

[Texto interpretado al español de Joudrey, D. N. ; Taylor, A. G. & Wisser, K. M. (2018). *Organization of information*. (4th ed.). Santa Barbara, CA.: Libraries Unlimited. I. Organization of Recorded Information, por el Prof. Gerardo Salta, para su uso en el dictado de clases en la Carrera de Biblioteología, ISET N° 18 “20 de Junio”, Rosario]

Cap. 1 Organización de la información registrada

El Merriam Webster’s Collegiate Dictionary define organización en distintos sentidos. Uno de ellos se asocia con “integrar”, constituir una unidad coherente o funcionar en una totalidad. Otro significado reside en ordenar elementos dentro de una totalidad de partes independientes. Este libro se dedica a explicar el proceso de brindar unidad y ordenar elementos, tal como lo efectúan las personas que trabajan en lugares donde se acumulan recursos de información usados por personas, tanto de manera inmediata como para la posteridad. [...]

La necesidad de organizar

Parece haber una urgencia humana para organizar. Los psicólogos establecen que el cerebro del bebe organiza imágenes por categorías, como caras o comidas. Los niños pequeños llevan a cabo organizaciones y correspondencias durante sus juegos. Por ejemplo, los niños pueden dividir sus juguetes por forma, color, tipo, antes de jugar o como una forma de juego. Muchos juguetes son clasificados y algunos de ellos se diseñan para ayudar a los niños a desarrollar la habilidad de categorizar una parte importante del funcionamiento cognitivo humano. En la medida que crecen, los humanos desarrollan habilidades cognitivas más sofisticadas para categorizar, reconocer patrones, clasificar, relacionar y crear grupos de cosas e ideas [...].

Muchos individuos tienen una necesidad de organización más fuerte que otros. Aquellos que comulgan con esta afirmación, no pueden comenzar a trabajar hasta que la zona de trabajo esté ordenada y los objetos estén en su lugar [...]. No obstante, aun aquellos cuyos espacios de trabajo aparecen como caóticos, pueden tener alguna organización en su cabeza. Estas personas, generalmente, tienen alguna idea o conocimiento de que cosas están en las diferentes pilas o colecciones. En términos de la gestión de información personal, las personas suelen categorizarse como archivadores (filers) y apiladores (pilers), dependiendo de cómo organizan sus propios materiales. (Las dos categorías se explican por sí mismas). Independientemente del estilo personal individual, no obstante, el aprendizaje humano está basado sobre la habilidad para analizar, organizar y recuperar datos, información y conocimiento, para reconocer patrones, comparar experiencias, conceptos e ideas y para procesar las relaciones entre todas ellas.

¿Porqué organizamos? Además del placer de poner cada cosa en su lugar (sea en un archivo, en un estante o en una pila particular), organizamos por razones más significativas:

- Comprensión (Understand): la organización nos ayuda a darle sentido a muchas cosas en la vida diaria. Por ejemplo, uno puede vivir en Cambridgeport, un distrito de Cambridge, una ciudad del condado de Middlesex, que es una división política en la Commonwealth de Massachusetts, que es un estado localizado en New England, en el noreste de EE.UU. existen sistemas de organización geográficos formales e informales

que nos ayudan a comprender donde estamos y para comunicar ubicaciones geográficas a otras personas.

- Ahorrar tiempo: nos organizamos para ser rápidos y eficientes. Queremos acceder a nuestra información, cosas, servicios, de manera oportuna. Por ejemplo, si los documentos en una caja de archivo no están agrupados de acuerdo a categorías o series, podemos tener que categorizar entre cientos o miles de piezas de papel para encontrar lo que necesitamos. Esto tomaría mucho tiempo.
- Reunión (Collocate): organizamos para reunir ideas o cosas similares juntas en grupos. Por ejemplo, si tenemos muchos libros sobre el mismo tema o grabaciones sonoras del mismo género musical, es útil etiquetarlos consistentemente de forma que puedan ser buscados y presentados juntos. Cuando encontramos un libro que es útil o alguna música que es atractiva, podemos descubrir recursos similares próximos.
- Recuperación (Retrieve): por encima de todo, organizamos porque necesitamos recuperar, el mundo está lleno de objetos e ideas. Necesitamos encontrarlos, lo que sea y donde sea que se encuentre. La organización hace esto posible.

La organización de la información es empleada en ambientes formales de información, tales como bibliotecas y archivos, aunque muchas de nuestras acciones diarias dependen del resto del mundo alrededor del que estamos organizados. Las cocinas se organizan de manera que los accesorios para cocinar estén fácilmente accesibles y que los alimentos y los especias se utilicen en la medida que se necesiten [...]. Los ítems en un periódico se organizan en secciones específicas de manera que las historias similares se presenten juntas y las historietas se separan de los obituarios. Las colecciones personales de música se organizan de manera que podamos encontrar la canción correcta cuando la queramos. Los lugares de trabajo se organizan de forma que los documentos sean recuperables, pudiendo llevarse a cabo el trabajo. El proceso de aprendizaje está organizado de manera que las relaciones entre ideas puedan usarse para asistir al estudiante con la finalidad de recordar el material aprendido [...]. La organización es una parte importante de la vida diaria, pudiendo ser clave para la eficiencia y el progreso.

La recuperación de información efectiva y eficiente depende de si ha sido organizada. La información se necesita en todos los aspectos de la vida—por ejemplo, por razones de salud, entretenimiento, para la comprensión de las cosas, para aprender acerca de una relación, o simplemente expandir nuestro conocimiento. Alguna de esta información ha sido ya asimilada y se encuentra reservada en el conocimiento, mientras que otra información tiene que ser pensada. Si esta información no es organizada, es difícil y a veces imposible encontrar (¿Podes imaginarte los inconvenientes de usar un diccionario impreso que no está ordenado alfabéticamente?). Así contamos con varias herramientas que están organizadas para ayudar en el proceso de encontrar información que necesitamos: directorios, diccionarios, enciclopedias, bibliografías, índices, catálogos, bases de datos.

Las razones por las que organizamos información puede ser resumida utilizando lo que expresa FRBR, un modelo conceptual de IFLA que identifica componentes clave del universo de la información registrada. FRBR establece que organizamos nuestras colecciones de recursos de forma tal que los usuarios puedan lograr las siguientes tareas o usos (tasks).

Tarea	Explicación	Ejemplo
Encontrar (Find)	Buscar entidades que se correspondan con criterios específicos	Encontrar <i>El hacedor</i> de Jorge Luis Borges en el catálogo a través de una búsqueda o a través del browsing
Identificar (Identify)	Confirmar que una entidad corresponda con los que es pensando; distingue entidades similares	Identificar que el autor encontrado es el argentino Jorge Luis Borges, que escribió <i>El hacedor</i> , y no el brasileño Jorge Borges, un biólogo que escribió acerca de la sustentabilidad
Seleccionar (Select)	Elegir un recurso que es apropiado a las necesidades de los usuarios	Seleccionar la versión de <i>El hacedor</i> que está en la lengua que comprenda el usuario
Obtener (Obtain)	Lograr acceso al recurso descrito	Obtener <i>El hacedor</i> de la biblioteca u obtenerlo de otra institución

Recientemente, en la actualización del modelo FRBR—conocido como IFLA Library Reference Model—se agregó otra tarea del usuario

Tarea	Explicación	Ejemplo
Explorar (Explore)	Descubrir recursos y entidades y lograr mayor comprensión de los recursos	Explorar otras obras creadas por Borges, lo que conduce a descubrir <i>Ficciones</i> y <i>Elogio de la sombra</i>

La organización de la información también nos permite guardar para la posteridad copias de todos los tipos de obras que resultan del esfuerzo humano (por ej., libros, obras de arte, grabaciones sonoras, películas, correspondencias, tweets, documentos gubernamentales). Las bibliotecas, los archivos, los museos y otras instituciones culturales han venido haciendo esto por muchos años y se han movido actualmente al mundo en línea. Este libro, en general, no está interesado en la organización de la información, los recursos o los materiales en empresas comerciales que puedan reunir colecciones de ítems con el propósito de venderlos, en lugar de coleccionar ítems para beneficio de la humanidad y la reserva de nuestro conocimiento colectivo.

La naturaleza de la información

Si se consideran los siguientes términos: comprensión (understanding), dato (data), conocimiento (knowledge), saber (wisdom) e información (information) y se pidiera brindarle algún orden ¿cómo se haría? Una forma de ordenarlos sería indicar desde el menos al más significativo, quedando el siguiente orden: dato, información, conocimiento, comprensión, saber ¿Cuáles de estos están presentes cuando se realiza la organización en bibliotecas, museos, archivos y otras instituciones? Existe el argumento oscilante entre aquellos que creen que se organiza información y aquellos que consideran que se organiza conocimiento. Nuestro sesgo es evidente a partir del título del libro. Pareciera que usamos nuestro conocimiento para escribir este libro, pero hasta que se lea, se comprenda y se integre al propio conocimiento, solo se trata de información. Esta es la razón por la que consideramos que organizamos información—para que otros puedan

encontrarla, leerla, y usarla al agregar a la propia reserva de conocimiento. En tal sentido, se sostiene que leemos, comprendemos y, entonces, integramos la información dentro del propio conocimiento. Comprensión y conocimiento son dos conceptos interrelacionados. Se necesita contar con alguna comprensión para incorporar algo en el conocimiento, pero debe haber una cierta cantidad de conocimientos para comprender nuevas cosas.

De acuerdo al significado establecido en varios diccionarios, el conocimiento existe en la mente de un individuo que ha estudiado un tema, lo comprendió y quizás lo ha agregado a él mediante investigación o por otros medios. Los mismos diccionarios sostienen que información es la comunicación o la recepción de conocimiento. Dicha comunicación ocurre en gran parte a través del registro de conocimiento de alguna manera. En tal sentido, las personas escriben, hablan, componen, pintan, esculpen, e intentan comunicar su conocimiento a los demás de muchas otras formas. [...] Este libro constituye información que puede ubicarse dentro de un esquema de organización del que puede recuperarse para ser estudiado por aquellos interesados en aumentar su conocimiento sobre el tema. De esta forma, hemos elegido usar el término información antes que conocimiento como nuestra expresión de lo que creemos organizar para el beneficio de los demás. Esto no constituye un rechazo de la organización del conocimiento. El conocimiento que existe en el cerebro de las personas se explota en diferentes situaciones. Los autores trabajan en la organización de su propio conocimiento cada vez que escriben. El conocimiento de los bibliotecarios de referencia se utiliza de manera organizada cuando ayudan a los usuarios en la respuesta de interrogantes. La frase gestión del conocimiento (knowledge management)—que se discute más abajo—ha comenzado a utilizarse en la administración de organizaciones. Recientemente, algunos profesionales de la información han comenzado a referir a los esquemas de clasificación como sistemas de organización del conocimiento (KOS, Knowledge Organization Systems).

La naturaleza de la organización de la información registrada

Como se menciona en el prefacio, este libro aborda la organización de la información registrada, ya que hace falta otro tipo de medios para organizar la información que solo ha sido hablada, escuchada o pensada. La información registrada, sin embargo, incluye mucho más que textos. Por lo tanto, en lugar de utilizar palabras como libro o ítem para referirnos a la unidad de información que puede ser organizada, se utiliza el término recursos de información. Un recurso es una instancia de información registrada; puede ser cualquier cosa, desde un tweet de 140 caracteres a una colección de 140 muestras de minerales y rocas. Más allá de las distintas formas en que puede registrarse la información, todos los recursos de información tiene algunos atributos o propiedades básicas en común—como título, creador, tema. Estos atributos como otros, se registran para ayudar a organizar la información; colectivamente, estos atributos pueden ser referidos como metadatos. Los metadatos en su definición más informal pero prevaleciente son “datos acerca de datos”. Esto significa que los atributos usados para describir un recurso de información son metadatos acerca o sobre tal recurso. Los metadatos se discuten a lo largo de este libro, particularmente en el capítulo 4, y en sus aspectos específicos en los capítulos 7-11.

Hagler ha identificado seis funciones del control bibliográfico. El control bibliográfico—que refiere crecientemente a la noción de organización de la información—es el proceso de describir recursos de información y proveer acceso a los nombres, títulos y temas de las descripciones, resultando en registros que sirven como sustitutos (surrogates) para los ítems de información registrada. Estos registros

sustitutos—llamados asientos (entries), registros bibliográficos, o simplemente metadatos—se localizan en herramientas de recuperación de información, donde los registros actúan como punteros (pointers) a los recursos de información actuales. La descripción encontrada en los recursos provee a los usuarios de suficiente información para determinar el valor potencial de recursos sin tener que ver los ítems directamente. Los metadatos se almacenan en una variedad de herramientas de recuperación—que incluyen bibliografías, catálogos, índices, instrumentos de localización (finding aids), registros de museos (museum registers), bases de datos bibliográficas y motores de búsqueda. El listado de Hagler refleja el énfasis de este libro puesto sobre el trabajo de los bibliotecarios. No obstante, la lista refleja las principales actividades involucradas en todo proceso de organización de información

1. *Identificar la existencia de todos los tipos de recursos de información a medida que estén disponibles.* Un libro puede ser publicado o un sitio Web creado o un video streaming lanzado, pero si nadie sabe de su existencia salvo las personas involucradas en su creación, no tendrá ninguna utilidad como información para nadie. La existencia y la identidad puede ser dada a conocer de muchas formas: anuncios de los editores, anuncios a través de e-mail, críticas, listados temáticos, La mayoría de los editores crean catálogos que enumeran sus productos conjuntamente con resúmenes de los mismos. Las fuentes de referencia como *Books in print* son producto de esta actividad. Algunas revistas en línea envían anuncios regularmente por e-mail, un esbozo de la tabla de contenidos, para facilitar a los lectores conocer cuando está disponible una nueva entrega. Algunas organizaciones piden a las personas alistarse para recibir anuncios por e-mail acerca de nueva información disponible en el sitio Web de la organización.
2. *Identificar las obras contenidas dentro de tales recursos.* En muchos casos, un recurso de información es equivalente a una obra artística, literaria o intelectual (por ej., un DVD que contiene una película). No obstante, en algunos casos, un único recurso puede contener numerosas obras individuales. Por ejemplo, un DVD puede ser una compilación de 12 cortometrajes o un portfolio en línea puede contener una colección de varios cientos de fotografías digitales. Tanto la compilación como el portfolio son considerados obras, pero los cortometrajes y las imágenes artísticas individuales contenidas dentro de ellas, también son obras. El nivel en que se describen los recursos depende de cuanta granularidad (i.e., cuanto nivel de detalle) se desea. Por ejemplo, un sitio Web que trata de la vida y obra de una persona famosa, posiblemente tenga obras digitalizadas de ella, material biográfico, anécdotas escritas por contemporáneos de dicha persona, así como otras partes. Los escritos sobre la persona y los acontecimientos posiblemente sean obras importantes por derecho propio y posiblemente tengan que ser identificados en forma separada.
3. *Recopilar sistemáticamente los recursos de información en colecciones.* La actividad de crear colecciones ha sido área de interés de instituciones como bibliotecas, archivos y museos. Pero las colecciones siempre han sido creadas en otras situaciones—por ejemplo, colecciones particulares formadas debido a un interés en un determinado tipo de información, colecciones de documentos internos almacenados en una oficina, colecciones de materiales de enseñanza

para una disciplina alojados en departamentos universitarios, etc. Estas colecciones se van dando a conocer paulatinamente al público a través de sitios Web. Las colecciones incluyen a menudo recursos electrónicos que no existen a nivel local, solo siendo accesibles en línea. Muchas instituciones brindan a sus usuarios el acceso en línea a sus recursos. Algunos de tales recursos son accesibles solo electrónicamente, mientras que otros también están disponibles impresos. Parte del proceso de organización consiste en determinar la necesidad de agregar dichos recursos de forma permanente a la colección. La mayoría de las instituciones desean coleccionar de una manera significativa. a menos que una colección este bien organizada, puede ser difícil determinar que nuevas obras mejorarán la colección. Quienes desarrollan las colecciones confían en los metadatos para evitar agregar algo que ya está en la colección.

4. *Producir listas de estos recursos de información preparadas de acuerdo a reglas normalizadas para citas.* Las listas creadas en esta actividad incluyen bibliografías, índices, catálogos de biblioteca, instrumentos de localización en archivos y registros de museos. Estos son importantes para la recuperación de recursos de información individuales, dado que si se busca un ítem conocido, especialmente uno de naturaleza tangible que requiere una ubicación física, es necesario encontrarlos listado en alguna parte. Dichas listas pueden estar impresas, aunque a lo largo de las últimas décadas, las preferencias se han movido gradualmente hacia la provisión de acceso electrónico, frecuentemente de manera exclusiva. Estas listas tienen niveles diversos de complejidad. Algunas, como las bibliografías, pueden ser muy simples y fáciles de usar; otras, como los catálogos de biblioteca, pueden contener información más completa y compleja, requiriendo mayor sofisticación en los métodos de acceso. La forma en que cada recurso se describe en una lista particular, es determinada por el conjunto de pautas usadas por la institución que crea la lista. Para asegurar la consistencia y la precisión, los recursos son descritos siguiendo las instrucciones de un estándar de contenido (i.e., reglas o un manual de instrucciones para la creación de metadatos). Diferentes tipos de instituciones siguen distintos tipos de estándares de contenido. En EE.UU., por ejemplo, las bibliotecas usan, generalmente, RDA; los archivos utilizan DACS; los museos pueden describir su colección con CCO; y los creadores de bibliografías pueden utilizar manuales de estilo—como Chicago Manual of Style: APA; MLA Handbook). La mayoría de los estándares de contenido, en un nivel mínimo, proveen una visión general del proceso de descripción, una indicación de que elementos de datos son obligatorios y/o deseables y algunas pautas sobre como registrar los metadatos. No todos los creadores de metadatos, sin embargo, adhieren a un estándar de contenido formal cuando producen descripciones.
5. *Proveer nombre, título, tema y otros puntos de acceso útiles a estos recursos de información.* La recuperación exitosa de recursos de información a través de listas depende de la inclusión de metadatos suficientes. Esta actividad que agrega más valor a la utilidad y a la recuperación potencial de una colección, es la provisión de puntos de acceso controlados autorizados de nombre, título y tema a las descripciones de los recursos de información. Un *punto de acceso* es un nombre, palabra, frase o identificador elegido por el catalogador o el indizador y puesto en un campo particular de un registro que describe un recurso. Puede ser usado para obtener tal registro de una herramienta de

recuperación u otro sistema organizado. El control de autoridad consiste de la práctica de utilizar una cadena de caracteres única para representar cada nombre, obra o tema, para lograr consistencia dentro del catálogo o de otra herramienta de recuperación. También involucra la creación de relaciones explícitas entre diferentes nombres, obras o temas. El acceso a través de palabras clave puede ser provisto de manera más o menos automática y al voleo, es decir cualquier información en formato electrónico puede ser encontrada al buscar una palabra que aparece en el recurso de información electrónico. Sin embargo, la búsqueda por palabras clave no siempre es eficiente o precisa, y no cualquier cosa sobre este tópico puede ser recuperado debido a una falta de controles semánticos. No obstante, los resultados de búsqueda a través de palabras clave se tornan cada vez menos satisfactorias cuanto más grande es la colección que es objeto de realización de la búsqueda. La recuperación es más satisfactoria cuando se puede buscar a partir de nombres, títulos y vocabulario controlado que se han creado bajo control de autoridad, generalmente por intervención de seres humanos. Si una persona ha sido identificada por diferentes nombres y/o formas del nombre (por ej., Jacqueline Bouvier, Jackie Kennedy, Jacqueline Kennedy Onassis y Jackie O.), y si este nombre de persona es formulado bajo control de autoridad, entonces un nombre o forma del nombre es elegido como preferido y los demás son considerados como referencias cruzadas. Así, una búsqueda por una forma del nombre recuperará recursos de información relacionados con la persona, independientemente de cual nombre o forma del nombre aparece en un recurso particular. Por ejemplo, el usuario puede buscar Kennedy, Jackie u Onassis, Jacqueline Kennedy, y recibir resultados de búsqueda relevantes. Esto puede lograrse a través del uso de una referencia directa como:

Búsqueda: Kennedy, Jackie

En este catálogo, busque por: Onassis, Jacqueline Kennedy, 1929-1944

O la herramienta de recuperación puede ser configurada para buscar por un nombre preferido automáticamente y continua cuando una de las referencias es usada en la búsqueda. Adicionalmente, si una obra cuenta con distintos títulos en diferentes manifestaciones, la forma más común del título es usualmente seleccionada como el punto de acceso para la obra. Todos los títulos variantes están conectados y la búsqueda por uno de ellos debería permitir a los usuarios navegar en cada versión de la obra, no importando el título de cada versión o edición individual del recurso. En otras palabras, una búsqueda por Hamlet debería recuperar también resultados como Amleto, Príncipe di Dinamarca; Hamlit; Gamlet, si todas estas variantes se encontraran en la lista que es objeto de búsqueda. Y si un sistema de recuperación utiliza vocabulario controlado para representar conceptos temáticos, una búsqueda por una palabra con más de un significado (lo que es característico para la mayoría de las palabras del español) frecuentemente permitirá la distinción entre diferentes significados (por ej., Puentes (Redes informáticas) vs. Puentes (Odontología)), y puede dirigir a los usuarios hacia términos generales, específicos y relacionados. También reúne juntos bajo un término (por ej., Publicaciones periódicas), todos los términos sinónimos y cuasi-sinónimos que pueden ser usados para expresar un concepto (por ej., revistas, magazines).

El acceso controlado por autoridad es poco usado, a menos que los sistemas se diseñen para tomar ventaja de él. Los Sistemas de Recuperación de Información (SRI) diseñados para la búsqueda y la visualización de información no son siempre sofisticados, intuitivos o fáciles de navegar. Sin el apoyo del diseño de sistemas orientados al usuario, los beneficios de la organización de la información pueden verse disminuidos y perdidos de manera íntegra.

6. *Proporcionar los medios de localizar un recurso.* Desde un siglo atrás, este principio ha constituido un valor agregado para las instituciones que cuentan con colecciones. Los catálogos impresos u otros listados creados en estas instituciones brindan información sobre la ubicación física del recurso. La mayoría de los catálogos de bibliotecas, alguna información de museos y varios instrumentos de localización de archivos, están disponibles en línea en la actualidad. En muchos catálogos en línea de bibliotecas, la información de circulación está disponible, de forma que, si un recurso ha sido sacado de la biblioteca, dicha información está disponible para el usuario. Las utilidades bibliográficas—como OCLC, RLIN, WLN—permiten averiguar que bibliotecas poseen físicamente un ítem determinado. Uno puede descubrir que bibliotecas tienen el recurso buscado y pueden averiguar si el recurso está disponible en los sitios de alguna biblioteca particular, o si esta prestado a un usuario (y, algunas veces, si el recurso está pedido y cuando se espera que llegue a la biblioteca). Tradicionalmente, las bibliografías y los índices no brindan información sobre ubicación. Las bibliografías enumeran lo que existe, aunque pocas veces indican su ubicación. Los índices proporcionan la obra mayor en la que puede encontrarse la obra menor que figura en la lista—por ejemplo, en que revista se encuentra el artículo listado—aunque tampoco proporcionan la ubicación física de la obra mayor. Todo esto sigue siendo válido para los recursos tangibles. No obstante, para los recursos electrónicos encontrados en Internet se vuelve cada vez más común indicar su ubicación—por ejemplo, la URL—la dirección Web para el recurso.

Para cumplir con las seis funciones del control bibliográfico de Hagler, son necesarios metadatos adecuados y suficientes. Sin estos metadatos, los recursos de información no pueden volverse conocidos al mundo, las obras pequeñas no pueden ser identificadas, las colecciones no pueden crearse, no pueden elaborarse listas consistentes, no puede garantizarse el acceso, y sería imposible ubicar los recursos. Los metadatos son necesarios para satisfacer estas y otras necesidades. No obstante, los metadatos no se verán de la misma manera en todos los casos. Existen diferencias en cómo se crean los metadatos para diferentes tipos de materiales, en distintos ámbitos, entre diferentes comunidades y en diferentes contextos.

¿Cómo se lleva a cabo la organización de la información en diferentes ámbitos?

Existen muchos ámbitos en los cuales se presenta el deseo de organizar la información de modo que se pueda llevar a cabo la recuperación de la información para diversos fines y que dicha información sea conservada para la posteridad. A continuación se mencionarán las bibliotecas, los archivos, los museos, los ambientes y contextos en línea—como Internet, la Web Semántica, las colecciones digitales, la arquitectura de la

información—la indización y resúmenes, la gestión de registros, la gestión de información personal y la gestión del conocimiento.

Bibliotecas

Se consideran las bibliotecas en primer término dado que son las que cuentan con la más larga tradición en organización de la información con fines de su recuperación y la conservación para la posteridad. En la mayoría de las bibliotecas, las actividades de organización de la información se centran en el departamento de servicios técnicos. El término *servicios técnicos* refiere a las actividades “detrás de escena” que tratan con el desarrollo, la adquisición, la organización y la preparación de la colección de recursos de la biblioteca. Adicionalmente, pueden existir otras unidades dentro del departamento, focalizadas en publicaciones seriadas, recursos digitales, preservación, conservación, archivos o tecnología. Sin embargo, no existe un acuerdo universal acerca de que unidades pertenecen al departamento de procesos técnicos, pudiendo tener cada biblioteca una configuración diferente de las unidades ubicadas bajo la denominación de servicios técnicos. Por ejemplo, en algunas bibliotecas, el desarrollo de la colección puede estar más estrechamente asociado con los servicios de referencia y dirigidos al público (las operaciones “frontales” de la biblioteca) o este puede contar con su propio departamento. En la Figura 1.1 se provee un diagrama muy básico de las actividades de los servicios técnicos [...].

Como se menciona al inicio, el proceso de organización de información registrada comienza con las colecciones. Las colecciones en las bibliotecas son creadas a través del proceso de **desarrollo de colecciones**, que puede utilizar algunos de los siguientes métodos:

- *Selección*, el proceso en el que los bibliotecarios de desarrollo de la colección—también conocidos como bibliotecarios temáticos o bibliógrafos—averigüen acerca de la existencia de obras a través de los catálogos de venta de los proveedores, las revisiones, los anuncios y los sitios Web de las editoriales, etc., de los que se eligen los materiales más apropiados para la colección.
- *Donaciones y Canje*, las donaciones abarcan recursos de información que son donados a la biblioteca por una parte interesada (por ej., un administrador, una organización local, o miembros retirados de la institución); el canje refiere a cuando una biblioteca tiene un acuerdo de intercambio mutuo con otra u otras bibliotecas para intercambiar sus recursos duplicados o no deseados por otros recursos buscados
- *Planes de aprobación (Approval plans)*, un método por el que la biblioteca contrata a uno o más proveedores para recibir nuevos recursos de acuerdo a perfiles pre-seleccionados representando las necesidades de la colección.
- *Adquisiciones conducidas por los usuarios (Patron-Driven Acquisitions)*, también conocida como *Adquisiciones conducidas por demanda (Demand-Driven Acquisitions)*; el proceso en el que una biblioteca obtiene ciertos recursos (usualmente, electrónicos), solo después que se han establecido definitivamente las necesidades por ellos (por ej., una biblioteca puede proveer acceso a un gran número de registros de e-book en el catálogo, pero la biblioteca solo adquiere un e-book individual después que un usuario ha accedido a él).

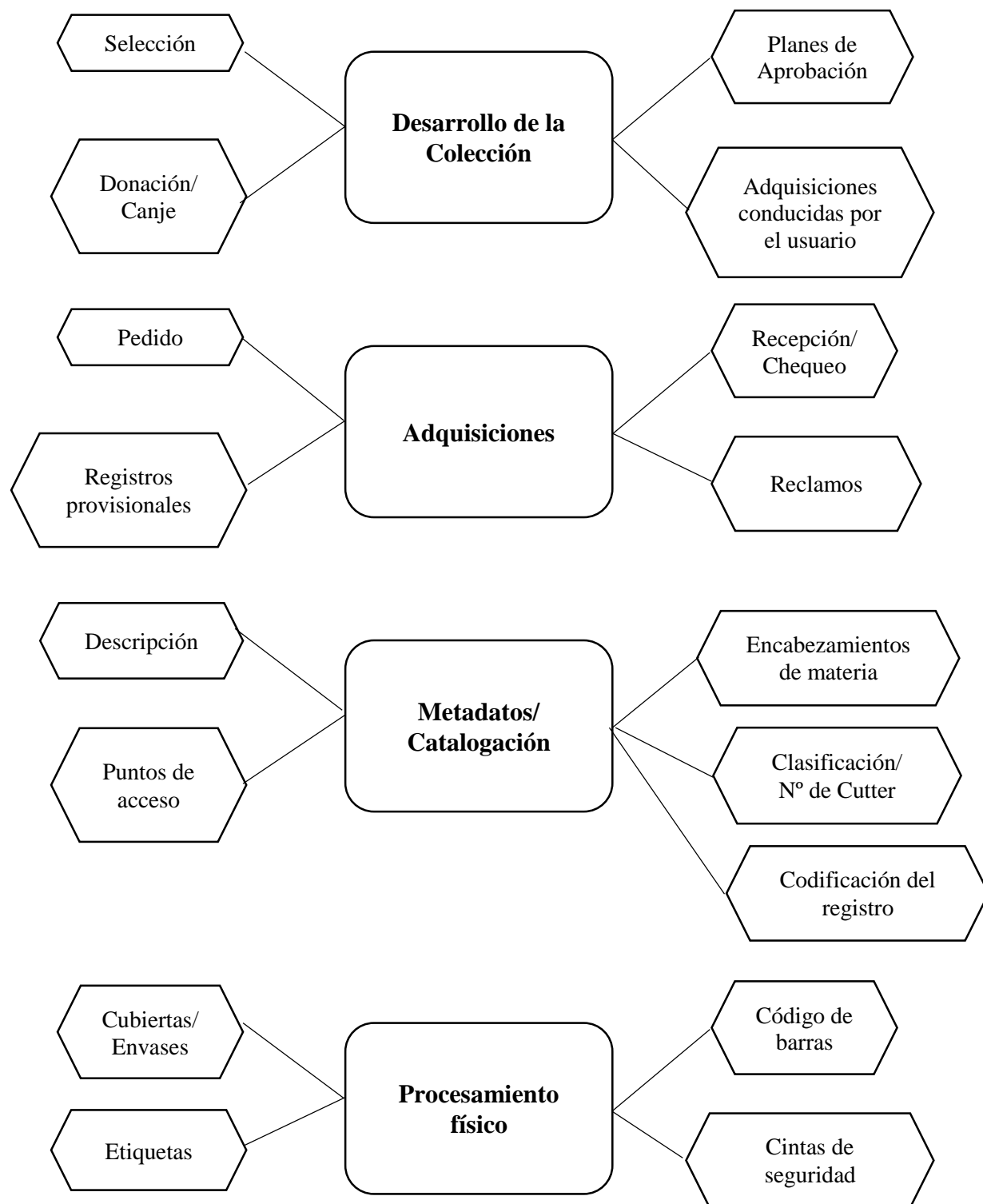


Figura 1.1 Diagrama de flujo de trabajo simplificado de los servicios técnicos de la biblioteca

Menos frecuente, pero aun practicado en algunas bibliotecas académicas y de investigación, los bibliotecarios encargados del desarrollo de la colección pueden realizar viajes de compra de recursos, para adquirir materiales especiales, particularmente materiales en lengua extranjera difíciles de encontrar.

Las **adquisiciones** son otra función relacionada con el desarrollo y la organización de la colección. La unidad de adquisiciones, entre otras cosas, es responsable de gestionar los pedidos y el presupuesto (por ej., solicitar los materiales, pagar las facturas, supervisar los pedidos de publicaciones en serie (standing orders), monitorear los presupuestos). Adicionalmente, puede ser responsable de recibir los recursos, chequearlos (check-in), reclamar ítems perdidos y gestionar devoluciones. Los bibliotecarios de adquisiciones pueden también ser responsables de obtener o crear registros de catálogo temporarios creados al momento de pedir los materiales., después de que los materiales han sido chequeados, estos registros provisionales son aumentados en el proceso subsiguiente.

Una vez que los recursos arriban a la biblioteca deben organizarse e integrarse dentro de la colección. Este proceso se denomina **catalogación**, y su objetivo es crear una lista multifacetada de los recursos a los cuales la biblioteca puede proveer acceso. Comprende dos actividades principales: la catalogación descriptiva y la catalogación temática (o clasificación). La catalogación descriptiva describe la composición de un recurso de información e identifica aquellas entidades responsables por su contenido artístico y/o intelectual, sin referencia a su clasificación temática o a la asignación de encabezamientos de materia, los que son dominio de la catalogación temática. La catalogación descriptiva involucra las siguientes actividades:

- **Creación de una descripción**—este proceso involucra el registro de los atributos más importantes de un recurso para permitir a los usuarios encontrar, identificar, seleccionar, obtener y explorar dicho recurso. Frecuentemente, la descripción es una representación directa de los “datos” del recurso que tratan de quien, que, donde y cuando (aunque quizás no tratan el porqué). Una descripción mínima incluye los siguientes atributos (si son aplicables): título del recurso; menciones de responsabilidad; información de la edición o versión; información de disseminación; tipo de recurso; extensión, dimensiones y tamaño; y los números normalizados asociados con el recurso.
- **Elección de los puntos de acceso**—esta actividad asegura que los nombres de la mayoría de las partes responsables, así como también los títulos varios asociados con el recurso, se enumeren en el metadato, y que dichos puntos de acceso puedan usarse para encontrar el recurso en el catálogo. Los puntos de acceso pueden incluir: nombre de los creadores (por ej., autores, artistas, entidades gubernamentales locales o nacionales, compiladores, compositores, familias, grupos musicales); nombres de contribuidores (por ej., editores, traductores, ilustradores, adaptador, intérprete, diseñador de sonido); los títulos brindados en el recurso; y, los títulos de recursos relacionados (por ej., películas, adaptaciones, títulos de series)
- **Asegurar el control de autoridad**—esta actividad provee consistencia en como aparecen los nombres y algunos títulos como puntos de acceso autorizados (i.e., secuencias de caracteres normalizadas elegidas para representar nombres o títulos). Por ejemplo, esto significa que los catalogadores usen consistentemente el punto de acceso autorizado Onassis, Jacqueline Kennedy, 1929-1994, en los metadatos para recursos por y acerca de Jackie Kennedy.

La descripción, la elección de puntos de acceso de nombres y de títulos, son llevados a cabo en la mayor parte de las bibliotecas de EE.UU. y de otros países a través del uso de RDA (Resource Description and Access; en español, Recursos, Descripción y Acceso). Este estándar de contenido reemplaza a las Reglas de Catalogación Angloamericanas, 2a

ed. (RCAA2) (Anglo American Cataloging Rules, 2nd ed., AACR2). RDA se basa en AACR2 y en los modelos conceptuales para datos bibliográficos y de autoridad (FRBR y FRAD), desarrollados por IFLA.

La catalogación temática, el segundo componente principal de la catalogación bibliotecaria, abarca dos actividades clave: el análisis conceptual y la traducción. El análisis conceptual es el proceso de determinar de lo que trata (aboutness) (i.e., el tema) del recurso. El catalogador analiza un recurso para determinar los conceptos, las ideas y las entidades discutidas por él, así como determinar su formato o género. Un recurso no puede ser precisamente descrito si no es examinado y comprendido cuidadosamente. La traducción es el proceso de transformar aquello de que trata un recurso a un lenguaje temático controlado. Incluye las siguientes actividades:

- **Elección de los términos del vocabulario controlado**—los catalogadores asignan términos del vocabulario controlado a partir de una lista para representar lo que trata un recurso, como parte de su metadato. El vocabulario controlado asegura consistencia en la representación temática y permite la reunión (collocation) (i.e., poner juntos materiales similares). Por ejemplo, a un recurso acerca de puentes y coronas dentales puede asignársele alguno de los siguientes conjuntos de encabezamientos de materia—cuales se apliquen dependen de la lista usada por la biblioteca:

Puentes (Odontología) y Coronas (Odontología), extraídos de Library of Congress Subject Headings (LSCHE)

- **Elección de las notaciones de clasificación**—este proceso involucra la representación de lo que trata un recurso a través de una notación normalizada. La clasificación asegura que los recursos similares se reúnan en el catálogo y en el estante. Por ejemplo, para los recursos indicados puede asignarse:

617.492 de CDD

RK666 de LCC

Cuando los materiales se organizan en colecciones, las entidades físicas tienen que organizarse de alguna forma. Estos pueden ser ubicados en estantes en el orden de llegada, o pueden ubicarse en un orden más significativo. Podrían ser ordenados alfabéticamente, la forma en que las secciones de biografía y de literatura de ficción se organizan en muchas bibliotecas escolares y públicas. Sin embargo, la mayoría de los recursos se organizan mediante la clasificación. Para los recursos tangibles, las notaciones de clasificación se utilizan para crear dispositivos de ubicación, denominados *signaturas topográficas* (call numbers). Esto se realiza para identificar el recurso físico y para proveer una ubicación única en el estante para él. Una *signatura topográfica* (sig. top.), cuenta generalmente con el agregado de un número de Cutter a la notación de clasificación (o *signatura de clase*) asignada al recurso. Un *número de Cutter* (o *signatura librística*) es una secuencia de caracteres alfanuméricos que usualmente representa el nombre del creador principal, o en ausencia de este, al título. Existen diferentes aproximaciones para establecer las sig. top., dependiendo del esquema de clasificación usado.

La mayoría de los registros creados por el proceso de catalogación se codifican con el formato MARC, de forma que puedan ser visualizados en un catálogo, el cual es parte

generalmente de una herramienta de automatización de la biblioteca, como el Sistema Integrado de Gestión Bibliotecaria (SIGB) o la Plataforma de Servicios Bibliotecarios (LSP).

El proceso de catalogación así descrito se denomina frecuentemente *catalogación original*. Afortunadamente, no es necesario para cada recurso de información en cualquier biblioteca ser catalogado originalmente en dicha biblioteca. Dado que las bibliotecas, frecuentemente, adquieren copias o ejemplares del mismo ítem, o deciden catalogar los mismos sitios Web, pueden compartir los metadatos adoptando una copia del registro original de catalogación creado por otra biblioteca para sus propios catálogos—este proceso se denomina catalogación por copia. Este tipo de catalogación se relaciona con la noción de catalogación cooperativa—el trabajo conjunto de instituciones independientes que permite compartir a los miembros de la red o crear la catalogación para que sea utilizada por otros. Existen varios modos de obtener copias de registros de catalogación existentes, como por ejemplo siendo miembro de un consorcio o red bibliográfica—por ej., descargando registros de OCLC—de proveedores—por ej. pagando por registros de catálogos que se envían junto con los recursos adquiridos—o de bibliotecas que ponen libremente disponibles sus metadatos—por ej., descargando un registro de catálogo creado por LC.

Una vez que los recursos han sido elegidos, recibidos y agregados al catálogo, los ítems físicos son enviados a la unidad de **procesamiento físico**, de manera que puedan prepararse para ser agregados a la colección. Esto implica agregar o remover la cubierta de los libros, colocar tiras de seguridad en los materiales, colocar las etiquetas de signatura topográfica (sig. top.) y los códigos de barra en los recursos, enviar un recurso al departamento de conservación/preservación si no está en buenas condiciones, entre otras cosas. Una vez que los recursos fueron procesados, pueden ser almacenados en los estantes para ser accedidos por los usuarios y puestos en circulación

Los principales resultados del proceso de catalogación son el ordenamiento de la colección y la creación y mantenimiento del catálogo que brinda el acceso a las colecciones. El catálogo puede mostrar lo que existe en la colección bajo determinados creadores, títulos particulares o sobre determinadas materias. También reúne juntas todas las obras de un autor y todas las ediciones de una obra, aun cuando es posible que estas no estén ubicadas juntas en las colecciones. Finalmente, el catálogo provee algún tipo de dispositivo de ubicación para indicar dónde puede encontrarse el ítem dentro de la colección. Antes que existieran los catálogos en línea, generalmente, el catálogo principal de la biblioteca, sea de fichas, en forma de libro o en microfilm, fuera complementado por otros catálogos. Los catálogos de bibliotecas departamentales, los registros de existencias de publicaciones seriadas, los catálogos de formatos especiales, y el catálogo topográfico (shelflist), contenían información de ubicación de copias específicas de un recurso hoy día, todos ellos fueron incorporados en una base de datos, en la mayoría de los catálogos en línea. Dado que la mayoría de los catálogos en línea forman parte de los SIGB, la información de circulación puede acompañar a cada uno de los registros del catálogo. Hasta hace poco tiempo atrás, el catálogo en línea solo siguió conteniendo registros de los ítems que estaban físicamente guardados por la biblioteca. A medida que las bibliotecas ingresan en entornos cooperativos, este principio de informar solo lo que la biblioteca tiene se ha ido perdiendo. En las uniones de catálogos que contienen los registros de más de una institución, el concepto se expande a lo que al menos tiene una de las bibliotecas cooperantes. Más recientemente, la adición de registros de recursos digitales encontrados en Internet significa que varios catálogos cuentan con registros de aquellos recursos a los que la biblioteca puede brindar acceso, abarcando también los recursos de información con los que cuenta la biblioteca. Los portales accesibles en los

catálogos en línea proveen a los usuarios de una forma de localizar toda la información para autorizar su acceso. El portal presenta una pantalla de autenticación para los usuarios; si su nombre de usuario y clave son aceptados, puede contar con acceso a cualquier recurso que se permita para su estatus. Una universidad, por ejemplo, puede tener licencias para que sus usuarios accedan a distintas bases de datos de texto completo en línea. Anteriormente, el usuario tenía que utilizar el protocolo de acceso para cada base de datos contando con un ID y una clave particular. Un portal de una biblioteca puede contar con vínculos de acceso a recursos locales—estos incluyen libros, revistas, bases de datos, colecciones digitales, cursos, todos ellos accesibles mediante la búsqueda en el catálogo—recursos remotos, ayudas de referencia—incluye herramientas de referencia en línea, vínculos a motores de búsqueda en línea y el escritorio de referencia virtual que permite el contacto en tiempo real o por e-mail con el bibliotecario de referencia—e información personal del usuario—incluye listas de materiales prestados, búsquedas grabadas y alertas personalizadas.

Los catálogos en línea pueden ser pasarelas (gateways) a recursos externos. Una mejora principal en los catálogos en línea fue el agregado de la posibilidad de acceder directamente a recursos de acceso remoto en Internet. Actualmente, una URL en un registro de catálogo puede proveer acceso inmediato a metadatos adicionales (por ej., la tabla de contenido), a contenido complementario (por ej., un suplemento, una imagen de la cubierta del libro) o al texto completo del recurso que se representa en la descripción, si está disponible (por ej., un vínculo a la versión pdf del recurso). Los catálogos pueden también conectarse con los catálogos de otras instituciones o aún a redes bibliográficas (por ej., OCLC) que pueden mostrar donde encontrar un recurso si este no está en el catálogo local. El recurso puede ser solicitado por Préstamos InterBibliotecarios (PIB) o alguna otra forma de distribución de documentos.

Además de la búsqueda en el inventario de los recursos tangibles encontrados en el catálogo, las bases de datos bibliográficas y de texto completo pueden ser accesibles a través de la capa o del estrato de descubrimiento (discovery layer) del catálogo (también denominado interfaz de descubrimiento, herramienta de descubrimiento o servicios de descubrimiento). Una capa o estrato de descubrimiento es un agregado tecnológico al catálogo. Permite a los usuarios tener una gran variedad de interacciones con los “almacenes de información” de la biblioteca. Por ejemplo, a través del estrato de descubrimiento, los usuarios pueden buscar libros, artículos de revistas y objetos digitales, todos en una sola búsqueda, o pueden realizar browsing en los resultados de búsqueda por palabras clave que han clasificado en facetas (i.e., aspectos fundamentales, características de los recursos), como género, tipo de recurso, creador, materias, fecha de publicación, ubicación, etc.

Una influencia sobre el proceso de organización en las bibliotecas se encuentra en el proceso de referencia. Las bibliotecas están organizadas de manera que la información pueda ser recuperada. En el proceso de referencia, se prueba el éxito de la organización. Si se encuentran dificultades en su uso, algunos de los procesos de la organización deben ser revisados. Los servicios administrativos de las bibliotecas también deben estar interesados con la organización de la información. Los administradores son responsables por las decisiones tecnológicas que afectan directamente la organización de la información registrada en dicho ámbito. Las decisiones de los administradores también afectan el futuro, en el que el caos reinará si las actividades y los procesos de organización de la información no son soportados.

Los archivos

A pesar de que a lo largo del siglo XX las bibliotecas se han normalizado crecientemente y que los recursos que colecciona son duplicados de los guardados por otras, no sucede lo mismo con los archivos. Los archivos generalmente coleccionan ítems únicos, por lo que se pensaba en la innecesidad de la normalización. Los archivos no podían aprovechar los metadatos previamente (i.e., usar metadatos creados por otras agencias) porque no describían materiales que también fueran poseídos en otro lugar. A pesar de esta situación, los archivos han experimentado intentos significativos en pos de la normalización. Los archivos preservan registros de valor duradero que documentan las actividades personales o de una organización, acumuladas durante el transcurso de su vida diaria y/o laboral. Los registros personales consisten de correspondencia, manuscritos, papeles personales, etc., o pueden consistir de una colección de recuerdos o de un álbum de recortes.

Los materiales de archivos pueden estar en diferentes formatos: papel, imágenes, grabaciones sonoras, grabaciones de imágenes en movimiento y formatos digitales. Los materiales de archivos se organizan desde hace siglos. A diferencia de los materiales coleccionados por la biblioteca, los materiales de archivo son coleccionados y descritos en conjunto. Hasta hace décadas atrás, cada archivo elegía una forma propia de organizar la información, especialmente respecto al nivel de control y de profundidad de la descripción. Los principios prevalecientes en la organización de la información de los archivos han sido la *procedencia* y el *orden original*.

La procedencia o *respect des fonds*, refiere al individuo, familia organización o institución que es responsable por la creación, mantenimiento o uso de los materiales. El principio dicta que los registros de diferentes orígenes deberían ser mantenidos separados para preservar su contexto. El orden original dirige a los archivistas a mantener la secuencia u organización de los materiales establecidas por el creador de tales materiales. Este puede o no ser evidente en la colección. Si el orden original se ha perdido, o la colección nunca fue verdaderamente organizada por el creador o coleccionador, entonces un archivista puede construir un orden lógico para la colección. Al establecer un orden lógico, sea en el orden original o en uno establecido por el archivista, los subagrupamientos (series) entre los materiales pueden llegar a identificarse. La descripción de la colección debería reflejar el orden de dichos agrupamientos [...].

La normalización y la cooperación llegan a los archivos a comienzos de la década de 1980, debido al aumento del interés de investigadores que requieren documentos y colecciones de archivos almacenados en distintas partes del mundo. En la década de 1990 y del 2000, crece el interés en la comunidad académica por ingresar las descripciones de las colecciones de archivo en las mismas bases de datos con los registros bibliográficos generados por las bibliotecas. Las descripciones de los materiales de archivo pueden adoptar una o más formas diferentes. Un registro de adquisiciones resume la información sobre la fuente de la colección, proporciona las circunstancias en las que fue adquirida (que se detalla más aún en el archivo de donaciones) y realiza una breve descripción de los datos físicos y de los contenidos de una colección. Los instrumentos de localización (finding aid) proporcionan una nota de contenido detallada sobre el contexto histórico y de organización de la colección, describe su contexto, proporciona un inventario que brinda una idea cabal de que contiene cada caja. Puede contener encabezamientos de materia, puntos de acceso controlados y autorizados y detalles físicos como si presenta materiales frágiles. Los instrumentos de localización son denominadas, a veces, registros (registers), inventarios o listas de contenedores (container lists). Un registro separado que

describe la colección o una parte de ella puede ser creado para incluirse en el catálogo de la biblioteca.

Para satisfacer la responsabilidad de cuidar los registros de valor permanente y al mismo tiempo provee acceso a dichos materiales. Los materiales de archivo se encuentran en estanterías cerradas, accesibles solo al personal. El público no puede hojear los documentos de modo que el ordenamiento del material no requiere ser clasificado—como ocurre en la biblioteca con estanterías abiertas. En todo caso, cualquier clasificación sería inútil dada su amplitud y la naturaleza variada de la colección.

En la década de 1980, los archivos se interesaron con la posibilidad de disponer que sus registros bibliográficos estén disponibles en bases de datos. En tal sentido, se desarrolló el formato MARC-AMC (Archival Manuscript Collections). A pesar de algunos problemas que aún persisten, se sigue utilizando este formato para codificar los registros de archivo (actualmente la denominación es *Mixed materials* en lugar de AMC). A fines de los 1990, el estándar EAD fue publicado e implementado internacionalmente para codificar instrumentos de localización de manera que puedan visualizarse y manipularse en línea.

La organización de materiales de archivos es una actividad fundamental en la gestión y provisión de acceso a las colecciones. Las descripciones archivísticas sirven a varios propósitos. Provee pautas a los investigadores sobre la utilidad del material para sus necesidades de información. También asiste a los archivistas en la gestión y preservación de sus colecciones. Cuando las colecciones están bien organizadas y documentadas, los materiales de archivo pueden satisfacer una amplia variedad de usos y necesidades de los usuarios.

Los museos (Colecciones de objetos y arte)

A pesar de que las bibliotecas y los archivos contienen algún tipo de material visual, las colecciones de museos, de galerías de arte y otras instituciones que coleccionan artefactos y objetos de la cultura material, consisten principalmente de material visual bidimensional o tridimensional. Estas colecciones han sido organizadas tradicionalmente para uso interno exclusivamente. No obstante, recientemente, las necesidades de investigación, los mandatos institucionales y el interés del público han dirigido a estas instituciones a pensar acerca de la organización de sus colecciones en los modos en que lo han realizado las bibliotecas y los archivos a lo largo de los años. Los esquemas de metadatos—como *Categories for the Description of Works of Art (CDWA)*—los vocabularios controlados—como *LC Thesaurus for Graphic Materials (TGM)*, los vocabularios del *Getty Institute* e *Inonoclass* (un sistema de clasificación para describir el contenido iconográfico y narrativo de las obras visuales—y estándares de contenido—como *Cataloging Cultural Objects (CCO)*—emergieron en las últimas décadas como respuesta a las necesidades de los museos para organizar y diseminar mejor la información de sus colecciones.

Las obras de arte o los objetos de museos y galerías se obtienen a través de un proceso de adquisiciones formal de la institución. Al igual que en los archivos, se crean registros de las adquisiciones. En los museos de historia natural esta práctica difiere, en tanto los objetos son adquiridos mayoritariamente a través del trabajo de campo, confeccionando un registro preliminar en el mismo campo. Al decidir guardar los objetos en la colección, se crean registros de adquisiciones. En algunos casos, grupos de objetos de naturaleza similar son descritos en un solo lote, al que se le brinda un único número de adquisición. La curación de objetos individuales resulta en registros de catálogo a nivel departamental con secuencias numéricas propias.

En aquellos museos que no se corresponden con la historia natural, los ítems se registran una vez que son ingresados. El registro se parece a la catalogación de las bibliotecas y de los archivos. El registro sirve como un catálogo, en tanto establece un control y organización sobre las obras de arte y los objetos. En los museos la procedencia constituye una información de importancia y esencial para determinar el nombre del objeto y otros elementos que describen el objeto particular. La procedencia tiene algunas leves diferencias en los museos y en los archivos. La procedencia en los archivos dirige los principios de la organización. Las colecciones se organizan de acuerdo a su procedencia. En la historia del arte, la procedencia de una obra se conecta con la posesión y la autenticidad. Tanto la procedencia como la condición física del objeto debe aparecer con toda la demás información acerca del objeto en el registro. Un aspecto de la creación de registros para arte y objetos de museos que es muy diferente de la creación de registros para recursos bibliográficos, es que los objetos son frecuentemente conocidos imperfectamente al momento de la adquisición o del registro. Puede existir una acumulación de información a lo largo del tiempo, alguna de ella puede ser conflictiva o contradictoria, al que se vuelve parte de la documentación.

La descripción del material visual es generalmente más difícil que la descripción del material textual, ya que depende en mayor medida de las percepciones de la persona que realiza la tarea. A menudo no se presentan o no existen palabras asociadas con los ítems descritos, siendo necesario que los que describen tales recursos utilicen sus propias palabras. Un registro individual comporta más campos que el típico registro de catálogo de una biblioteca. Algunos de los campos—no utilizados por la catalogación de bibliotecas—necesarios para la descripción de obras de arte son: material, técnicas, estudio de origen, tipo de equipo utilizado, colores, textura, simbolismo de diseño, procedencia, historial de exposición, consideraciones acerca de la instalación, valor estimado, etc. Aún con la presencia de estos campos adicionales, no es del todo posible anticipar todos los usos que un investigador podría hallar en una obra de arte o en objetos artísticos. Por ejemplo, una pintura que represente una escena callejera del siglo XIX puede revestir utilidad para historiadores, arquitectos, urbanistas, investigadores, historiadores de la cultura, médicos, sociólogos, etc. Se están desarrollando sistemas que comienzan con consultas que utilizan el texto de la descripción. Luego, los resultados de las consultas permiten que el usuario explore imágenes sustitutas.

El análisis temático de los materiales visuales también reviste dificultades, dado que una imagen no expresa mediante palabras escritas de que trata ni el título de la obra visual siempre informa claramente de que trata la obra. La dificultad mayor con este tipo de materiales estriba en trazar una línea divisoria entre descripción y análisis temático. En tal sentido, una obra pictórica puede ser descrita—por ejemplo—como una mujer vestida de azul mirando al niño que tiene en brazos. No obstante, al enunciar que el tema de la obra es “María junto con Jesús” se cruza la línea de la descripción hacia la interpretación—con la excepción de que este constituya el título de la obra brindado por el artista.

Una barrera percibida para compartir metadatos entre museos reside en la noción extendida de que cada uno de ellos poseen objetos únicos. Esta afirmación no reviste la misma consideración si se trata de colecciones de historia natural como en el caso de colecciones de obras de arte. Esto implica que si bien cada espécimen de insecto o de pájaro es único, cada uno de ellos representa una clase de organismos que puede identificarse a nivel de género y de especie. Si bien tendrían que presentarse notas específicas para cada uno de ellos, no se excluye absolutamente la idea de reutilizar metadatos para el acceso compartido. Hasta el advenimiento reciente de iniciativas como AMICO (Art Museum Image Consortium <http://www.amico.org>), ARTstor

<<http://www.artstor.org/index.shtml>> y WorldCat de OCLC <http://www.oclc.org/worldcat/>, los museos han rechazado abandonar su terminología y sus modos y prácticas de organización de la información para participar de uniones de catálogos o herramientas de recuperación federadas. Esta situación está cambiando. El capítulo 7 describe algunos proyectos cooperativos de museos que están actualmente poniéndose en marcha.

Además de sus colecciones principales, el museo o la galería de arte pueden contar también con un archivo, con una biblioteca o administrar registros. Una biblioteca de museo posiblemente contenga materiales publicados que documenten o se relacionen con las colecciones de dicho museo. Los archivos de museos documentan los registros institucionales del museo en la medida que se acumulan [...].

Similar a lo que ocurre en los archivos, las colecciones de los museos o de las galerías de arte son solo accesibles para el personal. Gran parte de ellas se almacenan sin acceso al público, mientras que solo una porción de ellas se encuentra en exposición en un momento determinado. Las colecciones que se almacenan son numeradas de manera tal que pueden recuperarse rápidamente cuando se las requiera. Las personas que son responsables de las exhibiciones del museo hacen frecuente uso del sistema de control de la organización. Estas colecciones se usan crecientemente para investigación por personas con diferentes necesidades de investigación, aunque el acceso mediato no requiere cambios en la estrategia de control organizacional.

Ambiente en Línea

En el siglo XXI, Internet se ha vuelto omnipresente en la vida personal y profesional. Internet—particularmente, la Web—es una herramienta que se utiliza para cualquier cosa. La usamos para llevar a cabo transacciones comerciales, para gestionar nuestras cuentas bancarias, para encontrar respuestas a nuestras preguntas y para investigar cualquier cosa sobre la que sentimos curiosidad. Sin embargo, Internet no es una panacea para todos nuestros problemas informativos. Internet es impredecible e inconstante; puede ser desagradable y falsa; esta llena de opiniones que pueden ser enmascaradas como hechos; y esta muy desorganizada. Internet ha sido comparada con una biblioteca en la que todos los libros han sido tirados en el piso y no cuenta con ningún catálogo. Por varios años, se han realizado esfuerzos para encontrar modos de lograr cierto control sobre ella. Sin embargo, no puede decirse que Internet este organizada y esto no es posible de cambiar de manera inmediata. Porciones de ella—ciertos tipos de espacios en línea y muchos sitios individuales pueden ser organizados de alguna manera, pero estos no son representativos de Internet en su totalidad. Las siguientes cuatro subsecciones discuten algunos de estos ambientes en línea: Internet en general, la Web Semántica, las colecciones digitales y la arquitectura de la información.

Internet

En la mitad de la década de 1990, algunos bibliotecarios e investigadores en Bibliotecología y Ciencia de la Información, intentaron diseñar proyectos para organizar Internet estos proyectos comenzaron a ser convincentes solo después que la Web de transformó en la cara de Internet, sin embargo, estos profesionales fueron rápidos al comprender que muchos esfuerzos estaban desactualizados en un corto tiempo. Inicialmente, las bibliotecas intentaron utilizar los medios tradicionales de organización, aunque se volvió claro muy rápidamente que la catalogación bibliotecaria a escala

completa no era particularmente apropiada para describir los recursos electrónicos efímeros, que cambian frecuentemente.

En lugar de crear registros individuales para cada sitio Web. Algunos bibliotecarios comenzaron a compilar bibliografías de sitios, algunos de los cuales se volvieron eventualmente pasarelas (gateways) o directorios de Internet más formales. Un *directorio de Internet* es una colección organizada de vínculos a sitios Web sobre tópicos particulares. En muchos directorios, los indizadores humanos son curadores de los vínculos al comienzo; aunque los desarrollos tecnológicos comenzaron a coleccionar vínculos a través de medios automáticos. INFOMINE establecido en 1994, es un ejemplo de directorio de Internet creado por bibliotecarios cooperantes de varias instituciones académicas. Contiene páginas dedicadas a varios tópicos (por ej., recursos culturales, información gubernamental, mapas y datos geográficos), cada uno de los cuales contiene listas de vínculos a sitios Web recomendados. Los vínculos podrían facilitar el browsing en listas alfabéticas o ser buscados por palabras del título, nombres de creadores, materias o por palabras clave. Su organización fue similar, aunque más simple, al directorio más conocido como fue el de Yahoo!, que también data de 1994. Yahoo! escogió un método de browsing con menú desplegable (drill-down) más jerárquico, donde uno comienza con una de las 14 categorías superiores, que se dividen en un gran número de facetas o de subtópicos. En cualquiera de las categorías, uno puede encontrar fácilmente 5 o 6 niveles de subtópicos y, en algunos casos, varios más. Sin embargo, los directorios de sitios Web estructurados jerárquicamente, dejan de ser preferidos frente a los motores de búsqueda, como Alta Vista y Google, después de que muchas mejoras incrementaron la eficacia de los motores de búsqueda. Después de dos décadas de operación, Yahoo! e INFOMINE cerraron sus directorios en Diciembre de 2014, luego de experimentar un significativo declive en su uso en los últimos años. Un ejemplo de un directorio de Internet duradero, al menos al momento de este escrito, es SciCentral, que provee vínculos actualizados sobre noticias, revistas, conferencias, recursos de trabajo, etc., de naturaleza científica. Aunque aún existe un pequeño número de pasarelas, los directorios jerárquicos son mayoritariamente, una reliquia del pasado de Internet. Como sostienen varios arquitectos de la información, las técnicas relacionadas con encontrar (findability) que fueron efectivas a fines de la década de 1990 (por ej., el directorio jerárquico curado por Yahoo!) son ineficientes hoy día.

Otro intento para controlar la información de Internet resultó en el desarrollo de uno de los esquemas de metadatos más ampliamente conocido en el mundo. Los bibliotecarios formaron parte del grupo de personas que desarrolló el estándar de metadatos para describir recursos en línea, denominado Dublin Core (DC). Fue diseñado como un conjunto pequeño de elementos básicos de metadatos, dado que la mayoría de los sitios Web y otros documentos como objetos (document-like objects) en Internet no requieren tanto detalle como el que generalmente se encuentra en el catálogo de la biblioteca. En la década de 1990, OCLC estableció una forma para que las bibliotecas cataloguen recursos en línea de manera cooperativa y contar con acceso a una base de datos de metadatos que describan recursos Web importantes. Este sistema llamado *Connexion* está disponible en la interfaz Web de OCLC. En Connexion, la catalogación puede realizarse tradicionalmente con AACR2 y MARC21 o en DC.

De hecho, mucha de la organización de Internet ha sido desarrollada por personas que no son bibliotecarios. Los motores de búsqueda—por ejemplo—han sido desarrollados por especialistas en computación y en programación. La mayoría de las personas aprecian los motores de búsqueda aunque comporten cierta frustración con el hecho de que no sean más selectivos y precisos en sus resultados. La mayoría de los programas—por ejemplo, robots, arañas—utilizados para buscar sitios Web y agregarlos a los índices de los motores

de búsqueda, solo reconocen texto; los videos, los archivos de sonido y las imágenes pueden ser reconocidas como tales, aunque no pueden ser indizados a menos que cuenten con etiquetas o texto que lo acompañe. Asimismo, dichos programas no pueden analizar el propósito, la historia, las políticas, etc., de un sitio Web. Para mejorar esta situación, se están llevando a cabo modificaciones sobre diversos tipos de metadatos. En tal sentido, los robots están recopilando información apropiada a partir de metadatos que se han agregado a un sitio Web por su autor original o por alguien capacitado en descripción y análisis de recursos de información.

En años recientes, la actitud de incluir metadatos en los recursos Web ha cambiado. Algunos de los motores de búsqueda más peculiares del mundo (i.e., Google, Bing, Yahoo! y Yandex) comenzaron a abrazar ciertas tecnologías semánticas que incrustan pequeñas cantidades de metadatos dentro del código HTML, para hacer que los datos se encuentren dentro de los sitios Web que son más significativos para los motores de búsqueda y los robots de la Web (véase la discusión sobre microdatos debajo). Usados propiamente, los metadatos pueden incluir información acerca de los creadores, títulos, materias y otros aspectos tradicionales de la descripción; las partes no textuales de un sitio; porciones textuales de un sitio; entidades, roles y relaciones; y, el propósito y la historia del sitio.

Este deseo inicial entre los bibliotecarios y los profesionales de la información de describir cada pieza de Internet, se disipó rápidamente. No es realista pensar que cada recurso Web pueda ser catalogado como un libro o un DVD en una biblioteca. En su lugar, los profesionales y el público en general abrazaron los motores de búsqueda para recuperar materiales en Internet. Aunque sus resultados son algunas veces solapados, frecuentemente llenos de falsas asociaciones y generalmente carente de control semántico (por ej., no son tratados los homógrafos y los sinónimos), los motores de búsqueda son excelentes para satisfacer las necesidades de búsqueda más generales, los profesionales de la información, los investigadores y otras personas que realizan búsquedas de forma consciente, no llevan a cabo investigaciones académicas utilizando solo motores de búsqueda, sino que suelen comenzar la exploración preliminar sobre un tópico de este modo. Los motores de búsqueda, al igual que otras herramientas de recuperación más estrechamente focalizadas, tienen un importante rol que jugar en la recuperación de la información y se anticipa que su alcance, poder y efectividad crecerá en el futuro en tanto más metadatos significativos se agreguen a los contenidos Web.

A pesar de la creencia de la imposibilidad de organizar Internet, si serán organizadas aquellas partes que revisten importancia respecto a la recuperación de la información y la preservación para la posteridad. Los principios aprendidos a lo largo de siglos para organizar información impresa—propios de la naturaleza humana—pueden ser utilizados para acelerar el proceso de organización de los recursos electrónicos. Los esfuerzos actuales dirigidos a crear la Web Semántica, donde los datos Web serán definidos semánticamente y vinculados con otros datos relevantes con el propósito de un descubrimiento más efectivo de la información, es un buen ejemplo.

Web Semántica y Datos Vinculados

En el ambiente tecnológico y de rápidos avances del tercer milenio, mucho de nuestras vidas se focaliza en Internet. Un gran volumen de la información mundial ha sido transferida a la Web, como también muchos de sus metadatos. Muchos de nuestros métodos de comunicación y procesos de trabajo dependen de estas tecnologías. Internet es indispensable en nuestra vida durante el siglo XXI, aunque esta es imperfecta. Las computadoras no son inteligentes y carecen de intuición. No son buenas en inferir

condiciones o en comprender el contexto sin surtir una cantidad significativa de información adicional. Consecuentemente, algunas veces nos sentimos frustrados con nuestra tecnología. En este momento, la WWW y sus motores de búsqueda carecen de control semántico. Cuando buscamos en Google información acerca de arañas en la Web—los programas que recorren Internet buscando nuevos sitios Web—nos vemos inundados también de páginas y de imágenes sobre telarañas creadas por los arácnidos para atrapar sus presas. Los sinónimos y homógrafos constituyen un obstáculo al realizar búsquedas en los actuales sistemas de correspondencia de palabras clave, causando falsas asociaciones. En su mayoría, los usuarios están acostumbrados a esto; aceptan esta situación como parte del proceso de búsqueda, aunque ninguno sienta que esto este bien.

En 1999, Tim Berners-Lee describió su sueño de que algún día las computadoras fueran capaces de analizar todos los datos en la Web (incluyendo las transacciones entre las computadoras y las personas), y cuando esto fuera posible, las comunicaciones entre computadoras traerían la llamada Web Semántica. La *Web Semántica* es una extensión propuesta de la WWW tradicional. Diferentes autores han expresado sus objetivos en distintos modos:

- La Web Semántica llevará estructura al contenido significativo de las páginas Web, creando un ambiente donde los agentes software deambulan de página en página pudiendo llevar a cabo tareas sofisticadas para los usuarios. Guiará hacia nuevas funcionalidades significativas en la medida que las computadoras se vuelven mucho más capaces de procesar y de “comprender” los datos que estas, en la actualidad, solo se limitan a mostrar
- La visión general de la Web Semántica puede resumirse con la frase: hacer mas accesible la Web a las computadoras
- El objetivo de la Web Semántica es moverse desde una Web de Documentos a una Web de Datos Abiertos, interconectada

Como indican estos párrafos, la Web actual carece de estructura, semántica (i.e., significado) y foco. Esto significa que la Web Semántica debe cambiar de la identificación y la recuperación de documentos para el consumo humano hacia la identificación y la recuperación de datos en aquellos documentos para que los humanos y las computadoras comprendan, manipulen y procesen de mejor manera. De varias formas, esto constituye un intento de transformar la Web en algo similar a una base de datos gigante de carácter global. Así, si regresamos a la analogía de Internet como una biblioteca donde todos los libros están tirados en el piso, la Web Semántica es una biblioteca donde todos los libros han sido despedazados en páginas, párrafos o aún oraciones individuales (i.e., piezas de datos), de forma que las piezas de datos puedan ser claramente identificadas y las relaciones entre ellas puedan ser explícitamente indicadas para que las computadoras se conecten e un nivel de mayor granularidad. La noción de Web Semántica no se basa sobre un documento como unidad; se dirige más a lo que está contenido dentro.

Existen algunas diferencias importantes entre la Web Semántica y la Web tradicional basada en documentos. En primer lugar, en la Web basada en documentos, generalmente, pensamos en páginas Web en tanto recurso—un documento codificado en HTML que contiene información, siendo una transacción estática y unidireccional. A lo largo del tiempo, este modelo ha evolucionado. Los recursos Web son actualmente algo mas que solo sitios o documentos, incluyendo canales de noticias, servicios comerciales (por ej., Amazon.com), herramientas interactivas (por ej., Facebook, Twitter) y demás. Sin embargo, en la Web Semántica, el término *recurso* esta definido aún de manera más

amplia; puede referir a sitios Web, bases de datos en línea, etc., sino que recursos puede referir también a objetos de la vida real (por ej., un libro físico, una pintura, un cáliz) así como personas, lugares, ideas y demás. En otras palabras, cualquier cosa, sea electrónica o no, puede representarse en la Web Semántica siempre y cuando exista un Identificador Único del Recurso (URI) o un Identificador Internacional del Recurso (IRI), que puede conectarse sin ambigüedad a él. Un URI es una secuencia compacta de caracteres que identifica un recurso físico o abstracto. En otras palabras, es una cadena de caracteres que ayuda a identificar cosas. La URL, que se ve en la barra de localización de un navegador Web, es un tipo de URI. Una IRI es similar a una URI, aunque trabaja con un conjunto de caracteres expandidos internacional.

En la Web tradicional, basada en documentos, las páginas constituyen contenedores de información. Esta información es marcada con HTML para subrayar algunas características principales (por ej., título, encabezamientos, párrafos) y para denotar como el contenido debería ser visualizado, aunque presenta una estructura semántica muy leve. Existen pocas indicaciones de que tipos de datos se encuentran en el recurso, quien es el responsable, de que trata, o que existen cientos de recursos que tratan sobre el mismo tópico (o que están conectados con él de alguna forma). La Web Semántica, por otro lado, espera suministrar mejores estructuras semánticas en los datos para mejorar la comprensión e incrementar las conexiones. Esto se logra por medio de la creación de metadatos adicionales que describan, de manera más significativa, la información encontrada en los recursos Web. Esto significa que los documentos necesitan ser descritos usando metadatos de forma estructurada, de manera que las computadoras puedan procesar estos datos y brindar una mejor “comprensión” del contenido. Por ejemplo, usando un formato altamente estructurado como Resource Description Framework (RDF), un recurso Web encontrado en <http://www.loc.gov/aba/cataloging/tools> puede ser codificado para expresar: Library of Congress (LC) es el creador, el título es “Resources for Cataloging” y el documento trata de *catalogación*. En RDF, se usarían IRI para comunicar esta información. Esos enunciados deberían verse como lo que se visualiza en la Figura 1.2. (Existen varios métodos de comunicación de enunciados RDF, siendo este el más simple). RDF es una estructura para comunicar metadatos en la Web. Se describe mejor debajo (y se explica con más detalle en el Capítulo 5).

Los documentos completos pueden describirse como se ve en la Figura 1.2, pero pueden describirse piezas individuales del contenido dentro de los documentos a través del uso de atributos adicionales. Para lograr una mejor comprensión de las partes individuales del contenido, los metadatos (una especificación HTML usada para anidar los metadatos dentro del contenido de los recursos Web) pueden incrustarse en una página Web. Por ejemplo, usando la semántica definida por Schema.org (un vocabulario diseñado para el marcado de contenido electrónico), los siguientes enunciados pueden adicionarse al código HTML de la página de Library of Congress.

Esto establece que el recurso Web es para la agencia gubernamental, cuyo nombre es Library of Congress, también conocida por su nombre de usuario en Twitter @librarycongress, y los dos nombres son aproximadamente equivalentes.

El uso de metadatos estructurales RDF o de microdatos incrustados, puede agregar cualquiera o toda la información siguiente para mejorar la comprensión del contenido de parte de la computadora:

- LC es el creador de la página
- LC es una biblioteca
- LC es parte de la rama legislativa del Gobierno de EE.UU.

- www.loc.gov es la dirección Web de un documento que es creado por y trata sobre LC
- www.loc.gov puede ser usado para representar la entidad real llamada Library of Congress
- LC tiene un número telefónico de contacto (202) 707-5000
- LC está en Washington, DC (una ubicación geográfica, la que podría ser descrita exhaustivamente)
- LC está ubicada en 101 Independence Ave., SE, Washington, DC 20540

En la Web de documentos actual, la codificación HTML provee menos instrucciones, como sigue:

- Presenta el documento usando HTML
- Identifica una cadena de texto particular como el título del documento
- Presenta otras cadenas como encabezamientos
- Formatea la mayoría del contenido en párrafos
- Usa una hoja de estilo particular para los detalles acerca de como se presentan las cosas
- Provee un vínculo al catálogo de la biblioteca, donde dice “*Search and browse records*”

Mientras que la mayoría de los sitios Web tienen algún grado de estructura, el lenguaje en el que son creados, HTML, se orienta hacia documentos textuales estructurados, antes que hacia los datos. En la medida que los datos se entremezclan en el texto circundante, es difícil para los softwares de aplicación extraer fragmentos de datos estructurados de páginas HTML.

Durante muchos años, las páginas Web tradicionales proveen un número limitado de vínculos escogidos a mano hacia otras páginas Web. Los creadores eligen puntear a otros documentos que consideran relevantes por la información que presentan—generalmente, al nivel de recurso completo o de una parte separada de él. En la medida que la Web se desarrolló, fueron diseñadas tecnologías que permiten vínculos para ser generados automáticamente. Un objetivo de la Web Semántica es aumentar las habilidades tecnológicas para conectar una amplia variedad de recursos relacionados a través de medios automáticos. La noción de datos vinculados es la base de la Web Semántica. Los datos vinculados (linked data) constituyen una aproximación para codificar datos de un amplio rango de fuentes diferentes, y publicarlos en la Web de forma tal que pueda ser comprendida por las computadoras relacionadas entre si al mismo recurso o concepto. Los datos vinculados:

- Son legibles por computadora
- Tienen un significado explícitamente definido
- Se vinculan a otros conjuntos de datos externos
- Se vinculan desde conjuntos de datos externos

Por ejemplo, si dos recursos diferentes contienen enunciados de metadatos donde cada uno de dichos recursos trata acerca de la catalogación (con cada uno de los recursos punteados a un IRI para una descripción autorizada de dicho concepto), entonces la computadora podría “comprender” que tales recursos tratan sobre catalogación y están, además, relacionados a cada uno de ellos. Las computadoras pueden explorar las conexiones e inferir que si uno de tales recursos fueran de interés, el otro debería ser de interés también para los usuarios.

```
01 <https://www.loc.gov/aba/cataloging/tools/> <http://purl.org/dc/terms/creator> <http://id.loc.gov/authorities/names/n78089035> .  
02 <https://www.loc.gov/aba/cataloging/tools/> <http://purl.org/dc/terms/title> "Resources for Cataloging" .  
03 <https://www.loc.gov/aba/cataloging/tools/> <http://purl.org/dc/terms/subject/> <http://id.loc.gov/authorities/subjects/sh85020816> .
```

Fig. 1.2 Un ejemplo de RDF


```

<div itemscope itemtype="http://schema.org/Government
Organization">
  <h1 itemprop="name">Library of Congress</h1>
  <span itemprop="alternateName">@librarycongress</span>
</div>

```

Con algún esfuerzo intelectual, las relaciones frecuentemente tácitas entre los recursos representados en la Web (por ej., creadores, materias, ubicación) pueden ser transformadas en más obvias, más explícitas y más fáciles de ser procesables por las computadoras a través del uso de los datos vinculados. Los datos vinculados usan RDF como su fundamento para asegurar claridad en el significado. RDF es una estructura de oraciones para enunciados de metadatos, conocidas como tripleta (triplet), en tanto contiene tres partes: un tema, un predicado y un objeto. Por ejemplo, un recurso puede contener un equivalente codificado para este enunciado:

Esta página Web	trata acerca de	quesos
Sujeto	Predicado	Objeto

Por supuesto, cuando ha sido codificado por computadoras antes que por la comprensión humana, ya no es más en lenguaje natural. En su lugar, utiliza IRI de fuentes autorizadas para reemplazar el sujeto, el predicado y, donde sea posible, el objeto (aunque algunas veces el objeto es una *literal*—una cadena fija de caracteres). Por ejemplo, el mismo enunciado se vería como sigue:

< http://www.cheese.com >	< http://purl.org/dc/terms/subject >	< http://dbpedia.org/page/Cheese >
Sujeto	Predicado	Objeto

El sujeto es el recurso Web que se está describiendo (www.cheese.com). El predicado aclara la relación entre el recurso y el objeto del enunciado; es un IRI del elemento *Subject* en DCMI. En otras palabras, dice que este recurso Web *trata acerca* del concepto representado en el objeto. El objeto es un IRI que puntea a la página dbpedia por el concepto queso. Más imple, la tripleta es un modo comprensible por la computadora de decir: “Esta página Web trata acerca de queso”, o “Esta página Web fue creada por Samantha Jones”, o “Este recurso se titula I Love You, Cheese”.

Si los datos de la Web Semántica se marcan más efectivamente y completamente usando tripletas que identifiquen a tosas las personas, lugares, cosas, ideas, etc., principales, las computadoras pueden usar estos metadatos para establecer vínculos entre recursos relacionados. Cuando muchos recursos de vinculan con IRI iguales o equivalentes, entonces dichos recursos están conectados son de interés potencial para los usuarios que buscan materiales acerca de un tópico particular; por un creador particular; un lugar o año particular; etc. Estas conexiones pueden ser cruzadas automáticamente, permitiendo a las computadoras ofrecer una amplia variedad de recursos como parte de los resultados de búsqueda (i.e., estos no serán dependientes únicamente de la correspondencia de las palabras clave en documentos con las palabras en las cajas de búsqueda). Esto permitirá no solo mejores resultados de búsqueda, sino también colecciones de información que pueden ser creados por medios automáticos y actualizados dinámicamente sin intervención humana; reúne juntos diversos recursos que,

en muchos casos, los seres humanos sería improbable que los encuentren, los datos vinculados hacen posible el descubrimiento de información acerca de entidades que de otra manera estarían separados por medios de codificación disociados o por distintos silos de datos.

Los aspectos prácticos del desarrollo de la Web Semántica están aún en sus primeras etapas. Mientras están poniéndose sus fundamentos, existen solo unos pocos productos concretos como demostración del concepto. La British Broadcasting Corporation (BBC) ha sido una de las primeras organizaciones en usar datos vinculados en varias partes de su sitio Web. Por ejemplo, en BBC Music se puede buscar una página dedicada a un músico favorito. En dicha página pueden encontrarse imágenes, datos biográficos de Wikipedia, archivos de audio y de video de la BBC, vínculos a las playlists que contienen música de dicho artista, información acerca de interpretaciones en vivo venideras de Songkick.com (un sitio dedicado a información sobre conciertos), noticias recientes de la BBC y vínculos a otros sitios acerca de músico elegido (por ej., el sitio MusicBrainz, páginas Web oficiales, páginas de Facebook, cuentas de Twitter, perfiles de Instagram, canales de YouTube) así como también información acerca de artistas similares. La cuestión más importante para mencionar acerca de este ejemplo es que las computadoras hacen el trabajo; las computadoras han reunido esta colección de datos desde diversas fuentes a los usuarios con una amplia variedad de información.

Las actividades actuales para abrazar los datos vinculados pueden llevar a una Web Semántica completamente fructífera, aunque debe completarse mucho trabajo para que esto ocurra. La adición de más metadatos y microdatos en la Web es una tarea enorme. Para que esto sea exitoso, los metadatos deben ser creados de forma más colaborativa. En otras palabras, la responsabilidad de crear los metadatos necesarios, debe compartirse entre muchas instituciones. La diversidad implicada por la utilización de metadatos desde múltiples fuentes y perspectivas pueden ayudar a crear una imagen más robusta de las entidades descritas por las computadoras y por los humanos. Y para que la imagen sea más completa pueden realizarse más conexiones. Muchos ven el desarrollo de los datos vinculados y al Web Semántica como la próxima frontera de la organización de la información, y se han dado pasos concretos para auxiliar al campo de la Bibliotecología y la Ciencia de la Información en el comienzo de esta tarea. Uno de los proyectos es BIBFRAME de LC, que constituye un intento de reemplazar el formato de codificación MARC actualmente usado en los catálogos con una aproximación de datos vinculados para representar e intercambiar datos bibliográficos en la Web y en el mundo de la red. Este proyecto, como muchos otros, está en su infancia y debería divisarse cuidadosamente por cualquiera que esté interesado en la organización de la información [...].

Existe un gran entusiasmo acerca del potencial de la Web semántica y con respecto a los cambios que pueden ocurrir como consecuencia de su maduración. Existirán efectos de amplio alcance sobre la catalogación y la organización de la información, pero en este momento, no es completamente evidente cuán exactamente cambiarán las cosas. Aunque algunos han expresado sus esperanzas acerca de este paso, existen aún muchos detractores que no pueden imaginar la visión de la Web Semántica que se manifiesta actualmente.

Colecciones Digitales

El desarrollo de Internet ha cambiado muchas cosas. Ahora contamos con medios mucho más rápidos de comunicación, compras instantáneas, noticias, entretenimiento bajo demanda, nuevas formas de encontrar el amor y muchas otras distracciones. También redefinió la noción de espacio. Actualmente vivimos mucho de nuestras vidas en espacios en línea—por ej., encontrándonos con amigos, asistir a clases, mirar televisión. Uno de

los usos más ambiciosos y benefactores de los espacios en línea ha sido el desarrollo de las *colecciones digitales*. Generalmente, estos espacios en línea contienen colecciones de recursos digitalizados o nacidos digitales que han sido seleccionados para su inclusión en base a propósitos de acceso y de preservación para la posteridad, una colección digital específica puede ser denominada *biblioteca digital*, *archivo digital* (*digital archive*) o *repositorio institucional*. Estos términos no son siempre definidos de manera precisa, lo que es un *repositorio* en una institución puede ser muy similar al *archivo digital* de otra. Como se denomina una colección digital individual, puede depender de la naturaleza del contenido y de la naturaleza de la institución que establece la colección. Por ejemplo, uno debería esperar encontrar mayoritariamente fuentes primarias (i.e., documentos originales) en los archivos digitales.

Las colecciones digitales son más sofisticadas y más accesibles que antes, aunque provienen de comienzos más humildes. En la década de 1990, la definición de biblioteca digital fue tema de debate. Existieron diversos experimentos sobre bibliotecas digitales, que no se consideran como tales hoy día. Por ejemplo, al nivel más simple están las colecciones de vínculos a recursos relacionados con un tema particular. Algunas veces tales colecciones—en realidad, bibliografías—son reunidas por bibliotecarios individuales de instituciones cooperantes. Para algunos, el término refiere a sitios como Project Gutenberg, el que en su forma inicial fue una colección extensa de “e-textos” para ser leídos en línea. Para muchos investigadores de Bibliotecología y Ciencia de la Información, la frase ‘biblioteca digital’ significa cosas diferentes para distintas personas. Algunos autores reportan haber hallado 64 definiciones formales e informales diferentes de *biblioteca digital*. No obstante, la frase biblioteca digital viene a significar colecciones en las que un sitio provee recursos de información digitalizados con una arquitectura y servicios para recuperar tales recursos. Resumiendo, las definiciones predominantes al inicio del siglo XXI, puede decirse que una biblioteca digital:

- Debe contener una colección organizada de recursos digitales—esto no constituye un conjunto de hipervínculos a otros materiales
- Es creada para un público, grupo de usuarios o comunidad particular
- Toma ventaja de la tecnología y de los recursos humanos (por ej., bibliotecarios)
- Provee acceso eficiente y rápido a los recursos digitales, frecuentemente si costo (aunque puede requerirse la membresía en una comunidad)
- Es propietaria, controla y posee los derechos sobre los recursos digitales distribuidos por ella

Las bibliotecas digitales son una extensión, una mejora e integración tanto de sistemas de recuperación de información como de múltiples instituciones que trabajan con información—de las que la biblioteca es solo una más. El alcance de las capacidades de las bibliotecas digitales no solo abarca lo vinculado con la recuperación, sino también con la creación y uso de la información. Las bibliotecas digitales se dirigen a comunidades de usuarios y constituyen extensiones de aquellos lugares físicos donde los recursos de información se seleccionan, coleccionan, organizan, preservan y brindan acceso, incluyendo bibliotecas, museos, archivos y escuelas. Hoy día, un ejemplo de una colección digital usualmente encontrada en un ambiente académico, es el repositorio institucional—un sistema en línea que colecciona, gestiona, disemina y preserva recursos digitales relacionados a la actividad intelectual de la comunicad académica.

Un proyecto exitoso en la historia de las bibliotecas digitales—que aun continua—es *American Memory Digital Library* de Library of Congress (LC). Comenzó como un

proyecto piloto en 1990 y evolucionó como una de las colecciones digitales más conocidas en el mundo. Esta biblioteca digital implicó un esfuerzo colaborativo entre bibliotecas académicas, bibliotecas públicas, bibliotecas de investigación, museos, sociedades históricas y archivos que digitalizaron colecciones sobre la historia norteamericana y la pusieron disponibles en el sitio de American History. otro ejemplo de proyecto cooperativo es Digital Public Library of America (DPLA), planificada para comenzar en 2010. Su objetivo es reunir juntas las riquezas de las bibliotecas, archivos, museos y sitios de patrimonio cultural de EE.UU., y ponerlos disponibles gratuitamente a los estudiantes, docentes, investigadores y el público en general. Un año antes de la apertura oficial, en abril de 2013, DPLA contenía más de 7 millones de ítems de patrimonio cultural digitalizados de 1200 instituciones contribuyentes a lo largo de EE.UU. un proyecto similar, Europeana Collections comenzó en 2008, actualmente cuenta con más de 53 millones de obras de arte, artefactos, libros, videos y grabaciones sonoras de toda Europa. Uno de los aspectos más innovadores de DPLA es la creación de una plataforma que permite nuevos y transformadores usos del patrimonio cultural digital de EE.UU. La interfaz de programación de aplicación (API) de DPLA y los datos abiertos, pueden ser usados por los desarrolladores de software, los investigadores y otras personas para crear nuevos ambientes dirigidos al aprendizaje, las herramientas de descubrimiento y las APPs. Por ejemplo, existen interfaces vinculadas al sitio Web de DPLA que permiten a los usuarios buscar imágenes por color dentro de la colección, comparar los resultados de búsqueda para dos términos o frases diferentes, jugar juegos de metadatos (por ej., agregar palabras clave a las imágenes como un juego) y encontrar al azar imágenes que contengan gatos en la colección.

¿Qué podría encontrarse en una colección digital? Casi cualquier cosa puede incluirse siempre y cuando pueda presentarse en formato digital. Aunque algunas colecciones digitales están focalizadas de manera muy estrecha, con límites a los tipos de recursos incluidos, otras son más amplias en su alcance. Por ejemplo, en FDR Presidential Library and Museum, en su colección digital se encuentran principalmente textos digitalizados—discursos, correspondencia, diarios, transcripciones de conferencias de prensa, órdenes ejecutivas, etc. Se puede encontrar grandes cantidades de texto entre la mayoría de las colecciones digitales, aunque mucha de la excitación acerca de estos espacios se da por la inclusión de recursos multimedia y objetos digitales complejos. En las colecciones digitales, como American Library o DPLA, uno encontrará una gran variedad de tipos de recursos. Por ejemplo, uno puede encontrar archivos de sonido de discursos, entrevistas, lecturas, sonidos naturales exóticos, etc. Los cortometrajes, videoclips, programas de tv y publicidades de servicios públicos, no son inusuales ni tampoco lo son las fotografías, dibujos, anuncios y datos geoespaciales digitalizados.

Las primeras colecciones digitales fueron desarrollos a medida, en tanto no existía software disponible para la venta para alojar estos proyectos digitales iniciáticos. Por ejemplo, La *University of California at Berkeley* contó con varios proyectos que contribuyeron con innovaciones para las bibliotecas digitales. Uno de ellos crea las especificaciones para codificar electrónicamente los instrumentos de localización (finding aids) para archivos y colecciones especiales. Este proyecto constituyó la base para lo que luego fue el esquema de metadatos *EAD (Encoding Archival Description)*. Otro proyecto propuso la codificación estándar de objetos digitales—denominado *Making of America II*. Este proyecto evolucionó en el esquema de metadatos *METS (Metadata Encoding and Transmission Standard)*.

En la medida que el énfasis ha virado de la experimentación a la implementación masiva, ha cambiado el foco enfatizando la estandarización, la organización, la usabilidad y la producción de software comerciales dirigidos a las instituciones que ingresan en el

terreno de las bibliotecas digitales. Muchas empresas de automatización bibliotecaria ofrecen soluciones para bibliotecas digitales y muchas empresas de gestión de tecnología multimedia ofrecen paquetes que incluyen el desarrollo de cursos de educación a distancia—que incluyen medios para crear bibliotecas digitales como apoyo a un curso o a un conjunto de cursos. Por ejemplo, ExLibris ofrece *DigiTool*, una solución que trabaja junto al SIGB y permite catalogar, gestionar, compartir, buscar y recuperar una colección digital. Otros productos son Portfolio ofrecido por SirsiDynix y Vital de Innovative Interfaces. Existe también software independiente para que las bibliotecas construyan sus propias bibliotecas digitales. Un producto es Omeka, un proyecto del Roy Rosenzweig Center for History and New media de la George Mason University. Omeka es un software de código abierto, gratis, que integra varios formatos de documentos diferentes, aunque con metadatos normalizados. Permite que bibliotecas, museos, archivos y otras instituciones que guardan patrimonio cultural, visualizar sus recursos digitales. Fue creado para contar con facilidad en su uso para aquellos que no conocen nada acerca de bibliotecas digitales, pero quieren crear colecciones o exhibiciones digitales. Para una institución con recursos tecnológicos significativos, puede ser una herramienta muy poderosa, y existiendo una amplia comunidad de programadores que pueden proveer ayuda así como agregados para mejorar el sistema. Otros softwares para colecciones digitales son ContentDM disponible en OCLC, DSpace, Fedora y Digital Common.

La organización de colecciones digitales es frecuentemente realizada usando algunos de los mismos métodos de organización que se utilizan en las bibliotecas, archivos y museos físicos; pero, estas aproximaciones a la organización de la información complementadas con el uso de esquemas de metadatos y de herramientas adicionales, así como también otros vocabularios controlados, taxonomías, formatos de codificación y estándares de contenido. Las aproximaciones a la organización pueden variar de institución a institución y entre paquetes de software. Como con los recursos tangibles, se describen los atributos importantes de los recursos digitales, se eligen puntos de acceso y los recursos son analizados de forma que el tema y el género/formato pueda ser descrito. Además, de los metadatos generados por estas actividades pueden necesitarse metadatos posteriores para tratar el ciclo de vida del recurso digital. Por ejemplo, los usuarios necesitan saber que el recurso es textual, que tiene una extensión de 120 páginas y los formatos en que esta disponible (por ej., HTML, PDF). Si el recurso es descargado o si requiere un software para interactuar con él, el tamaño y los requerimientos del sistema, implican metadatos necesarios. Si hay restricciones de acceso o uso, deberían establecerse con claridad. La descripción, además de identificar las partes responsables de la creación del contenido, debería incluir también a quienes digitalizaron el recurso y cuando lo hicieron, así como a quienes crearon el metadato, cuando fue creado y cuando fue actualizado por última vez. Algunos metadatos estructurales podrían necesitarse también para ayudar a juntar las piezas de objetos digitales complejos para operar sin complicaciones (por ej., así el video puede ser visto, o las páginas de un e-book pueden ser pasadas). Los recursos digitales pueden ser descritos en varios niveles.

Para contener los metadatos necesarios para cada recurso, algunas colecciones digitales usan esquemas establecidos como METS, DC, MODS, o aún la catalogación bibliotecaria tradicional completa (i.e., crear registros MARC con RDA, LCSH y otros estándares). Por ejemplo, LC usa tanto metadatos MARC como metadatos no-MARC en American Memory. Sin embargo, otras instituciones desarrollan sus propias aproximaciones, frecuentemente parcial o completamente basadas en uno o más esquemas de metadatos existentes. Por ejemplo, el Perfil de Aplicación de Metadatos de DPLA identifica varios esquemas de metadatos existentes como fuentes para su propio conjunto de elementos de metadatos. Antes de “reinventar la rueda” reutilizan elementos

de DC, Europeana Data Model (EDM), Simple Knowledge Organization System (SKOS) y otros esquemas al igual que otras colecciones digitales. DPLA emplea estrategias que cumplen con los protocolos OAI (Open Archives Initiative) para asegurar la facilidad al compartir los recursos digitales y los metadatos. Proyectos como estos también pueden utilizar una o más listas terminológicas controladas para describir temas, nombres, lugares, títulos, formatos, géneros, tipos de recurso, etc. Por ejemplo, American Memory usa nombres controlados de LC/NACO Authority File (LCNAF) y temas de LCSH en sus metadatos. Resumiendo, la organización de las bibliotecas digitales esta siendo lograda utilizando muchas de las aproximaciones propias usadas por las bibliotecas, archivos, museos y otras instituciones de preservación cultural, con herramientas como estándares de contenido, esquemas de metadatos, codificación XML o MARC, vocabularios controlados, etiquetado del usuario, taxonomías, etc. Estos componentes se describen más completamente en los capítulos posteriores de este libro.

Arquitectura de la información

Las bibliotecas digitales, los sitios Web generales, las fuentes de noticias en línea, las librerías y los bancos en línea, las aplicaciones móviles, Facebook, entre otros, son todos ejemplos de espacios en línea. Al igual que los arquitectos determinan las necesidades de las personas que usarán un espacio, y crean un modelo o maqueta para satisfacer dichas necesidades para el diseño de edificios u otras estructuras que servirán a las necesidades de las personas, los arquitectos de la información determinan los usos que presentará la información y crearán modelos para los modos de encontrar información requerida, además de crear interfaces atractivas para la información. La arquitectura de la información es mucho más que el diseño Web, aunque su desarrollo y emergencia como campo disciplinar se encuentra estrechamente asociado con la creación de sitios Web. La arquitectura de la información (AI) es el proceso de diseño, implementación y evaluación de espacios de información que son humana y socialmente aceptados para sus destinatarios interesados. Esta definición queda abierta para cubrir los aspectos organizativos, empíricos y de acción, permitiendo que los roles de AI cubran dichos aspectos. Otra definición de AI se focaliza en: el diseño estructural de entornos de información compartida; la combinación de sistemas de organización, de etiquetado, búsqueda y navegación de sitios Web e Intranet; modelar experiencias y productos informativos para sostener la usabilidad y la capacidad de encontrar información; una disciplina emergente y una comunidad de prácticas focalizada sobre principios de diseño y de arquitectura digital.

Existe aún desacuerdo acerca de que forma parte de AI. Las diferencias de puntos de vista radican en la ausencia de una definición formal. Una línea divisoria ha sido trazada entre dos visiones en competencia dentro del campo, conocidas como Gran AI vs. Pequeña AI. *Gran AI* es el proceso de diseño y construcción de recursos de información que son útiles, usables y aceptables. Desde esta perspectiva, la AI debe cubrir la experiencia del usuario y la aceptación del recurso de parte de la organización. Por otro lado, la *Pequeña AI* refiere a una actividad mucho más constreñida que trata con el mantenimiento y la organización de la información, aunque no involucra el análisis de la respuesta del usuario o el diseño gráfico del espacio de información. La Gran AI tiende a ser vista como un enfoque arriba-abajo, concibiendo el producto completo y su impacto organizacional o humano. La Pequeña AI es vista desde un enfoque abajo-arriba, considerando los metadatos y el vocabulario controlado de la organización de la información, sin tratar directamente con, y nunca evaluando formalmente, la experiencia del usuario del espacio resultante. Algunos arquitectos de la información rechazan la

noción de que la arquitectura de la información constituye un nuevo enfoque de organización de la información que se ha practicado en bibliotecas, archivos y museos por un largo tiempo. No obstante, los paralelismos son llamativos. Los bibliotecarios han comprendido la necesidad de adquirir recursos de información selectivamente y de organizarlos de manera que puedan ayudar a los usuarios a ganar acceso a ellos en tanto lo necesiten. Sin embargo, parece haber algún acuerdo en el deseo de diseñar una “ecología de la información” útil y completa. Para lograr esto, el arquitecto de la información debe:

- Crear espacios en línea con las necesidades, comportamientos y límites de los usuarios en mente
- Comprender el contexto específico del sitio (por ej., la misión, objetivos, estrategias, etc., que no son únicas para la institución que crea el espacio)
- Organizar la información en línea (por ej., las cosas que buscan los usuarios) lógica y claramente para proveer un fácil acceso a esta información

El proceso incluye diseñar interfaces y sistemas de navegación utilizables (por ej., mapas del sitio, taxonomías, menús), además de crear un diseño gráfico en general agradable. [...]. El campo se dedica a lograr que la información se encuentre y comprenda por la creación de estructuras de información únicas y lógicas en ambientes en línea.

Diversos autores identifican las siguientes etapas en el proceso de AI: investigación, estrategia, diseño, implementación y administración:

- **Investigación**, consiste en una revisión de los materiales con antecedentes, lograr la comprensión de los objetivos y del contexto del negocio, examinar la AI existente, el contenido y el público al que va destinado y, finalmente, conducir estudios necesarios para explorar la situación. En esta etapa debe desarrollarse una comprensión del contenido. Esto incluye reunir información acerca de la propiedad, los tipos, formatos y la estructura del contenido; los metadatos existentes; y, la cantidad o volumen del contenido.
- **Estrategia**, se erige de la comprensión contextual desarrollada en la primera fase y define los niveles altos de la organización y la estructura de navegación del sitio, mientras también considera tipos de documentos y esquemas de metadatos. Por ejemplo, uno debe desarrollar ideas acerca de como los usuarios accederán a la información del sitio (por ej., por medios alfabéticos, cronológicos, tópicos u orientado a tareas)
- **Diseño**, involucra la creación de proyectos de acción detallados, un esquema de metadatos, etc., para ser usados por diseñadores gráficos, programadores, autores de contenido y equipo de producción. En esta etapa, se crean categorías de contenido, menús de browsing, vocabularios controlados, funciones de búsqueda y sistemas de etiquetas.
- **Implementación**, consiste de diseños que se utilizan en la construcción, la prueba y lanzamiento del sitio—la organización y el etiquetado de documentos, la solución de problemas y el desarrollo de documentación—ocurren en esta fase.
- **Administración**, involucra la evaluación continua y la mejora de la AI del sitio.

Las etapas de estrategia y de diseño son las únicas que requieren una comprensión minuciosa de las bases teóricas de la organización de la información, incluyendo el

conocimiento de los metadatos, la provisión de puntos de acceso con todas las relaciones asociadas a ellos; la aproximación temática por categorías, clasificación y/o etiquetas alfabéticas; y, el diseño de sistemas que permitan la visualización de los resultados de manera lógica y utilizable.

Indización y Resúmenes

La indización y el resumen son dos aproximaciones para reducir el contenido informativo a una representación abreviada y comprensiva de un recurso informativo. La indización tiene una tradición extensa y cambiante en términos de que es, quien lo ha hecho, porque lo ha realizado y como se ha hecho. La historia del resumen es menos volátil y ha evolucionado en los siglos XX y XXI hacia formatos específicos con públicos particulares.

Indización

La indización es el proceso por el cual el contenido de un recurso de información se analiza y se determina y expresa de manera concisa la atinencia (aboutness) de dicho ítem. La indización también está interesada en describir el recurso de información de forma que los usuarios sean conscientes de los atributos del documento—como su autor, título, extensión y ubicación del contenido. La indización se aplica típicamente solo a textos. No obstante, la indización de imágenes es un área en desarrollo. Existen tres tipos básicos de indización: indización de final de libro, indización de bases de datos e indización Web.

En la *indización de final de libro*, el índice es una lista de términos o frases ordenadas alfabéticamente con referencias de ubicación que posibilitan al usuario recuperar el contenido requerido. El lenguaje de los términos de indización es derivado del lenguaje del texto—por ello, la indización realizada en este contexto se denomina *indización derivada*. Un buen índice de final de libro también incluye subencabezamientos, variantes y referencias cruzadas. La indización en listas es realizada principalmente por especialistas independientes que establecen contratos con las editoriales, aunque algunas de ellas mantienen personal de indización estable, o pueden ser desarrolladas por los mismos autores.

En la *indización de bases de datos*—también denominada indización de publicaciones seriadas—cada ítem de una base de datos se representa por un conjunto de descriptores y, en algunas instancias, por un código de clasificación. Esta indización utiliza normalmente un tesoro o vocabulario controlado, desde el que el indizador selecciona o asigna los términos apropiados. El alcance y el número de descriptores asignados a un ítem se determina mediante políticas del proveedor de las bases de datos e indizadores entrenados especialmente o que forman parte de la planta de personal.

La *indización Web*—o *indización de Internet*—aún se encuentra en desarrollo en lo concerniente a la jerga y a su práctica actual. Actualmente, este tipo de indización cae en dos categorías básicas:

1. *Indización de final de libro* (back-of-the-book style indexing), frecuentemente denominada Indización A-Z, que utiliza vínculos de índice codificados dentro del sitio Web;
2. *Indización de motores de búsqueda* (search engine indexing), más precisamente descrita como indización automática de sitios Web, donde:
 - (a) los sitios Web se buscan en base a términos de búsqueda generados por

los usuarios; (b) se mantiene un índice de palabras encontradas, junto con su ubicación; (c) las búsquedas futuras sobre los mismos enunciados utilizan los índices guardados.

La mayoría de los índices A-Z son realizados por profesionales independientes, o eventualmente, por indizadores de libros que expandieron la prestación de sus servicios a la creación de índices de la Web. Existen varios software que pueden ser utilizados para elaborar índices. Utilizan una variedad de técnicas, cuya eficacia depende de variables como el costo, el tipo, el tamaño de los archivos a ser indizados, las preferencias individuales, etc. La *American Society of Indexing (ASI)* lista algunos tipos de herramientas para indización:

- *Herramientas dedicadas o aisladas* (standalone or dedicated tools), que facilitan la creación de índices de final de libro a partir de la prueba de galera con las páginas numeradas.
- *Herramientas de indización incrustadas* (embedded indexing tools), que permiten la inserción de asientos del índice en forma de texto invisible en archivos electrónicos.
- *Herramientas de palabras clave y etiquetado* (tagging and keywording tools), que permiten que el código de indización—en lugar del texto invisible—se incruste en el texto electrónico y facilite la creación de datos directamente incrustados (hard coded jump) similares a los vínculos Web. Estos descansan sobre las palabras usadas por los autores del texto, no en los conceptos del indizador.
- *Software de indización automática*, que acompaña a la mayoría de los procesadores de texto y construye concordancias o listas de palabras directamente desde el texto usando el lenguaje del autor (nuevamente, no constituyen auténticos índices ya que incluyen conceptos clave que no utilizan las palabras del autor).
- *Herramientas de búsqueda de texto ponderado o de texto libre* (free-text and weighted-text searching tools) que permiten asignar valores a las palabras y a las frases.

Resumen

El resumen consiste en el proceso de analizar el contenido de un recurso de información y escribir un sumario o sinopsis sucinto de la obra. Generalmente, el resumen se realiza para publicaciones académicas o revistas especializadas. La extensión, estilo y nivel de detalle en un resumen puede variar de acuerdo al público que se destina. Generalmente, un resumen no es una revisión de la obra, ni evalúa o interpreta la obra que esta siendo resumida, aunque los resúmenes críticos incluyen algún texto de carácter evaluativo. Aunque contiene palabras clave y conceptos provenientes del texto del documento, el resumen es un documento original antes que un pasaje del texto. Existen varios tipos de resúmenes:

- **Indicativo:** descriptivo del contenido, pero sin proveer resultados—también se lo denomina *resumen descriptivo*.
- **Informativo:** resume los resultados o los resultados enfatizados.
- **Crítico:** revisión crítica condensada—es un tipo poco común.

- **Estructurado:** comporta un formato no narrativo; incluye factores específicos, como objetivos, métodos, resultados, etc.
- **Modular:** incluye cinco secciones separadas—cita, comentario, resúmenes indicativo, informativo y crítico.

Técnicamente, un resumen es el sumario del texto; en la práctica, un resumen formal consiste tanto del título y la cita de la obra resumida así como del sumario del texto. El resumen es realizado por los autores y por profesionales de la información especialmente entrenados. Las revistas especializadas requieren frecuentemente que un resumen acompañe los artículos que los autores envían para ser publicados. Los resúmenes también son escritos por profesionales que son parte del personal o son contratados por las editoriales. Las editoriales elaboran políticas para guiar la confección de resúmenes, las que no son idénticas entre sí pero se diseñan en respuesta a las necesidades de públicos específicos.

Los resúmenes cuentan con un número de usos en la organización y la recuperación de información. Los investigadores encuentran que revisar resúmenes en lugar del texto completo, constituye un ahorro de tiempo. Los resúmenes ayudan a decidir que artículos requieren leerse de manera completa y cuales pueden leerse por encima o ser evitados. Los bibliotecarios y otros profesionales de la información encuentran que el uso de resúmenes asiste en la velocidad y en la utilidad de las búsquedas de los usuarios. Los indizadores de bases de datos, que generalmente utilizan en su labor el título y el resumen del texto, requieren un resumen bien escrito y preciso.

Administración de registros en oficinas

La administración de registros es la terminología aplicada al control de la explosión de la información electrónica en oficinas y en otros entornos administrativos. Tiene su origen en los sistemas de archivo de oficinas desarrollados a lo largo del siglo XX. Estos sistemas se han visto afectados por los avances tecnológicos—máquinas de escribir, fotocopadoras y computadoras. El uso de computadoras en este contexto ha sido denominado, algunas veces, como *administración de datos*. Los sistemas de gestión de registros poseen una fuerte relación con los archivos en tanto ese es el lugar donde se depositan los registros de la organización cuando su vida activa ha transcurrido y dichos registros se han vuelto inactivos.

Al igual que otros sectores de la sociedad, la administración de datos implicó guardar, archivar y mantener registros de papeles. En tales épocas se archivaba copia de un registro en un solo lugar. Las etiquetas de las carpetas generadas por un administrador de registros generalmente no le resultaban lógicas al siguiente administrador. En la medida en que se comienza a guardar información en carpetas y archivos electrónicos, los puntos de acceso—es decir, las etiquetas de las carpetas—comenzaron a volverse invisibles. Esta situación no acarrea problemas, en tanto y en cuanto quien elaboraba el archivo electrónico documentara lo que estos contenían. La situación empeoró en la medida en que se desarrollan y adquieren computadoras personales más poderosas que permiten guardar la información en carpetas y archivos propios, dado que se presentan inconvenientes con la continuidad de dichos archivos cuando eran abandonados.

Durante mucho tiempo se han automatizado diversas operaciones, cada una con sistemas propios—como, por ejemplo, las planillas de sueldos, el libro mayor, las cuentas a pagar, los inventarios, etc. En años atrás se ha llevado a cabo la integración de estos sistemas, dando como resultado la presencia de muchos campos redundantes con poca documentación acerca de su contenido. Da la sensación que dichos campos deberían

contar con la misma información, no obstante los datos ingresados allí son muy diferentes—por ejemplo, en el campo *nombre* de la lista de sueldos se presenta el nombre completo mientras que en el mismo campo de la carpeta que enumera los recursos humanos se presenta el nombre abreviado con iniciales.

En 2001, ISO publica un estándar para administración de datos. Esta norma define administración de registros (record management) como el campo responsable de la gestión para el control sistemático y eficiente de la creación, recepción, mantenimiento, uso y disposición de registros. Además, en la descripción de las características de los sistemas de registros (records system), el estándar establece que estos sistemas deberían: (a) capturar rutinariamente todos los registros dentro del alcance de las actividades del negocio que cubre; (b) organizar los registros de modo que reflejen los procesos del negocio del creador de los datos; (c) proteger los registros de alteraciones o disposiciones no autorizadas; (d) funcionar rutinariamente como fuente principal de información acerca de las acciones que se documentan en el registro; (e) proveer un fácil acceso a todos los registros relevantes y a los metadatos relacionados. Existe un número de sistemas de gestión de registros (records management systems) disponibles en el mercado, que siguen el curso y almacenan registros, proveen funciones de auditoría y seguridad, tienen módulos de gestión de contenido e identidad de usuarios, etc.

Los administradores de registros gestionan la explosión de información mediante la aplicación de principios de organización de la información. Las unidades que deben organizarse en el entorno electrónico administrativo son: directorios, archivos, programas y valores de los campos. La organización puede ser mediante sistema—por ejemplo, la lista de sueldos, el presupuesto—o por tipo de registro—por ejemplo, los nombres de personas, los registros de inscripción. Los administradores de datos deben estar al tanto de aquella información que se ingresa en varios archivos diferentes. También deben contar con métodos para gestionar conceptos que tienen la misma denominación pero diferente propósito o aplicación—por ejemplo, la noción de dedicación parcial en una universidad puede ser diferente dependiendo de si refiere a la lista de sueldos, al personal docente, a los estudiantes de posgrado o a los estudiantes de grado.

Gestión de Información Personal

La *gestión de información personal* se define como la práctica y el estudio de las actividades que una persona lleva a cabo para adquirir o crear, almacenar, organizar, mantener, recuperar, usar y distribuir información necesitada para cumplir con tareas o usos y con varios roles o responsabilidades. Dado que la gestión de la información personal esta conectada directamente con la vida cotidiana de un individuo, existe un creciente interés en el desarrollo de herramientas y dispositivos que faciliten la gestión de información personal, así como también más investigación acerca de como las personas administran su información personal.

La gente organiza su información personal en varios formatos—incluyendo documentos en papel, e-mails, fotos, música, álbumes, recetas y diferentes tipos de archivos digitales—es sus espacios personales tales como oficinas u hogares. Cuando organizan información personal, los individuos tienen diferentes hábitos y procesos. Por ejemplo, algunas personas tienen organizadas ordenadamente sus oficinas, mientras que otras están desorganizadas con pilas de papeles.

En el caso de la organización de archivos digitales, han existido debates sobre la necesidad de organizarlos en carpetas (folders), particularmente desde que las personas los buscan en sus dispositivos personales. Sin embargo, muchas personas que organizan sus archivos digitales en carpetas, indican que el acto de organizar archivos tiene mas

funciones que solo encontrar ítems específicos. Estas funciones incluyen recordarle a las personas tareas pendientes y ayudarlas posteriormente a comprender las relaciones entre los ítems de información. Además, la función de búsqueda puede ser de menor utilidad cuando existe una cantidad de archivos con nombres similares o la palabra clave exacta no puede ser debidamente recuperada.

Existen varios factores que influyen las decisiones de organización, tales como donde y como organizar ítems. Los principales factores incluyen: el uso/propósito de un ítem de información; el formato del ítem de información y el tema del ítem de información. La organización de la información personal puede presentar desafíos en la medida que involucra varias decisiones que necesitan hacerse en base al uso futuro de, la necesidad de, el interés en y el valor de la información, todas ellas difíciles de predecir y que pueden cambiar con facilidad. Actualmente, dos problemáticas hacen aún más desafiante la organización de la información personal. Estos son la *sobrecarga de información*, que ocurre cuando la persona recibe más información de la que puede procesar y la *fragmentación de la información*, que se presenta al contar con ítems de información dispersos entre múltiples dispositivos personales y herramientas de recuperación de información en diferentes formatos. Sin embargo, independientemente del formato, la organización de la información personal efectiva facilita las tareas de encontrar y utilizar la información eficientemente, situación que aumenta la productividad individual.

Gestión del conocimiento

Originalmente, la frase “*El conocimiento es poder*” se aplicó a los individuos e implicó que las personas que incrementaban su conocimiento, aumentaban su poder en la sociedad. Durante la década de 1980, se amplió la misma idea a las organizaciones. En ese momento se dio una drástica reducción de personal en las organizaciones para bajar los gastos e incrementar las ganancias. En dicho proceso fue obvio que las organizaciones perdieran importantes dosis de conocimiento en la medida que los empleados eran despedidos o se retiraban, llevándose consigo el conocimiento acumulado. En la misma época, existió un gran desarrollo tecnológico dirigido a ahorrar costos emanados del reemplazo de mano de obra humana. Otra vez, el conocimiento sostenido y aplicado por humanos no puede ser reemplazado por las máquinas. Para que una organización sobreviva, el conocimiento es reunido para asumir los desafíos que enfrente la organización. La gestión de dicho conocimiento aumenta su poder.

La idea de transmitir el conocimiento generado por un trabajo ha pervivido por siglos. Los aprendices han tomado varios oficios trabajando al lado de un trabajador experimentado. Los jóvenes, frecuentemente, siguen a sus padres en la empresa familiar. Recientemente, apareció la figura del mentor. También, suele solicitársele a las personas que dejan su trabajo que entrenen a su sucesor.

La *gestión del conocimiento* (KM, Knowledge Management) es el proceso de capturar, desarrollar, compartir y utilizar información organizacional para la toma de decisiones bien informada. El concepto de gestión del conocimiento se transforma en un intento de capturar el conocimiento de los empleados con tecnología avanzada, de manera que el conocimiento pueda almacenarse y compartirse fácilmente. En la medida que la gente se ve abrumada por el aumento de la información disponible a través de los rápidos desarrollos tecnológicos, el KM emprende el rol adicional de arreglárselas con la explosión informativa. En el contexto KM, el proceso comprende tres componentes principales: personas, procesos y tecnología.

Gestionar el conocimiento requiere definir conocimiento. Este ha sido caracterizado de varios modos—por ejemplo, residiendo en la mente de las personas antes que almacenado de cualquier forma; como una combinación de información, contexto y experiencia; como aquello que representa la experiencia compartida entre grupos y comunidades; o, como una información de alto valor que se aplica a decisiones y acciones. Es importante observar que el conocimiento no constituye una cosa o un sistema, sino un proceso activo de naturaleza efímera. Si se toma esta perspectiva nadie—menos una corporación—puede poseer el conocimiento. El conocimiento en sí mismo no puede almacenarse, ni puede medirse como capital intelectual, y ciertamente nada de él puede gestionarse. No obstante, los gestores del conocimiento desarrollan herramientas, políticas e incentivos para alentar a las personas a compartir lo que conocen.

El KM parece tener su origen en 1995 con la extensión de ideas acerca del conocimiento tácito versus el conocimiento explícito—puestas a discusión por Ikijuro Nonaka e Hirotaka Takeuchi. Ambos autores plantearon que el conocimiento tácito esta oculto en la mente humana y no puede representarse fácilmente por medios electrónicos. No obstante, puede explicitarse con el grado necesario para lograr una innovación específica. Describen el proceso espiralado de compartir conocimiento tácito con otros por medio de la socialización—seguido por los oyentes que internalizan el conocimiento—y, consecuentemente, se crea nuevo conocimiento, el que a su vez es compartido. De aquí se sigue que todo el conocimiento en la mente de las personas pudiera o debiera explicitarse. No obstante, los primeros programas de KM intentan descorporizar todo el conocimiento de sus poseedores para hacer de él una ventaja organizacional. Fueron creados software para ser usados con este propósito—por ejemplo, *Knowledge Base Software* de Novo Solutions que brinda herramientas de entrenamiento para nuevos empleados, centraliza y retiene el conocimiento de los empleados y crea y actualiza artículos sobre KM, categorizados y disponibles para la búsqueda. El software *Lotus* de IBM busca crear, organizar, compartir y gestionar contenidos de negocios para “proveer la información correcta para la gente indicada” y “reunir e intercambiar información entre redes profesionales y comunidades de expertos para auxiliar en la rapidez de la ejecución de las tareas”. Algunos autores proveen una lista de herramientas y de mecanismos tecnológicos y no tecnológicos para implementar diferentes fases del ciclo de gestión del conocimiento (por ej., creación y captura del conocimiento, transferir y compartir el conocimiento y usar y aplicar el conocimiento). Otros autores ponen en consideración una cantidad de aspectos acerca del lugar de la tecnología en la gestión del conocimiento:

- Un único conjunto de herramientas no puede ser obligatorio, dado que cada organización y sus empleados necesitan decidir por sí mismos que herramientas y tecnologías son fáciles de usar y son útiles para sus necesidades actuales.
- Las herramientas tecnológicas se encuentran cambiando continuamente, por lo que no puede haber un conjunto permanente de recomendaciones que sean verdaderas a lo largo del tiempo. Lo que si permanece consistente es la necesidad de crear, compartir y usar conocimiento en las organizaciones.
- Una organización necesita considerar el costo de adoptar cualquier conjunto particular de herramientas o de tecnología (i.e., compra/licencia y el costo de mantenimiento)
- La tecnología no es el componente más importante en las implementaciones KM. La gestión del conocimiento trata acerca de personas, no acerca de

herramientas y de tecnología. La tecnología se necesita para soportar las necesidades de las personas, y no a la inversa.

Algunos autores observan que KM involucra, frecuentemente, en una empresa: la identificación, selección y descripción de recursos de información pertinentes a las necesidades de la empresa; identificar los patrones de flujo de información entre individuos y grupos—por ejemplo, averiguando quien hace que consultas, aprendiendo que información se obtiene de que fuentes, determinando que tipo de información no esta fácilmente disponible o accesible—y, diseñando y desarrollando sistemas amigables para el acceso a la base de conocimiento de la empresa.

Las problemáticas centrales para personas en el ámbito de la organización de la información consisten en describir, clasificar y recuperar lo que ha sido almacenado. En el contexto de KM, la organización de la información significa que el conocimiento de la organización debe ser ordenado, etiquetado—es decir, descrito—y clasificado en diferentes temas o grupos, si se requiere recuperarlo cuando se necesite. Las acciones de capturar el conocimiento tácito y describirlo, es muy parecido al proceso de organización de información de las bibliotecas. Esencialmente, la catalogación descriptiva y la catalogación temática son logradas cuando los consultores o especialistas trabajan describiendo una pieza de conocimiento completando una plantilla normalizada con un número de atributos y, posteriormente cuando usan facetas para categorizar dicha pieza de conocimiento. Si esto se llama análisis facetado, etiquetado o provisión de metadatos, el significado nuclear de la catalogación y de la clasificación existen tanto en las empresas consultoras como en las bibliotecas.