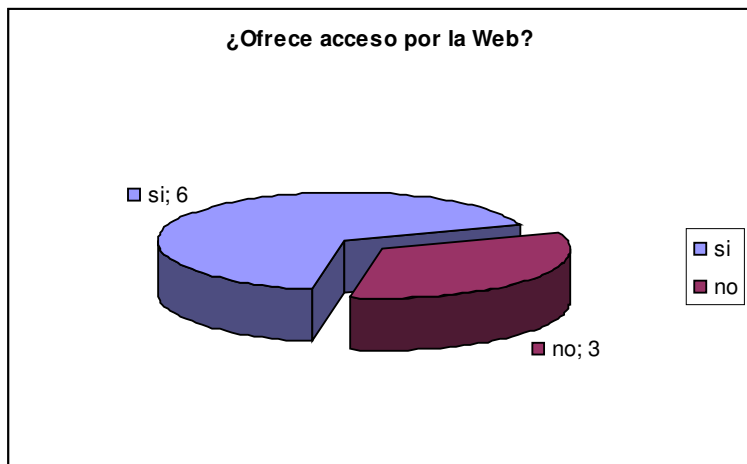
Se preguntó a las instituciones que indicaran los criterios de selección aplicados para decidir cuáles originales digitalizar. En el siguiente cuadro se puede ver el resultado. Es interesante destacar que la respuesta más representada, “Mejorar el acceso”, también fue el criterio predominante que guió los esfuerzos de digitalización en otras instituciones del mundo, en las épocas en que comenzaron a procesar originales, por lo general antiguos y únicos, libros raros, etc. Es sabido que una de las principales razones que motivaron el inicio de los proyectos de digitalización en el mundo estuvo asociada con el interés de mejorar el acceso a originales, una vez que la web fue convirtiéndose en un medio eficaz de comunicación a distancia. Sin embargo, los criterios de “Importancia académica”, “Reducir la manipulación de originales”, “Valor histórico-cultural” y “Preservar los originales” han tenido una importante cantidad de respuestas positivas, lo que muestra que esas preocupaciones no son ajenas a los proyectos en curso.



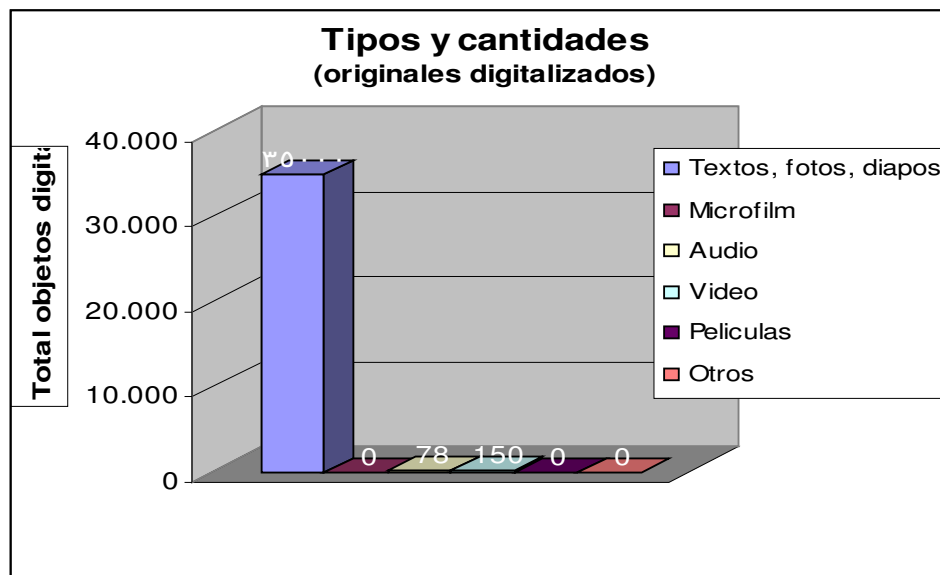
Como una muestra más de la “juventud” de este campo en nuestro país, se advierte que ninguna institución ha superado la cantidad de 10.000 originales digitalizados, y al mismo tiempo, la mayoría de los proyectos están por debajo de los 1.000 ítems digitalizados. Sería injusto, y de poca utilidad, comparar estas cifras con las de otros proyectos pioneros en esta área, ya sea de los EEUU, Canadá, Europa, y en general, de los países más avanzados, en los que el número de ítems digitalizados alcanza cifras superiores al millón de unidades. En verdad, puede considerarse a esta etapa inicial que vivimos en Argentina como un buen punto de partida, donde todavía estamos a tiempo de consensuar buenas prácticas, aprender de otras experiencias más avanzadas, y tender hacia el desarrollo de colecciones digitales sustentables en el largo plazo.



Considerando la importancia atribuida al criterio de “mejorar el acceso”, es interesante observar en qué medida esta motivación se tradujo en la disposición a brindar acceso por la web a las colecciones creadas. De las 9 instituciones que digitalizaron material, 6 lo ofrecen por la web, y 3 sólo por otros medios.

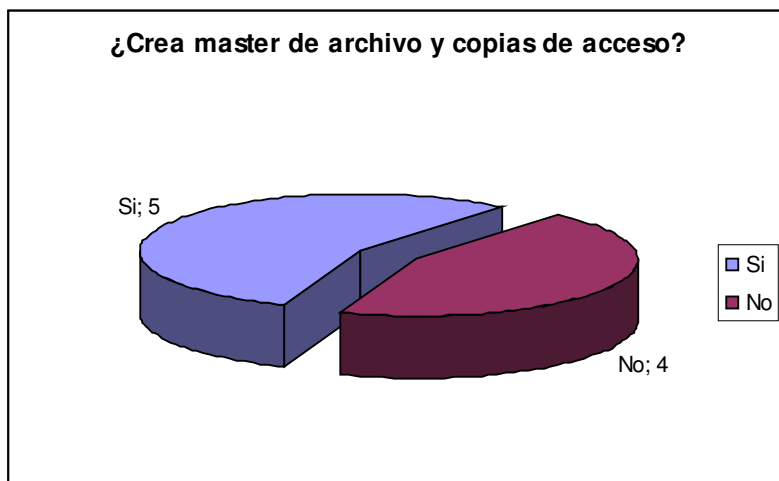


Entre los tipos de originales digitalizados predominan ampliamente los que se resuelven mediante imágenes fijas, esto es, textos impresos y manuscritos, fotografías, negativos y diapositivas. Sólo una institución comenzó a digitalizar información de flujo de datos, tales como archivos de audio y video. Esto no puede sorprender, debido a la complejidad adicional que supone el procesamiento de ese tipo de información, y los desafíos que implica sobre los sistemas de acceso web así como de almacenamiento.

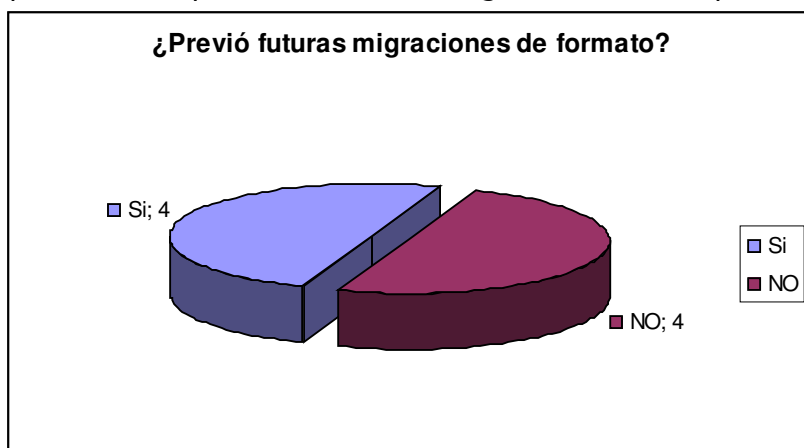


Se preguntó si en el proceso de digitalización se utilizaba una estrategia de creación de archivos masters, de mayor calidad y resolución, con producción derivada de copias de acceso para la web, o bien para los sistemas de consulta en sede. En términos prácticos, la mitad así lo hace y la otra mitad no. Este punto es de gran valor para evaluar la probabilidad de una larga vida de las colecciones digitalizadas, ya que la realización de un archivo master de alta calidad facilita el mantenimiento de la utilidad del objeto digital a lo largo del tiempo, entre otras

razones, porque permite enfrentar los rápidos cambios en las tecnologías de acceso y representación electrónica con nuevas copias derivadas.

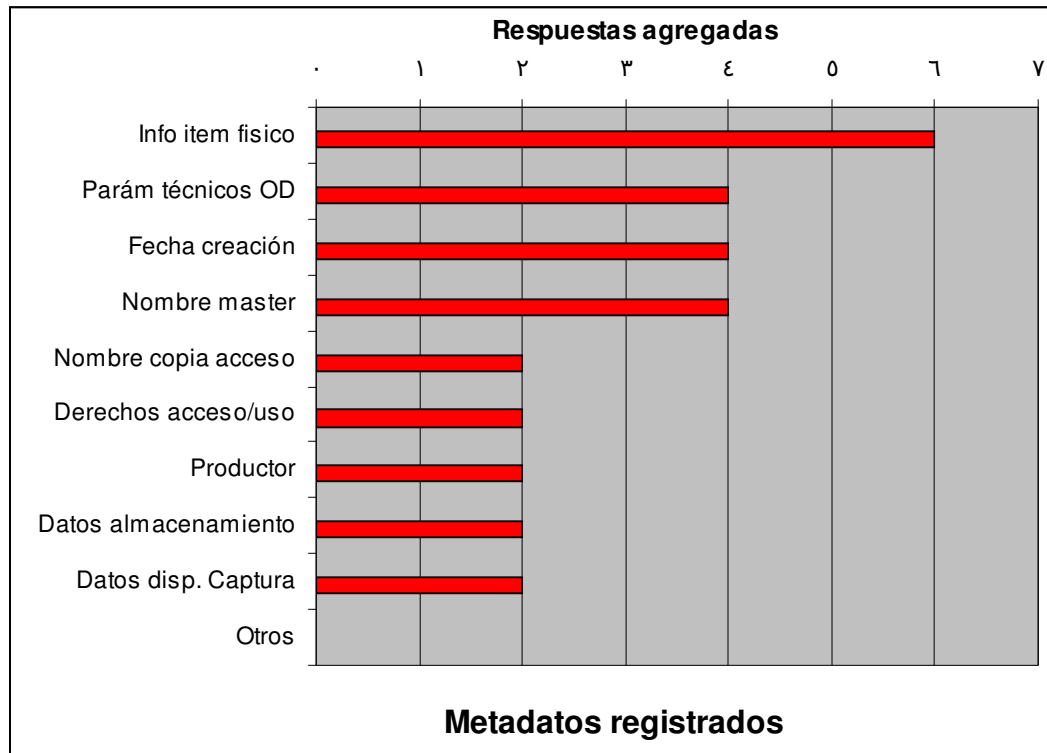


Otro punto consultado fue acerca de si se tuvo en cuenta, al digitalizar los archivos, la eventual necesidad de cambiar de formato digital en el futuro, para evitar los riesgos de la obsolescencia. En general, la elección de formatos abiertos y universales, bien documentados, dejan la puerta abierta para futuras migraciones, cuando estas sean necesarias. En forma similar a los resultados de la pregunta anterior, las respuestas positivas alcanzan a la mitad de los proyectos. También esta pregunta ayuda a orientar decisiones futuras para las instituciones que aún no aplican estas estrategias básicas de prevención.

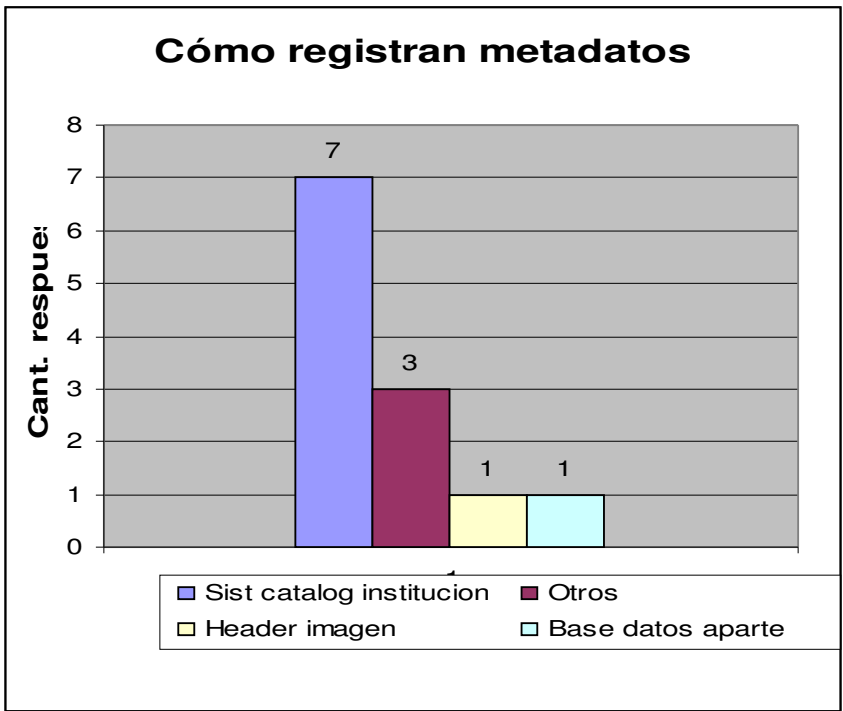


El siguiente cuadro resume las respuestas agregadas sobre el tipo de metadatos que se registran al momento de digitalizar originales. Las instituciones podían marcar más de una opción, hasta alcanzar la combinación que mejor representara su práctica real. No puede sorprender, dada la preferencia mayoritaria por “mejorar el acceso” ya comentada, que el tipo de metadatos más empleados se refiera a las características del ítem físico que se digitaliza (6 respuestas positivas). Por otro lado, dado que el universo relevado aquí es de las bibliotecas, y éstas ya cuentan con descriptores para los originales en sus sistemas de catalogación, este tipo de metadato probablemente sea el más sencillo para generar, pudiéndose recuperar de -o vincular a- los sistemas de descripción en uso. Sólo la mitad de las instituciones encuestadas registra metadatos relacionados con los parámetros técnicos de los objetos digitales

(básicamente, imágenes), y menos aún los referidos a derechos de acceso y uso, datos técnicos del dispositivo de captura o bien datos sobre el almacenamiento. En este punto se debería recordar la reciente emergencia de software libre desarrollado para capturar en forma automática los datos técnicos de los archivos digitales, que puede facilitar en gran medida la producción de esa información tan necesaria para la gestión de largo lazo de las colecciones digitalizadas.

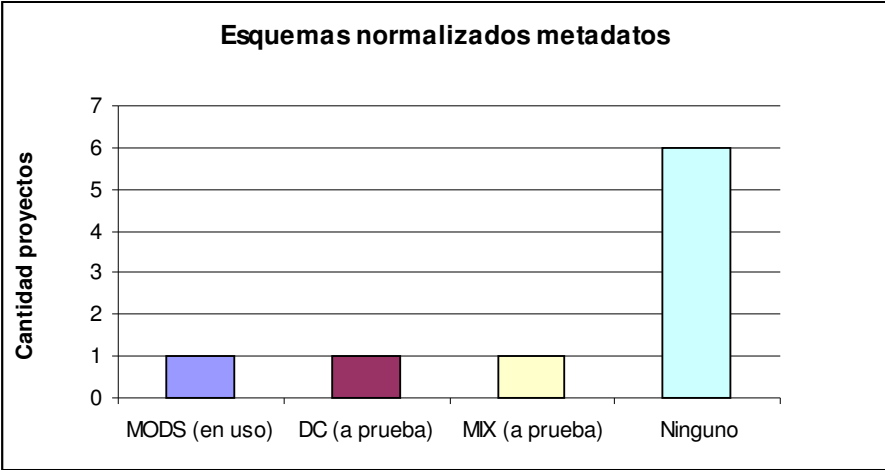


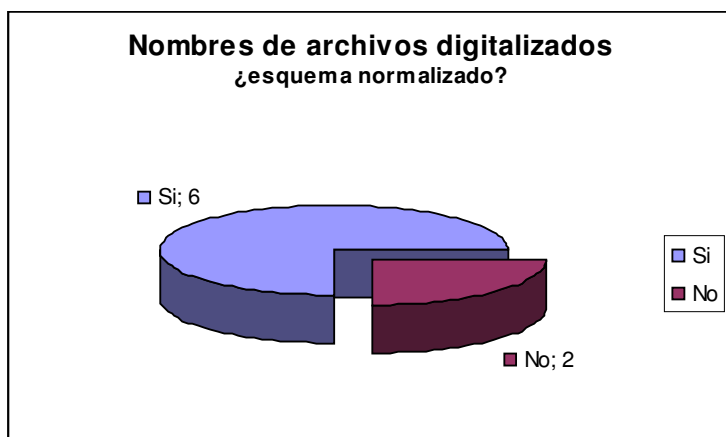
Vinculado con la anterior pregunta, se preguntó de qué modo registraban los metadatos antes descriptos. También en esta pregunta las instituciones podían marcar más de una opción, por lo que la suma de respuestas excede a la cantidad de instituciones relevadas. Por amplia mayoría, las instituciones registran esta información en los sistemas bibliográficos que ya tienen en uso (por ej.: ISIS), lo cual es muy razonable, y una práctica extendida en el mundo entero. Cuando los requerimientos de la preservación digital se hayan difundido ampliamente, es de esperar que tales sistemas ofrezcan modos normalizados de registrar metadatos específicos para tal fin. Sólo en un caso contestaron guardar metadatos en el encabezador de la imagen, y en otro en una base de datos aparte, específica para esa función.



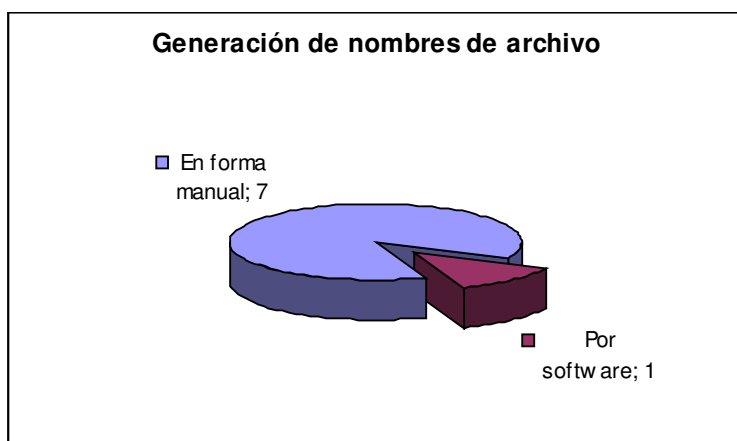
A su vez, como muestran los gráficos contiguos, sólo 3 instituciones declaran emplear un esquema de metadatos normalizado (referido a algún tipo de estándar internacional). Entre ellas, dos lo están haciendo a modo de prueba, y sólo una como práctica normal establecida. La baja frecuencia de respuestas positivas nos muestra uno de los campos donde es

esperable que haya avances en el futuro, a medida que se difundan las buenas prácticas de preservación de objetos digitales.



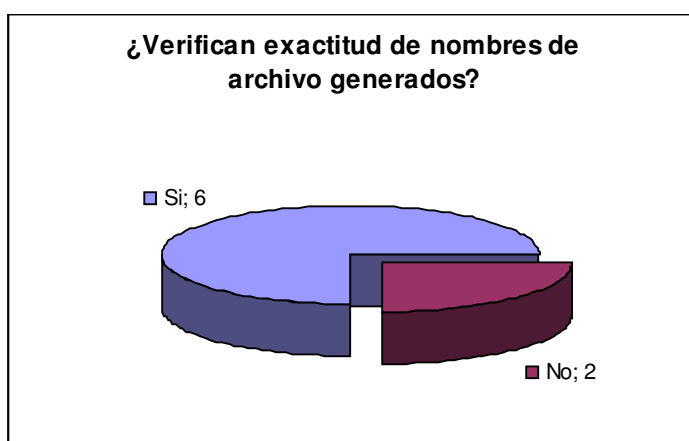


Respecto de la identificación de los archivos electrónicos, 6 instituciones declaran utilizar un esquema normalizado (ad hoc, no hay normas internacionales para esto), y sólo 2 no.



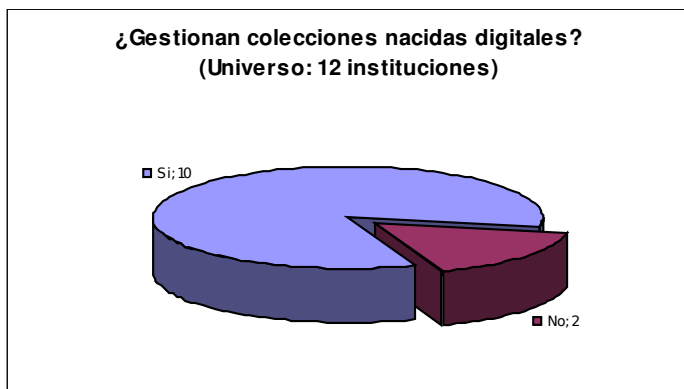
En conexión con lo anterior, y como era previsible, la creación de los nombres de los archivos digitales se realiza manualmente (siete respuestas positivas). Sólo en un caso se responde “por software”. Esto es muy esperable, ya que la aplicación de un esquema de nombres no suele ser resuelto por los softwares de operación de los escáneres, que solo ofrecen una base

muy genérica para este fin.

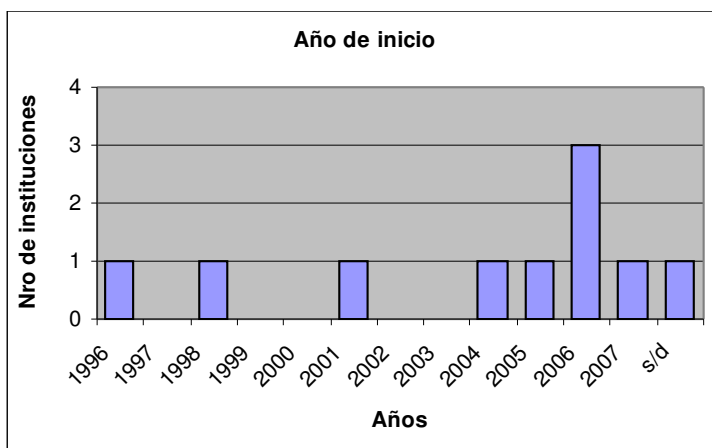


La realización de verificaciones respecto de la corrección de los nombres atribuidos a los archivos digitales es necesaria, ya que por ser una tarea manual suele ser propensa a errores. Las 6 respuestas positivas reflejan que la mayor parte de las instituciones son concientes de la necesidad de realizar esta verificación.

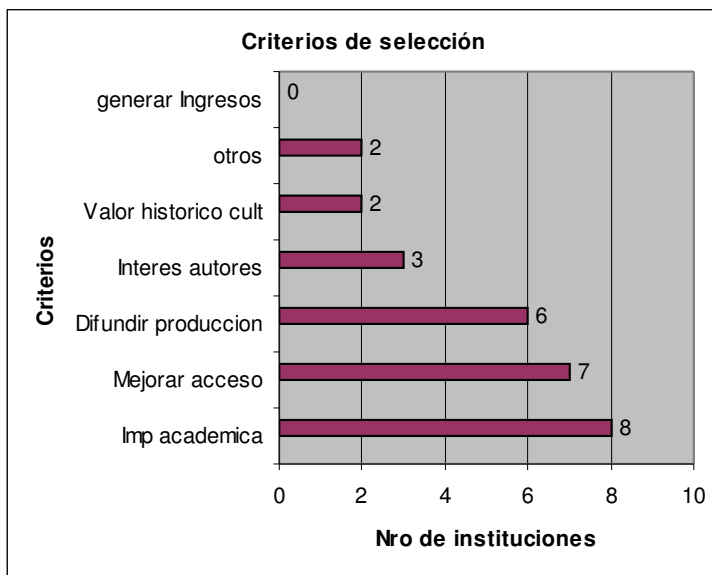
## Segunda Parte - Objetos nacidos digitales



Del universo de 12 instituciones, 10 mantienen colecciones nacidas digitales, y 2 no lo hacen. En general, se trata de materiales académicos o de investigación que reflejan la producción central de las instituciones encuestadas.



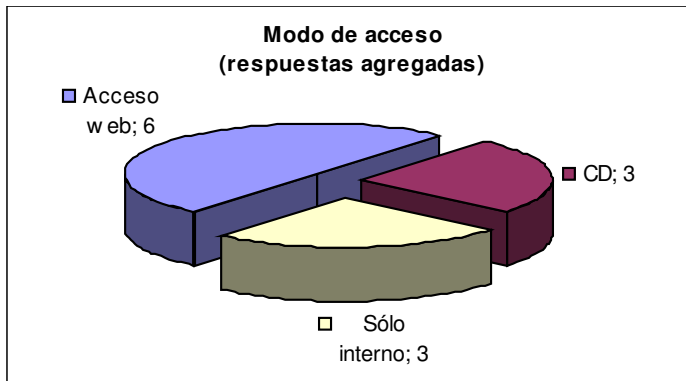
Entre ellas, 5 se iniciaron entre 1996 y el año 2005, 4 entre el 2006 y el 2007, y una no presentó esta información. Si algo muestran estas cifras es que parecería haber un incremento a partir del año 2006 (único año en que la cantidad de proyectos iniciados supera la cantidad de uno), pero es difícil con estos pocos números arriesgarse a adivinar "tendencias".



Respecto de los criterios de selección para estas colecciones, la mayoría se inclina por la importancia académica atribuida al material, seguido de cerca por el deseo de mejorar el acceso a los mismos, y difundir la producción institucional. De menor importancia resultaron criterios tales como el interés de los propios autores por difundir sus escritos y el valor histórico o cultural de los mismos. Tal como ocurría con las colecciones digitalizadas, ninguna institución parece considerar a estas iniciativas

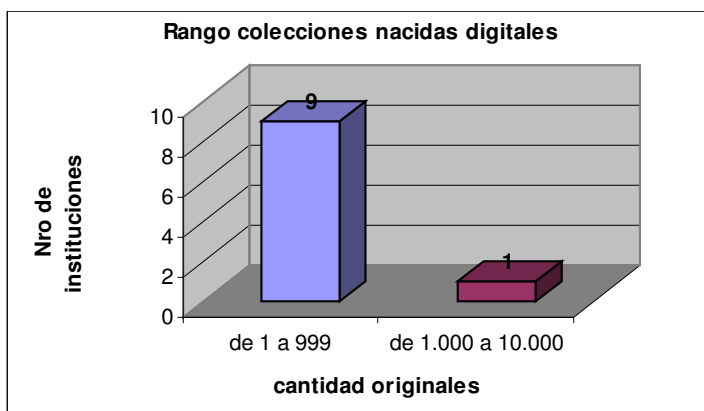
como posibles generadoras de ingresos económicos.





Aunque una de las principales motivaciones para desarrollar colecciones nacidas digitales consiste en el deseo de mejorar el acceso, sólo 6 instituciones de las 10 ofrecen acceso por Internet (y en algún caso sólo a los catálogos, no a los textos mismos). En este caso se permitía marcar más de una opción, de modo que vemos que el acceso a Internet es

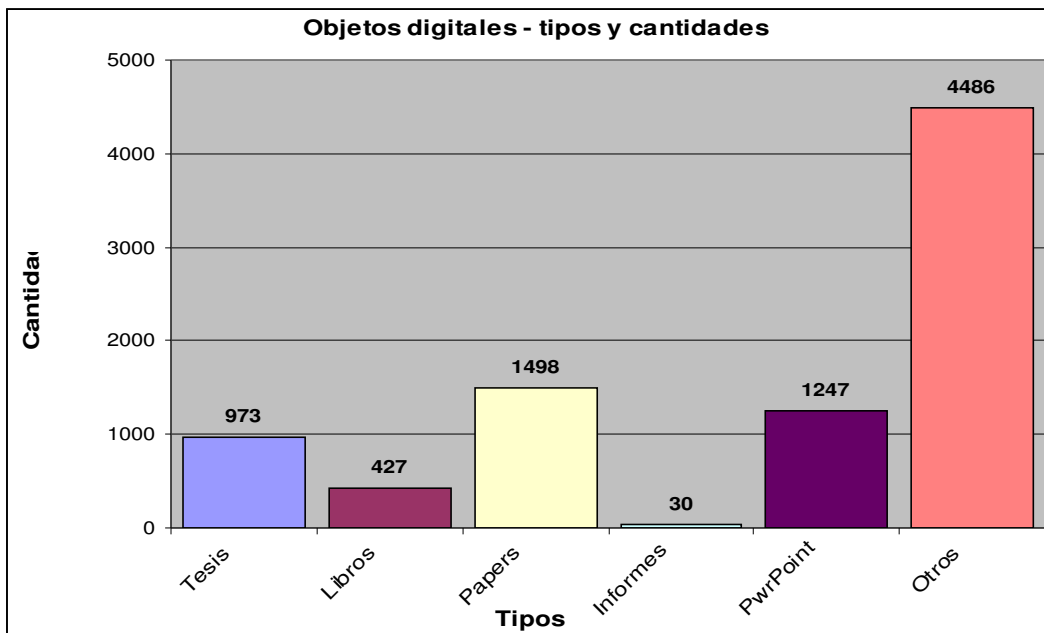
complementado a veces con la distribución en CD de la información. Sin embargo, 3 instituciones contestaron que por el momento sólo tienen un acceso interno a la colección, limitado a puestos de consulta dentro de la institución.



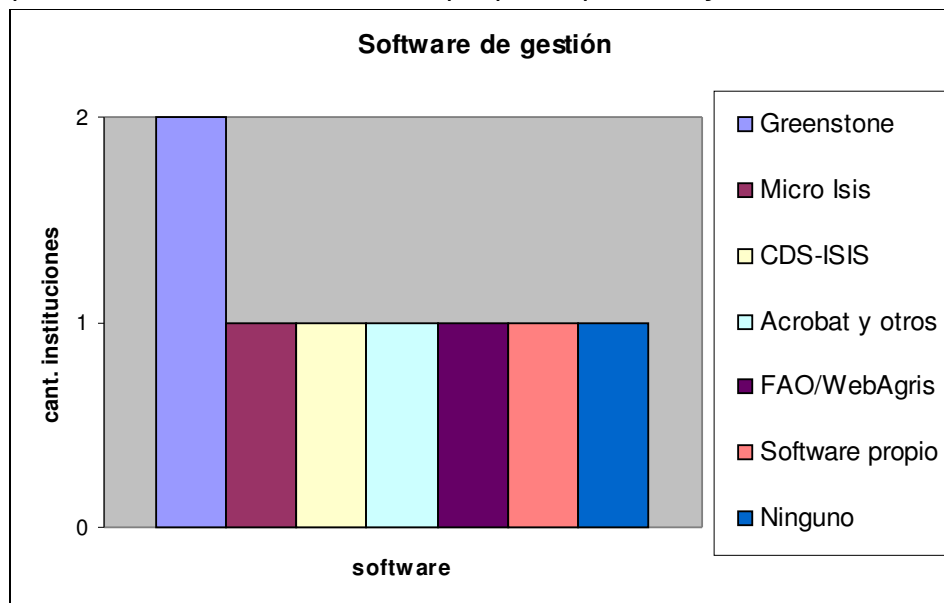
Nuevamente vemos aquí, considerando el rango de "originales digitales" que componen las colecciones, que todavía estamos en un momento inicial de estos esfuerzos en nuestro medio. Así, la gran mayoría de las instituciones (9) mantienen colecciones nacidas digitales compuestas por menos de 1.000 originales, y sólo una institución se ubica en el rango

siguiente, de 1.000 a 10.000 originales digitales.

Desagregado por tipo de original, los resultados totales son los siguientes:

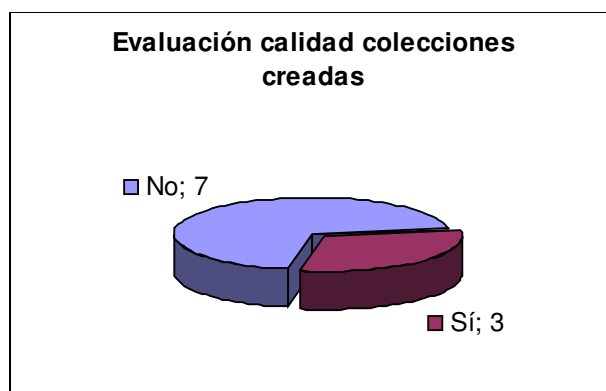


*Herramientas de software utilizadas para la gestión de los materiales nacidos digitales:* sobre 8 respuestas válidas (2 instituciones no contestan esta pregunta), las dos aplicaciones más representadas son Greenstone e ISIS, con dos respuestas positivas cada una respectivamente. Las otras respuestas muestran que cada institución hace sus propias opciones y combinaciones de diversas

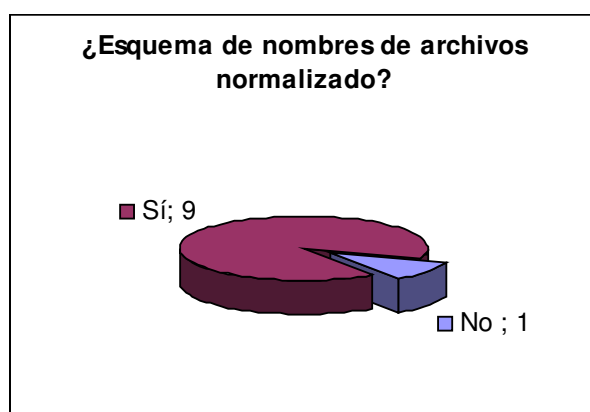


aplicaciones de software, lo que incluye un caso con desarrollo de software propio para la gestión de la biblioteca digital. Como ejemplo de esas

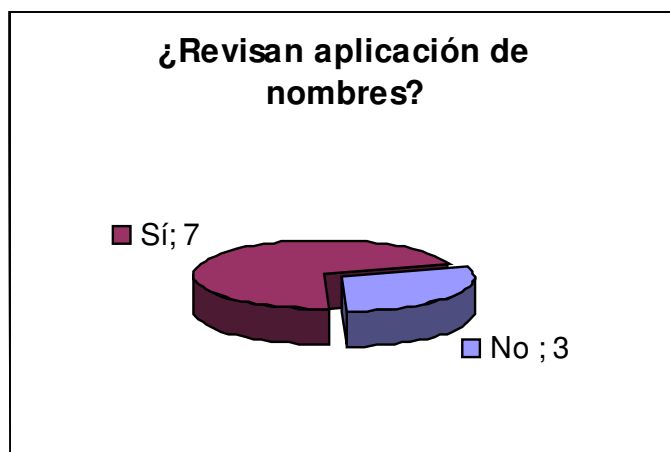
combinaciones, una de las instituciones que utiliza Greenstone, también recurre al uso del software libre Open Office.



Respecto de la evaluación de calidad de las colecciones nacidas digitales, la mayoría contestó negativamente, como se ve en el gráfico que acompaña. Algunas de las respuestas positivas se vieron facilitadas por tratarse de archivos PDF, donde se controla que estén todas las páginas del documento, por ejemplo.



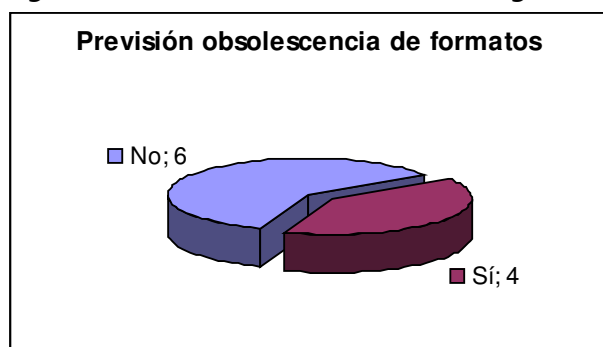
En lo tocante a los nombres de los múltiples archivos que componen las colecciones nacidas digitales, la casi totalidad de las instituciones encuestadas aplica una convención normalizada (ad hoc), y sólo en un caso se registra una respuesta negativa. En la totalidad de los casos (10 respuestas) la aplicación de esos nombres se realiza en forma manual.



Ya que la aplicación del esquema de nombramiento es totalmente manual, y por lo tanto propenso a errores, se preguntó si realizaban un control de calidad de los resultados de esa tarea. Aunque la mayoría de las respuestas contestan positivamente (7 casos), existen tres casos donde tal control no se realiza. Consistentemente, las instituciones que declaran realizar un control de calidad de las colecciones nacidas digitales

bajo gestión, repiten la respuesta positiva en esta pregunta.

*Previsión de obsolescencia de los formatos de archivo al momento de crear o ingresar las colecciones nacidas digitales:*



Una ligera mayoría (6 casos) contestó haber tenido en cuenta este riesgo, mientras que una minoría (4 casos) no lo tuvo en cuenta. Sin duda, este es un punto sobre el que debería trabajarse para mejorar las posibilidades de mantenimiento en valor de las colecciones. Aunque no se lo preguntó, sería muy revelador averiguar a partir de qué momento comenzó a cobrarse conciencia de la

existencia de este riesgo. Preguntados acerca de si tomaban otros recaudos adicionales para prevenir el problema de la obsolescencia de los formatos, las mismas 6 instituciones contestaron afirmativamente, respondiendo:

Caso	Recaudo adicional
1	Guardado en XML y en HTML
2	Se realiza un backup
3	Se establece una fecha de "refreshing"
4	Elección de formatos estándares y universales
5	Se guardan en PDF y HTML simultáneamente
6	Greenstone convierte a formato XML

Comentarios:

Caso 1: XML es un formato abierto y universal apto para preservación de textos electrónicos y otros tipos de archivos, evitando tener que migrar de formato en el futuro. Efectivamente, ese recaudo tiende a prevenir la obsolescencia digital.

Caso 2: Realizar backups no está relacionado con la obsolescencia del formato.

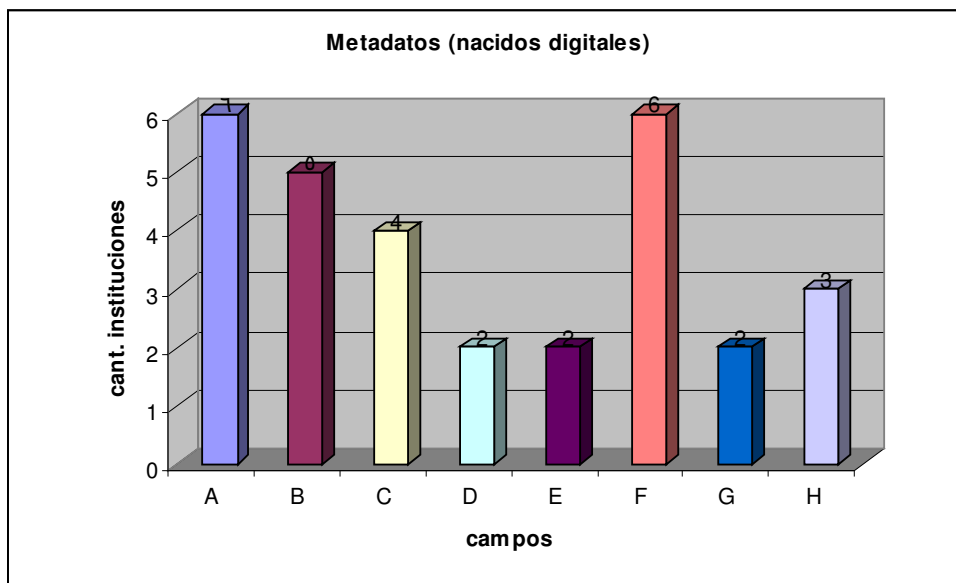
Caso 3: Refreshing, o copia de refresco, sirve para mantener en buenas condiciones a los soportes, no se vincula con los formatos particularmente.

Caso 4: OK

Caso 5: PDF es un formato que puede cambiar. Eso ya ocurrió en el pasado. HTML es más estable, pero no tiene la flexibilidad de XML.

Caso 6: Está OK.

Metadatos registrados con las colecciones nacidas digitales (respuestas agregadas)



Referencias del gráfico

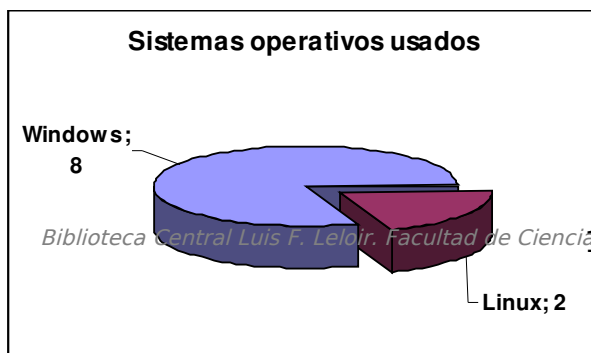
- a) Nombre del archivo digital
- b) Fecha de creación del archivo
- c) Extensión del archivo
- d) Software de creación
- e) Software de lectura
- f) Información sobre el documento físico equivalente, si existe
- g) Información sobre derechos de acceso y uso
- h) Otros

¿Cómo se registran esos metadatos?

- I) en el sistema de catalogación de la institución (indique el software)= **7 casos**
- J) en una base de datos aparte (indique el software) = **1 caso**
- K) en el header de la imagen = **ningún caso**
- L) De otro modo = **1 caso**

Desagregado por software:

- I) ISIS: 4 casos; FAO/WebAgris: 1 caso; Software propio: 1 caso; sin especific.: 1 caso. **Total 7 casos**
- J) Lotus Notes: 1 caso
- L) Automático por Greenstone: 1 caso; Excel: 1 caso; sin especific.: 1 caso. **Total 3 casos**

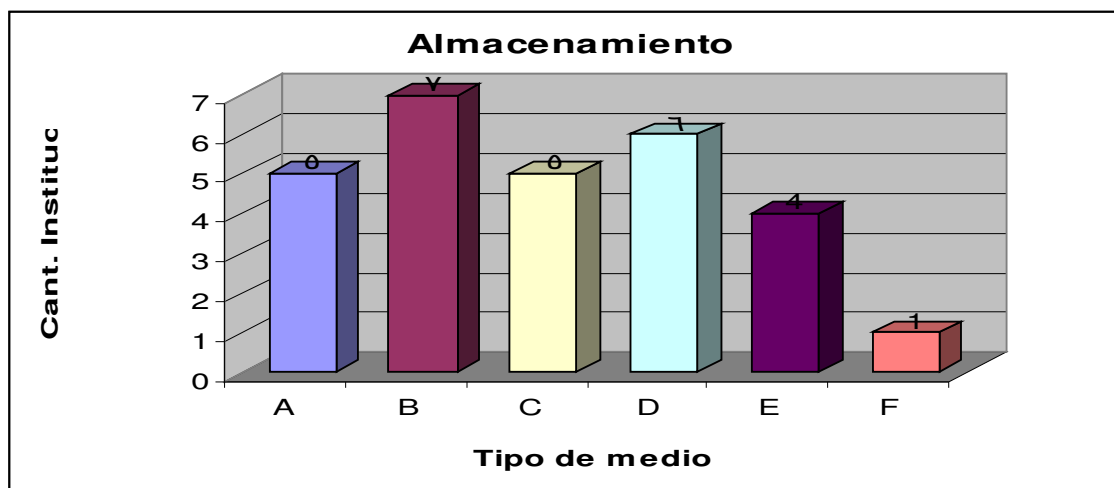


La gran mayoría de las instituciones encuestadas utiliza el sistema operativo Windows para la gestión de

sus colecciones nacidas digitales, pero hay dos de ellas que están usando alguna versión de Linux.

### Tercera Parte. Almacenamiento de los archivos digitales (digitalizados y nacidos digitales)

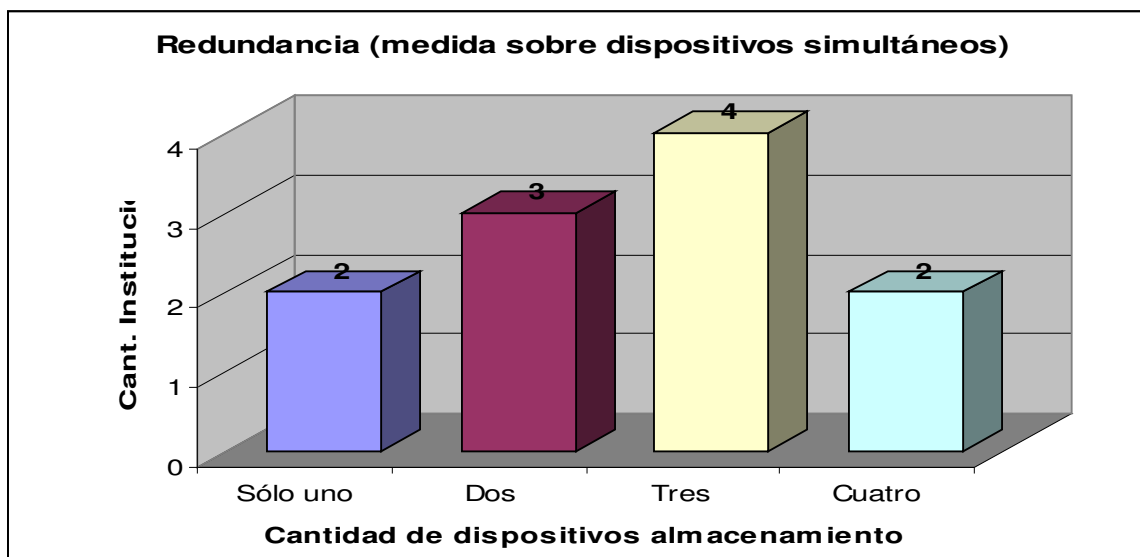
En esta parte de la encuesta hemos relevado distintas cuestiones vinculadas con las estrategias de almacenamiento, y de preservación de los archivos y medios de soporte, en uso por parte de las instituciones. En el primer gráfico, *Almacenamiento*, se presentan las respuestas agregadas (podían marcar más de una) para las opciones de almacenamiento en uso. La opción más representada aquí es el almacenamiento en un disco rígido de un servidor local, seguido de cerca por el uso de CDs, y luego por el uso de discos rígidos de un servidor general de la institución así como de una computadora común de trabajo (5 respuestas positivas en ambos casos). Por último, el uso de DVD está representado por cuatro respuestas positivas, y la opción "otros" por una sola respuesta. Si agregáramos el uso de Cds y DVDs, veríamos que los medios ópticos serían la categoría principal de almacenamiento, con 10 respuestas positivas.



#### Referencias del gráfico

- A) En discos rígidos de un servidor general de la institución
- B) En discos rígidos de un servidor local (del área o departamento)
- C) En disco rígido de una computadora común de trabajo
- D) En CD
- E) En DVD
- F) Otros dispositivos (discos WORM, ZIPdrive, etc.).

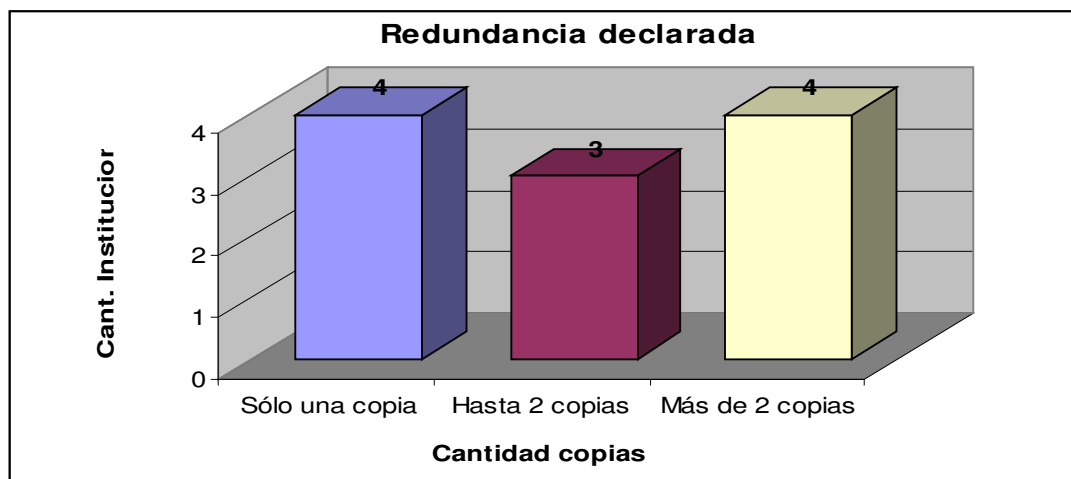
Pero esto sólo nos brinda un panorama general de las tendencias en cuanto al almacenamiento. Dado que la redundancia de copias es una de las estrategias más accesibles para la preservación digital, hemos elaborado el siguiente cuadro, *Redundancia*, para observar mejor cómo actúan las instituciones en este punto. Para ello hemos agrupado a las instituciones bajo una clasificación ordenada por la cantidad de medios de almacenamiento (discos de servidor, de una PC común, CD, etc.) en uso simultáneo, reagrupando las respuestas a esta misma pregunta.



El universo de respuestas refleja las estrategias de once de las instituciones encuestadas, ya que una de ellas no contestó esta pregunta. Aclaremos que esta “redundancia” incluye masters, copias de acceso y backups, sin discriminar. De todos modos, el gráfico es útil para evidenciar que la mayoría de las instituciones comprende la importancia de realizar backups y copias redundantes, aunque no se refieran exclusivamente a los archivos master. El hecho de que 9 instituciones sobre 11 mantengan al menos dos copias de los objetos digitales, y entre esas 9, 6 realicen más de dos copias, muestra una tendencia en favor de la redundancia de copias que merece ser señalada, y permite suponer un terreno favorable para el desarrollo y la aplicación más consistente de esta estrategia. Sólo dos instituciones guardan una sola copia de los objetos digitales, en ambos casos dentro de un servidor local del área que contiene al proyecto. Sería deseable, en este último caso, la realización de al menos una copia de resguardo en medios ópticos. Pero dado el bajo costo de estos medios ópticos, no sería imposible pensar en la realización de más de un juego de copias, y entrar en una estrategia de redundancia más fiable.

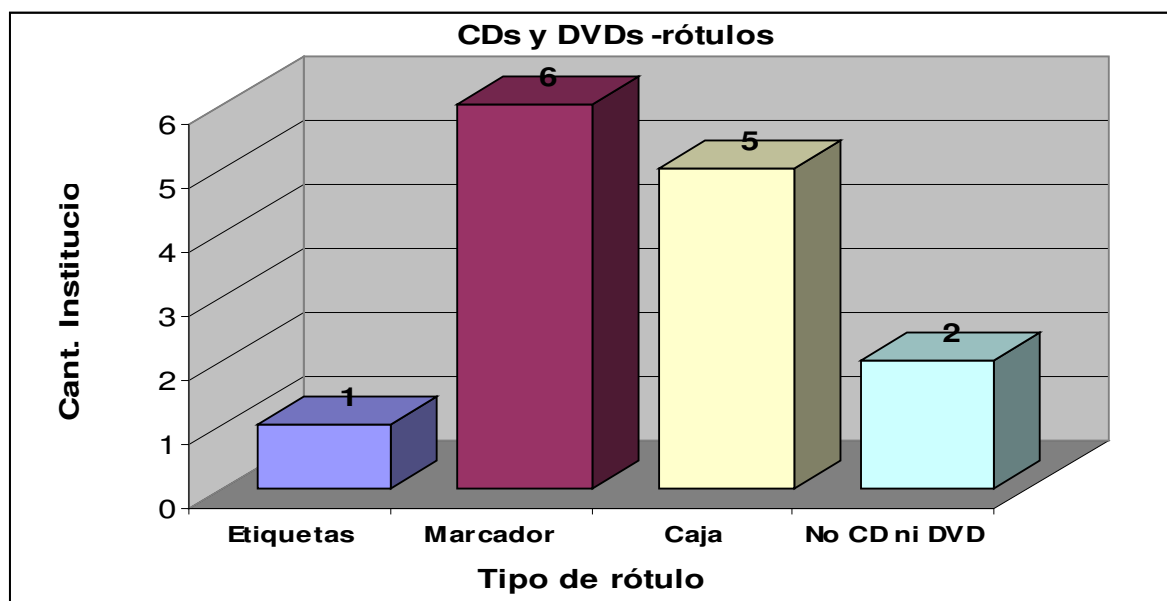
Los datos del cuadro anterior surgen de la pregunta acerca de dónde guardan los objetos digitales (sin discriminar entre masters, copias de uso, etc.), pudiendo marcarse más de una opción. La siguiente pregunta pedía respuestas específicas sobre la cantidad de copias producidas sobre los mismos archivos digitales, pese a que esa información ya estaba implícita en las respuestas a la pregunta anterior. El cuadro siguiente, *Redundancia declarada*, expresa los resultados tal como fueron presentados por las instituciones.

La nueva distribución que se aprecia en el gráfico *Redundancia declarada* se debe a un pequeño juego semántico que introdujo alguna confusión, ya que la pregunta pedía que se defina cuántas **copias** se realizaban. Naturalmente, los que sólo realizan una copia, se reflejan en el cuadro anterior como usando dos dispositivos de almacenamiento simultáneos, ya que la existencia de una copia supone un “original” o master digital ya presente y almacenado. La presencia de 4 respuestas positivas a la opción “sólo una copia” se debe a que las 2 instituciones que en la pregunta anterior respondieron tener almacenado sólo el original, contestaron aquí como si realizaran una copia.



Pese a la confusión señalada se vuelve a advertir que, de las 11 respuestas válidas, en 7 casos se realizan al menos **dos copias**, en distintos soportes (lo que sugiere la existencia de tres “instancias” de cada objeto digital: un master más dos copias). Considerando la juventud de este campo emergente en nuestro medio, es auspicioso que la mayoría de las instituciones intente acercarse a una estrategia basada en la redundancia de copias.

Se preguntó a las instituciones si utilizaban medios ópticos (CD, DVD) de almacenamiento, y cómo los identificaban, ya que aquí se presentan riesgos sobre la integridad de estos soportes. En el siguiente gráfico, *CDs y DVDs - rótulos*, podemos observar que de las 11 instituciones que dieron respuesta a este punto, 2 no usan estos medios ópticos para el almacenamiento. De las 9 restantes que sí lo hacen, sólo una recurre a las etiquetas autoadhesivas para identificar sus discos, y las demás usan tanto marcadores especiales para CDs y DVDs como rótulos dentro de la caja del disco (no sobre el disco mismo). Conviene remarcar aquí que las instituciones encuestadas siguen, en este punto, buenas prácticas internacionales en su gran mayoría, y sólo una utiliza un procedimiento desaconsejado (etiquetas autoadhesivas sobre el disco) dados los eventuales efectos negativos que podría producir el adhesivo químico sobre la capa externa de los discos ópticos.





## Referencias

**Etiquetas:** etiquetas autoadhesivas sobre el disco

**Marcador:** rótulo sobre el disco con marcador especial para CD y DVD

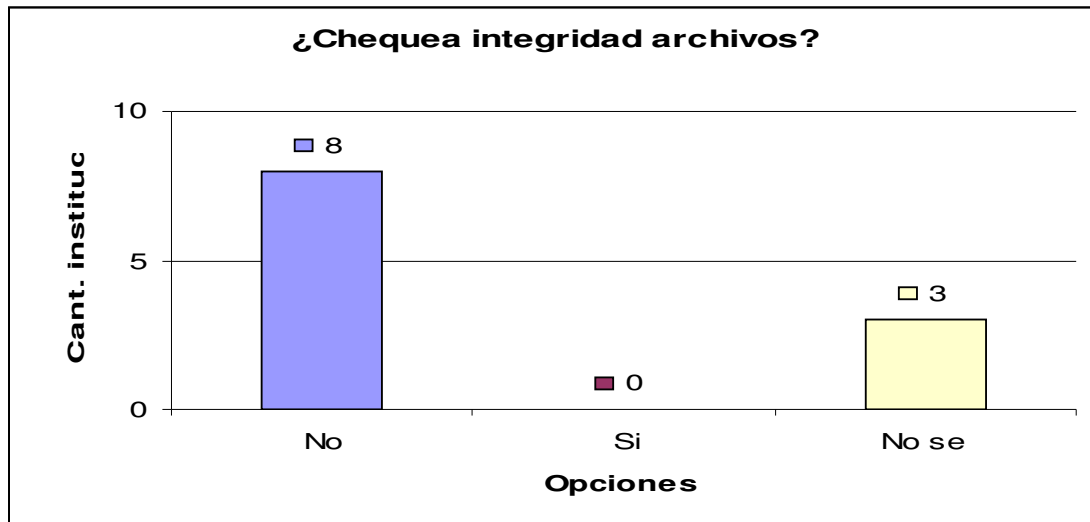
**Caja:** rótulo sobre la caja del disco (no sobre el disco)

**No CD ni DVD:** No usan discos CD ni DVD

### 44) ¿Se realizan chequeos de integridad de los archivos digitales almacenados?

NO SI NO lo sé

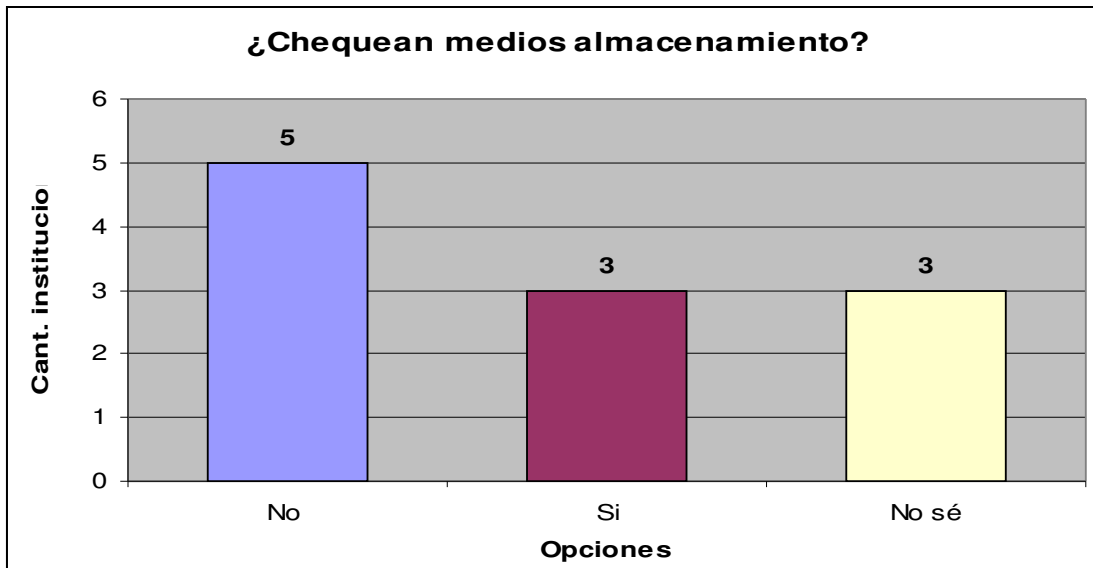
Si contestó afirmativamente, describa el método (checksum, hash, md5, etc.)



De las 11 respuestas válidas a esta pregunta, se observa que no parece haberse implementado aún ningún procedimiento para comprobar que los archivos digitales se mantengan inalterados. Aunque las rutinas de software para realizar esta tarea son relativamente sencillas, tienen el problema de utilizar mucho tiempo de proceso -cuando las colecciones alcanzan magnitudes de decenas de miles de objetos-, y sólo brindan una detección temprana de problemas, pero no ofrecen modo alguno para resolverlos.

### 45) ¿Se realizan chequeos de integridad de los dispositivos de almacenamiento (discos rígidos, CD, DVD, etc.)?

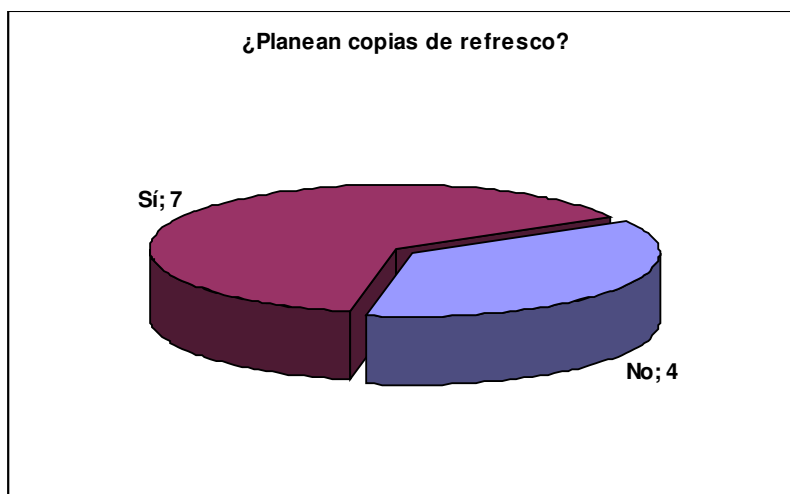
NO SI NO lo sé



Sobre la realización de chequeos del buen estado de los medios de almacenamiento, sólo tres respuestas lo afirman positivamente. Es posible que las tres respuestas a la opción “no lo sé” oculten chequeos realizados en discos fuera del alcance de los encuestados (por ej.: de un servidor general). De todos modos, con las cinco respuestas negativas más las tres que lo ignoran, y conociendo por los gráficos anteriores el alto uso de servidores locales y discos ópticos, este punto debería revisarse para mejorar las condiciones de detección temprana de problemas que podrían afectar la persistencia de la información digital almacenada.

46) Si utilizan CD o DVD u otros dispositivos removibles para almacenar archivos, ¿tienen planeado realizar copias de refresco de esos soportes?

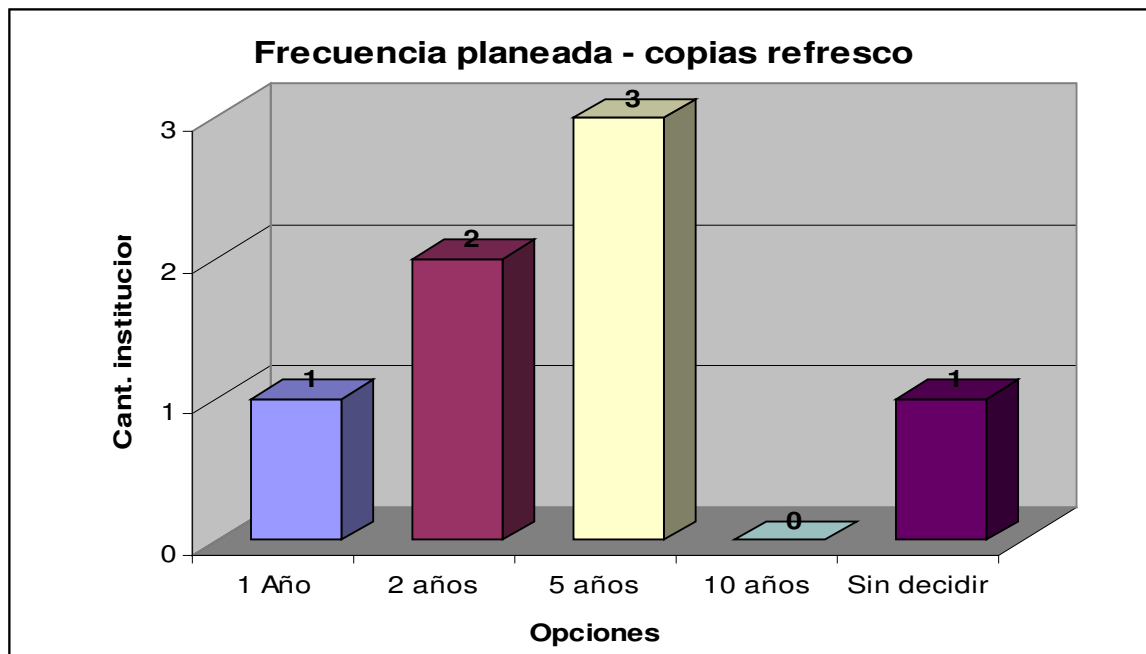
NO SI



En este punto, sobre 11 respuestas válidas, 7 planean realizar copias de refresco y las 4 restantes no. Entre estas 4 últimas, en dos casos la negativa se explica porque no usan medios de almacenamiento removibles.

Si contestó afirmativamente, ¿con qué frecuencia se proponen realizar las copias de refresco?

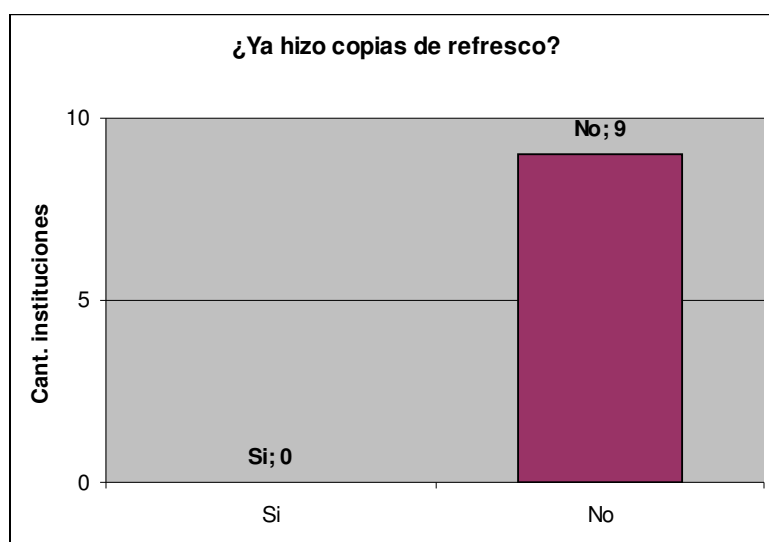
Cada año - Cada 5 años - Cada 10 años - Otro plazo



De las 7 instituciones que se proponen recurrir a las copias de refresco como un medio de prevenir los efectos del decaimiento de los soportes removibles, la mayoría (3) piensa hacerlas cada 5 años, 2 planean refrescar cada 2 años, y 1 anualmente. Finalmente, en un caso no han decidido aún la frecuencia que habrá de aplicarse.

47) **A la fecha, ¿ya se han producido copias de refresco de los soportes removibles en uso (CD, DVD, etc.)?**

SI NO No usamos soportes removibles



Es interesante, como forma de comprobar por otra vía la juventud de estas iniciativas en Argentina, observar que aunque muchas instituciones tienen planes para realizar copias de refresco, e incluso algunas en ciclos anuales o bienales, a la fecha ninguna de ellas ha realizado todavía su primer ciclo de refresco. De todos modos, sin duda es muy auspicioso que las instituciones hayan

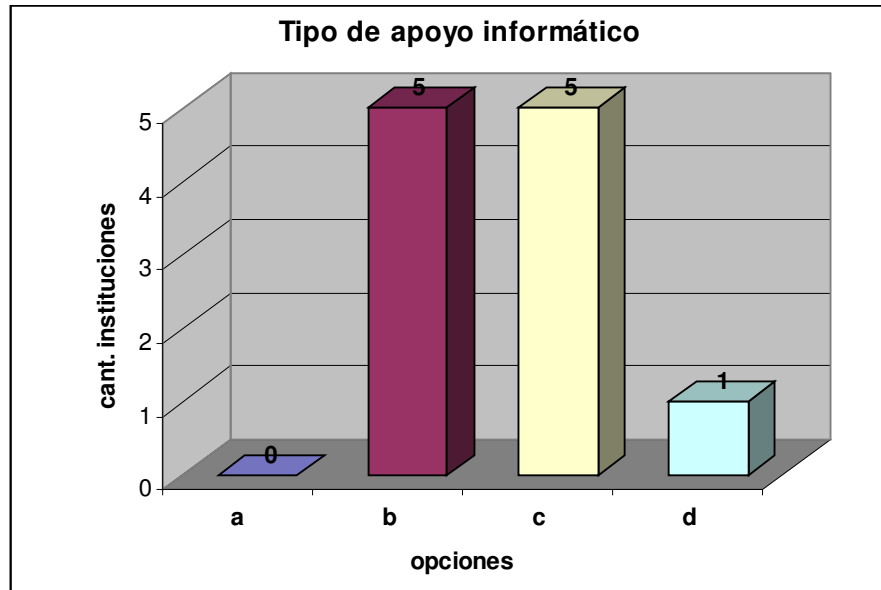
previsto estas tareas preventivas en sus agendas de trabajo.

49) **Defina el alcance del apoyo informático con que cuenta el programa o proyecto, para la creación, gestión y mantenimiento de las colecciones digitales o digitalizadas:**

- a) Informático dedicado exclusivamente al "proyecto digital"
- b) Informático no exclusivo del proyecto, pero específico del área mayor (p. ej.: informáticos propios)

- de la Biblioteca, dentro de una Facultad, etc.) que cobija al proyecto
- c) Informático general de la institución
- d) No contamos con apoyo informático

La primera evidencia que surge de este gráfico es que ninguno de los proyectos o actividades vinculados con la creación y gestión de colecciones digitales cuenta con apoyo informático dedicado en forma exclusiva.

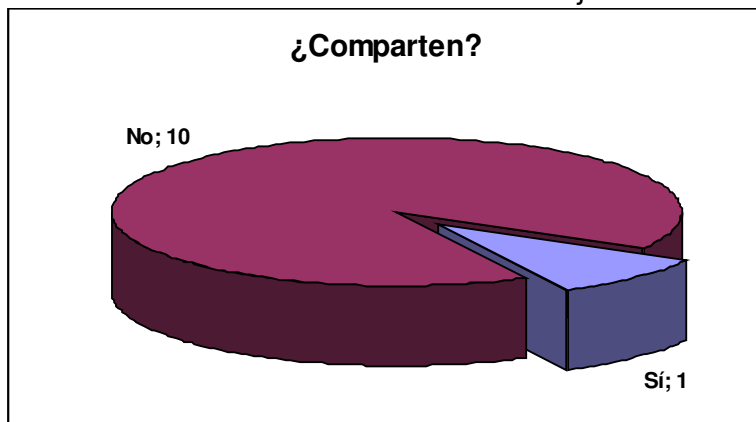


Sin embargo, 5 cuentan con apoyo informático

específico del área mayor que cobija al proyecto, lo que podría significar que acceden con cierta facilidad o inmediatez a ese apoyo, y que el o los informáticos están bastante compenetrados con las funciones y tareas requeridas por el área mayor (p ej.: la biblioteca de una universidad). En tercer término, otras 5 iniciativas reciben apoyo informático general de la institución. Puede que esto determine una menor inmediatez en el acceso a ese soporte, pero esa situación depende mucho de situaciones particulares no evaluadas en este cuestionario. Finalmente, en el caso de una institución se contesta que no reciben ningún tipo de apoyo informático.

**50) ¿Comparten la producción, la gestión y/o el almacenamiento de los archivos digitales con otra institución?**

En otra muestra más de la relativa juventud de estas iniciativas en nuestro



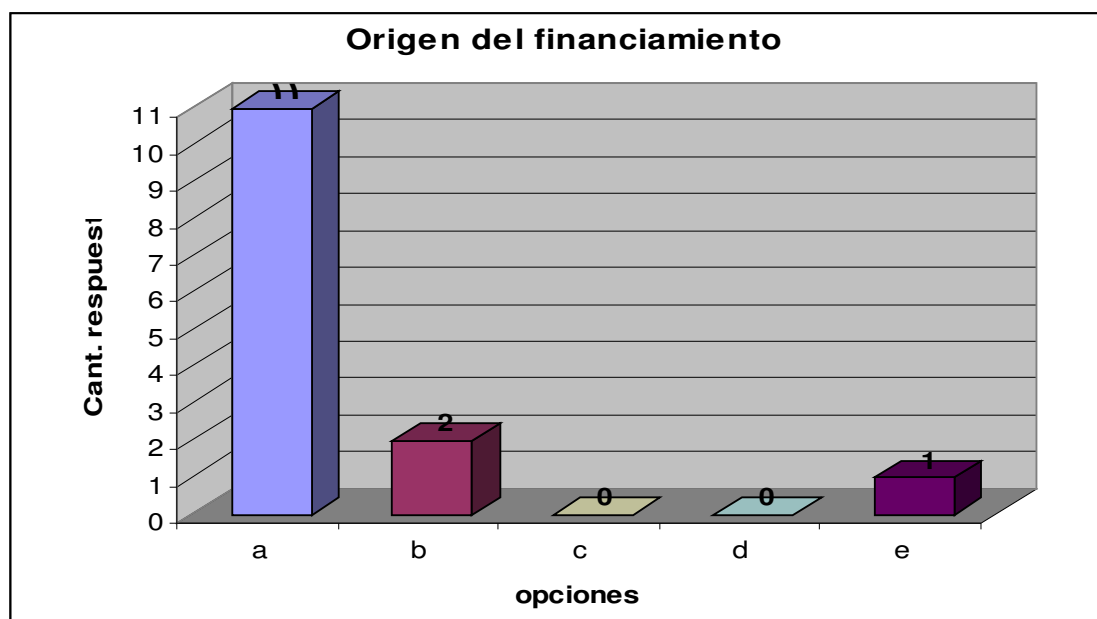
medio, la gran mayoría (10 respuestas) desarrolla sus proyectos en sus propios ámbitos institucionales, sin compartir la producción, la gestión o el archivo de las colecciones digitales con otras instituciones. Sin embargo, en un caso sí se lo hace, aunque esto se explica por el carácter mismo de la organización a la que pertenece, la que

funciona como una red de instituciones a escala supranacional. En los países más avanzados en esta problemática, es común el desarrollo de iniciativas conjuntas, buscando la suma de fuerzas y conocimientos aportados por las distintas partes.

Por otro lado, es natural, hasta que no se logre un desarrollo mayor de este campo en Argentina, que no sea tan fácil pensar en esos emprendimientos en conjunto –más allá de las “culturas institucionales” que no siempre favorecen esa forma de trabajo–. Sin embargo, sería muy beneficioso que los proyectos actuales, y los que se inicien de aquí en más, busquen la forma de trabajar compartiendo estándares y buenas prácticas emergentes en el mundo, lo que ayudaría a generar colecciones más “preservables”, y tal vez más “interoperativas” también. Dado que el desarrollo de estas colecciones ocurre dentro del mundo de las computadoras y de Internet, las cuestiones de la preservación, del acceso distribuido y de la interoperatividad deberían ser preocupaciones centrales de cada uno de los proyectos en este campo.

**51) ¿Reciben financiamiento de alguna de las siguientes fuentes para desarrollar el proyecto / programa? (Marque todas las que correspondan)**

- a) de la propia institución
- b) de un ente nacional estatal (ej: Ministerio de Educación)
- c) de un ente nacional no estatal (fundaciones sin fines de lucro, etc.)
- d) de una empresa privada
- e) de una fundación o programa internacional (ej: UNESCO, Mellon Foundation, etc.)



La última de las preguntas nos permite observar que la gran mayoría de los proyectos reciben financiación exclusivamente de la propia institución a la que pertenecen (se da por descontado que siempre existe financiación de la propia institución, por obvias razones). Sólo en dos casos se accede a fuentes de financiación adicionales, tratándose éstas de organismos públicos nacionales. Sólo una iniciativa recibe fondos de fundaciones internacionales –ayudada sin duda por la capacidad y experiencia de su institución madre para gestionar este tipo de subsidios–, y no configura un ejemplo representativo del conjunto. Por último, ningún proyecto declara recibir fondos, subsidios o donaciones, provenientes de fundaciones (OSFL) nacionales ni de empresas privadas. Si vemos estas cifras con la óptica del “vaso medio lleno”, sin duda queda un largo camino por recorrer en lo tocante al financiamiento de las iniciativas de digitalización, acceso y preservación digital en Argentina.

## Información agregada en cuadros

Hemos dejado para el final el análisis de las respuestas a las preguntas 5, 9 y 27 de la encuesta, en donde se pedía información agregada mediante la confección de diversos cuadros que relevaban el estado de las colecciones digitalizadas y de las nacidas digitales. Entre 8 y 9 instituciones completaron estos cuadros, con una gran dispersión en las respuestas, por lo que se impone un cuidado especial al tratar de evaluar esta información.

Respecto de la pregunta 5, como ya se ha dicho más arriba, se observa que el total de imágenes producidas asciende al número de 35.000, y por amplia mayoría se trata de la digitalización de imágenes fijas de material textual, y en mucho menor medida de materiales fotográficos (sólo una institución digitaliza archivos de video y audio, todavía en cantidades modestas).

El siguiente cuadro agrupa las respuestas a la Pregunta 9, referida a los parámetros de producción de las imágenes digitales.

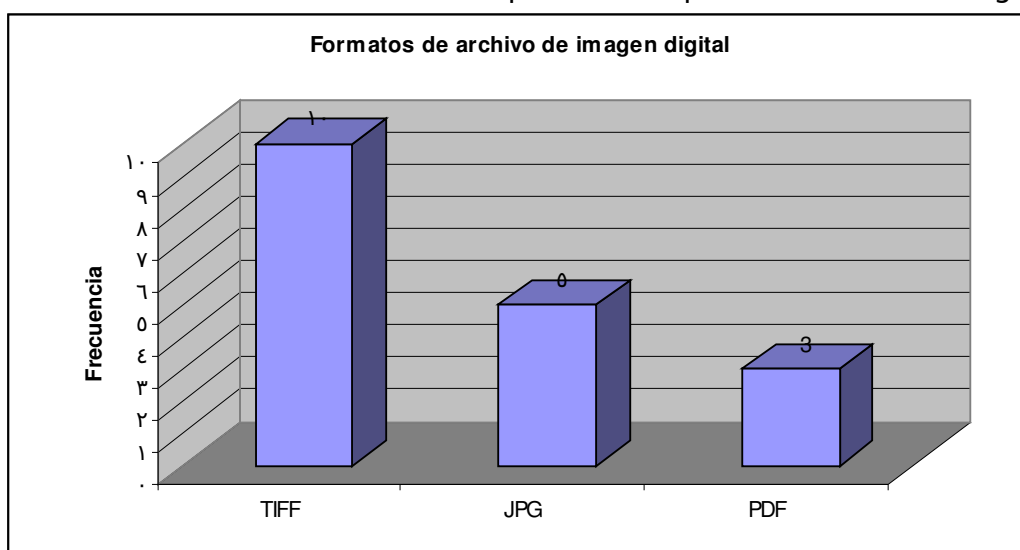
Tipo de imagen	Resolución (dpi)	Formato archivo	Compresión	Tipo de original reproducido	Cantidad imágenes
<b>Bitonal</b>	300	TIFF y PDF	NO	Textos impresos	17.964
	600	TIFF		Textos Impresos	45
	300	PDF	NO	Programas académicos	430
	300	TIFF	Grupo IV (declara "sin compresion")	Publicación original	s/d
	<b>Subtotal</b>				<b>18.439</b>
<b>Tonos de grises</b>	300	TIFF	NO	Ilustraciones en impresos	50
	150	JPG	"Estandar"	Impresos (libros)	250
	300	TIFF		Impresos texto e imagen	s/d
	No sabe	JPG		Paginas de revistas	s/d
	300	TIFF y JPG	No y 10:1	Impresos, libros antiguos	s/d
<b>Subtotal</b>				<b>300</b>	
<b>Color</b>	200	JPG y PDF	"Estandar"	Libros y revistas	3.658
	300	TIFF	NO	Impresos en color	612
	300	TIFF		Impresos varios	s/d
	3000	TIFF	NO	Copias foto, negat y Diago	1
	300 / 600	TIFF y JPG	NO y 10:1	Diapos, libros antiguos	s/d
<b>Subtotal</b>				<b>4.271</b>	

La suma total de 23.010 imágenes, inferior a la de 35.000 ya mencionada, se explica porque algunas instituciones que digitalizan originales no contestaron esta pregunta, provocando un descenso del número total de imágenes declaradas en este cuadro. Lo primero que salta a la vista es una confirmación del tipo predominante de originales digitalizados, constituido básicamente por diversas clases de textos impresos.

Respecto del uso de imágenes bitonales, en tonos de grises o en color, la diferencia básica que se observa está vinculada con la naturaleza de los originales, tales como presencia de ilustraciones u otras características de las

publicaciones que pueden representarse mejor por el uso de grises y de color. Con todo, las imágenes bitonales parecen predominar, lo que no es extraño ya que la mayoría de los textos impresos sin ilustraciones pueden resolverse con este tipo de imágenes. Entre las cinco instituciones que declaran digitalizar imágenes en color, dos lo hacen sobre materiales fotográficos, con lo que se evidencia que la mayoría procesa a los textos impresos con imágenes bitonales o en tonos de grises, lo que es consistente con las prácticas internacionales sobre este tipo de originales.

En el siguiente gráfico vemos que entre los formatos en uso para guardar las imágenes no hay una gran dispersión, ni tampoco aparecen formatos propietarios de poco uso, sino que las preferencias se reparten entre TIFF, PDF y JPG, todos de amplia difusión en el mundo digital. Se debe recordar que las instituciones suelen usar estos formatos simultáneamente, por lo que el gráfico mide la frecuencia relativa de uso para cada tipo de archivo de imagen digital.



Hilando un poco más fino vemos que al momento de producir imágenes bitonales, una institución aplica una resolución de captura de 600 dpi y guarda en formato TIFF (práctica internacional para este tipo de masters), mientras que las tres restantes capturan a 300 dpi y guardan en TIFF y/o PDF (JPG no es aplicable en las bitonales). Para las de tonos de grises, en tres casos aplican 300 dpi y usan TIFF (se puede ver el intento de generar un master de archivo digital), una aplica 150 dpi y guarda en JPG (orientado sólo al acceso), y en el último caso no se sabe la resolución y usan sólo JPG. Entre las cinco que producen imágenes en color, una declara capturar entre 300 y 600 dpi, otras dos en 300 dpi, una en 200 dpi, y por último una institución digitaliza materiales fotográficos, no textuales, por lo que emplea resoluciones más altas.

Pese a lo exiguo de estos datos, una proporción interesante de respuestas afirma que no usan compresión al guardar las imágenes digitalizadas, lo que también habla del intento de generar un archivo master.

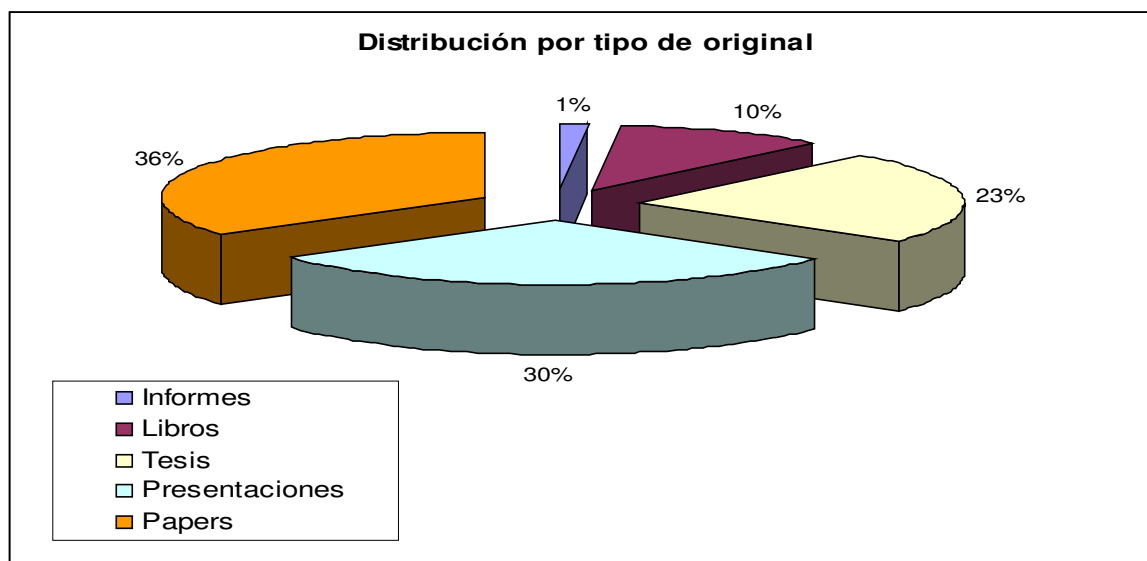
Por último veremos las respuestas a la pregunta 27, relativa a las características de las colecciones nacidas digitales.

Tipo de original	Cantidad documentos	Cantidad archivos digitales
Informes	62	30

Libros	429	102
Tesis	973	1.196
Presentaciones	1.247	157
Papers	1.435	1.045
Otros	4.590	2.944
<b>Total</b>	<b>8.736</b>	<b>5.474</b>

En este cuadro, con la totalidad de las respuestas agregadas, vemos la composición de las colecciones nacidas digitales. Hay cierta inconsistencia numérica entre la cantidad de documentos y la cantidad de archivos digitales que los representan. Esto podría deberse a que no todas las instituciones pudieron brindar información detallada para cada uno de los ítems solicitados en la pregunta, aunque no podemos descartar que existan archivos digitales (un PDF, por ejemplo) que contengan más de un tipo de documento en el mismo archivo digital. Sea como fuere, el total de documentos electrónicos se acerca a 9.000, mientras que la cantidad de archivos digitales que los componen está por debajo de 6.000.

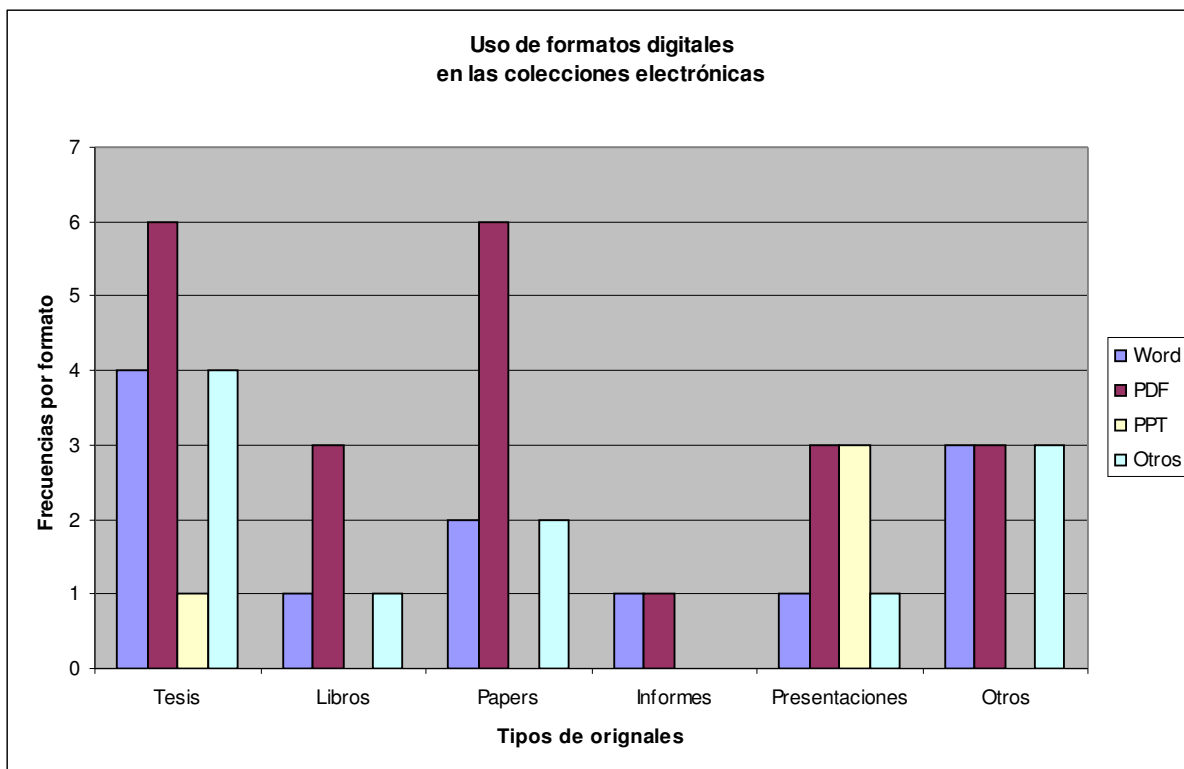
Considerados por tipo de original, y eliminando la categoría "otros", se observa el predominio de los papers y las presentaciones nacidas digitales (66 % del total entre los dos tipos), seguidos por las tesis electrónicas (casi un cuarto del total), los libros electrónicos y los informes, en ese orden.



Esta distribución parecería razonable, dado que se trata de documentos nacidos digitales, y en ese caso las presentaciones de tipo PowerPoint y los papers académicos en formatos de procesadores de texto o bien en PDF van poblando las bibliotecas a partir de la contribución de los autores. El caso de las tesis electrónicas es más complejo y suele requerir una gestión y publicación digital más elaborada (por ejemplo, es un material más propenso a presentar múltiples archivos digitales por cada unidad intelectual).

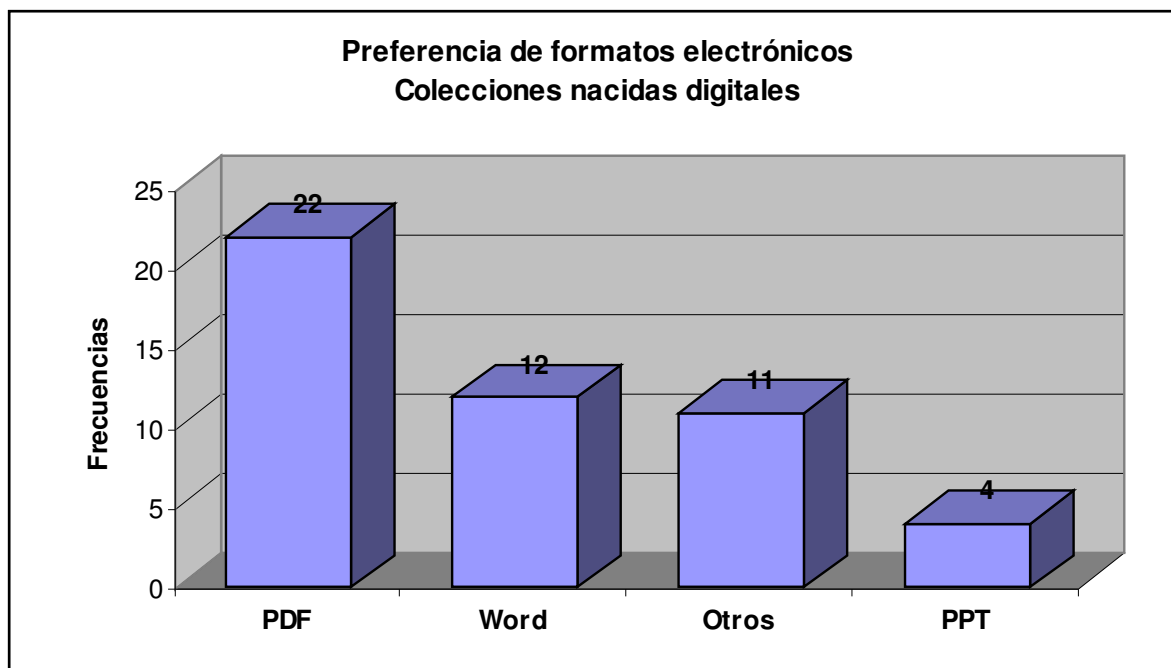
Vamos a ver entonces cómo son las preferencias de formatos electrónicos para cada tipo de documento original, recordando que las instituciones sostienen simultáneamente diversos tipos de "originales electrónicos".





El gráfico nos muestra, para todos los tipos de originales electrónicos, que la preferencia se inclina hacia el uso de PDF como formato de distribución, seguido por el uso de formatos de procesadores de texto y HTML (este último incluido en la categoría “otros”, en color celeste, que significa “otros formatos digitales”). Con todas las precauciones ya mencionadas sobre números tan pequeños de respuestas, se observa que el uso de formatos tipo PowerPoint sólo es relevante en el tipo “Presentaciones” (lo cual es tautológico), pese a que allí comparte la preferencia con el PDF, y sólo aparece una vez en el resto de los tipos de documentos, específicamente en el caso de las tesis electrónicas. Es interesante notar que los formatos tipo HTML o de texto puro (RTF) están bien representados (por la barra celeste de “otros formatos digitales”), lo que seguramente debe relacionarse con la necesidad de publicar estos textos a través de Internet.

Reagrupando esta información respecto de los formatos predominantes en las colecciones electrónicas, obtenemos un “ranking” de preferencias (respuestas agregadas) que presentamos en el siguiente gráfico:



Aquí se puede apreciar mejor lo que antes afirmábamos sobre el predominio del formato PDF. Es interesante notar que la categoría "Otros" agrupa formatos tales como HTML, XML y RTF (formatos más "preservables" que los de procesadores de textos propietarios). Desde el punto de vista de la preservación de las colecciones de textos electrónicos, habría que enfatizar el uso de formatos abiertos y bien documentados (XML, por ejemplo). Dada la preferencia por el PDF, y aunque algunas instituciones lo usan en conjunto con formatos de texto puro, se podría intentar acercarse a la especificación de PDF/A, desarrollada con el objetivo de preservar el acceso a los archivos así creados.

Lic. Fernando Boro  
Marzo de 2008