

ORGANIZACIÓN
Y DESCRIPCIÓN DE RECURSOS
DE INFORMACIÓN
DIGITAL

Robert J. Glushko (ed.)





ORGANIZACIÓN Y DESCRIPCIÓN DE RECURSOS DE INFORMACIÓN DIGITAL

v.1.0

Traducción parcial del libro
The discipline of organizing, 4ª
ed. 2016 Core Concepts Edition,
editado por la red internacional
iSchools, bajo licencia Creative
Commons CC-BY-NC, disponible
en acceso abierto en
[https://ischools.org/
Discipline-of-Organizing](https://ischools.org/Discipline-of-Organizing)

Este libro recoge

El capítulo 1 “**Foundations for Organizing Systems**”, del autor Robert J. Glushko;

El capítulo 4 “**Resources in Organizing Systems**”, de los autores Robert J. Glushko, Daniel D. Turner, Kimra McPherson y Jess Hemerly

Y el capítulo 5 “**Resource Description and Metadata**”, de los autores Robert J. Glushko, Kimra McPherson, Ryan Greenberg, Robyn Perry, Matthew Mayernik, Graham Freeman y Carl Lagoze.

Traducción realizada por Inés Megido González y Silvia Saorín Miralles, bajo la supervisión técnica y coordinación de Tomás Saorín, Universidad de Murcia.

La obra se encuentra disponible en acceso abierto en la sección de publicaciones de la web de FESABID
<http://www.fesabid.org/federacion/publicaciones>
en doble formato:
libro electrónico en PDF interactivo y fichero editable en formato PDF de alta calidad.

Edita: FESABID, diciembre de 2020
Licencia de reutilización del texto en español:
Atribución-NoComercial 4.0 Internacional (CC BY-NC 4.0),
<https://creativecommons.org/licenses/by-nc/4.0/deed.es>
Diseño y maquetación:
www.ikonovisual.com

Publicación realizada con la colaboración del Ministerio de Cultura y la Cátedra de Transparencia y Gestión de Datos de la Universitat Politècnica de València.





ORGANIZACIÓN
Y DESCRIPCIÓN DE RECURSOS
DE INFORMACIÓN
DIGITAL

Robert J. Glushko (ed.)



ÍNDICE

- 7** **TL;DR » PRESENTACIÓN DE LA EDICIÓN “LITE” EN ESPAÑOL**
- 15** **CAPÍTULO I.**
SISTEMAS DE ORGANIZACIÓN, MARCO DE REFERENCIA
- 1.1 Organizar, una disciplina global
 - 1.2 El concepto de “Sistema de Organización”
 - 1.3 El concepto de “Recurso”
 - 1.4 El concepto de “Colección”
 - 1.5 El concepto de la “Ordenación Intencionada”
 - 1.6 El concepto de “Principio de Organización”
 - 1.7 El concepto de “Agente”
 - 1.8 El concepto de “Interacciones”
 - 1.9 El concepto de “Recurso de Interacción”
- 51** **CAPÍTULO II**
LOS RECURSOS EN LOS SISTEMAS DE ORGANIZACIÓN
- 2.1 Introducción
 - 2.2 Cuatro distinciones sobre los recursos
 - 2.3 La identidad de los recursos
 - 2.4 El nombre de los recursos
 - 2.5. Los recursos y el paso del tiempo
- 103** **CAPÍTULO III**
DESCRIPCIÓN DE RECURSOS Y METADATOS
- 3.1. Introducción
 - 3.2. Visión general de la descripción de recursos
 - 3.3. El proceso de descripción de recursos
 - 3.4 La descripción de recursos no textuales
- 164** **DOS EPÍLOGOS Y UN DESEO**
- GIGO » LOS DATOS NO SON NÚMEROS, SON ORGANIZACIÓN**
por Antonia Ferrer Sapena
 - DIY » ORGANIZAR INFORMACIÓN NO TIENE FIN**
por Alicia Sellés

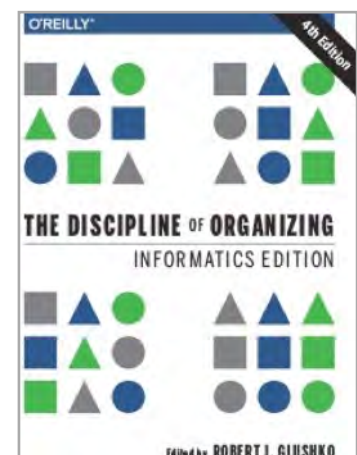
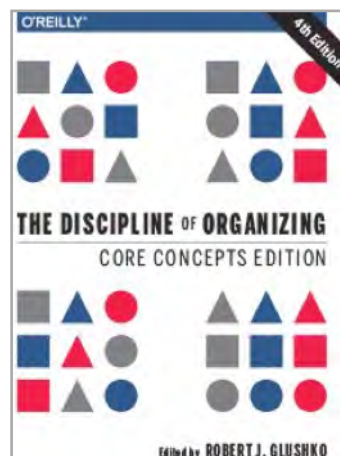
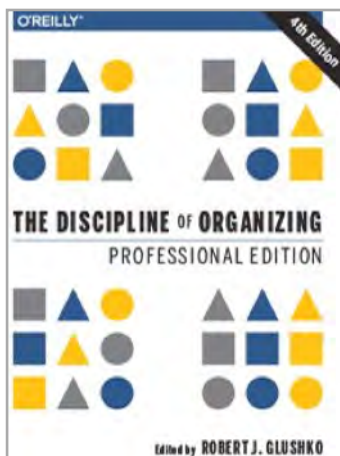




TL;DR » PRESENTACIÓN DE LA EDICIÓN

“LITE” EN ESPAÑOL

Organización y descripción de recursos de información digital es el título por el que finalmente nos hemos decidido para la traducción al español de tres de los capítulos del manual de referencia de la red internacional de iSchools **The discipline of organizing**, en el que se aborda el marco genérico de los sistemas de organización aplicables especialmente sobre objetos informativos digitales. Se trata de una obra que presenta de forma integrada los aspectos conceptuales que subyacen en diferentes campos –representación del conocimiento, bibliotecas, archivos, museos, publicación digital, recuperación de información, gestión de contenidos, diseño de servicios de información digitales– y que explican los fundamentos compartidos del continuo reto social de organizar información para permitir su acceso, enriquecimiento, almacenamiento y descubrimiento. La obra, coordinada por Robert J. Glushko, ha ido incorporando progresivamente los enfoques de áreas en auge, como la ciencia de datos, el posicionamiento web o la web de datos, creando un manual que propone un utillaje léxico compartido que enlaza desde ámbitos clásicos como la biblioteconomía y documentación hasta el de la gestión corporativa de activos de información, la gestión de datos públicos abiertos, pasando por las ingenierías de la información y las ciencias cognitivas.





El libro que extractamos y traducimos es una obra ambiciosa. La primera edición de 2013 del libro ***The discipline of organizing***, conocido como **TDO book**, fue reconocida por ASIS&T (*Association for Information Science and Technology*) como el libro del año en el campo de *information science*. Ha sido reeditado en sellos editoriales de referencia en el campo de la información digital, como O'Reilly y el MIT o la Universidad de Berkeley, y va ya por su cuarta edición, en su voluntad de no perder el paso a los acelerados cambios de tendencia en organización del conocimiento y explotación de información digital. Además, está concebido de una forma atrevida: existe una versión profesional, donde cada cita se enmarca en un área (Ciencias cognitivas, informática, LIS), junto a otra edición para lectores del área de la "informática", junto a la denominada edición *core concepts*, que reduce el aparato bibliográfico y busca una lectura más directa, que es la que tomamos como punto de partida en esta traducción.

¿Qué es "la disciplina de la organización"? Si bien está claro el sentido del título de la obra, en su traducción hemos optado por desviarnos de la literalidad del término "disciplina", y preferido el de "campo", con resonancias a campo profesional, a conjunto de técnicas aplicadas sobre un mismo objeto, y frente a la ambivalencia de "organización" como entidad y como actividad, hemos preferido usar la forma verbal de "organizar". En español podríamos explorar un título como el de "Organizar información: un campo abierto". Dado que además esta edición representa solo unos pocos aspectos del libro completo, añadir el concepto de "descripción de recursos de información" ayuda a delimitar su alcance y, por otro lado, darle relevancia al objeto de la "información digital", el cual nos gustaría que se leyera como "contenidos digitales". La información digital no es una abstracción útil, necesita existir materializada en contenidos con una alta dosis de inteligencia incorporada: contenidos que expresan sus significados, relaciones y estructura de forma clara al resto de actores del medio digital. El contenido digital precisa ser "ingenierizado" con precisión para funcionar con fluidez en la red; es un contenido con una intensa carga de diseño que suponen decisiones tecnológicas dependientes del contexto, de estándares, plataformas y servicios de mediación, que lo enriquecen y multiplican sus prestaciones.

Los autores usan el término "Sistemas de organización" como un continuo histórico y como un enfoque compartido en diferentes dominios:

"La elaboración de conceptos, métodos y tecnologías para describir y organizar los recursos ha sido una actividad humana esencial durante milenios, que ha evolucionado tanto en respuesta



a las necesidades humanas como para permitir otras nuevas. Los Sistemas de Organización permitieron el desarrollo de la civilización, desde la agricultura y el comercio hasta el gobierno y la guerra. Hoy en día, los Sistemas de Organización están incorporados en todos los ámbitos de la actividad económica, incluida la investigación, la educación, el derecho, la medicina, los negocios, la ciencia, la memoria institucional, la memoria sociocultural, el gobierno, la rendición de cuentas, así como en los actos ordinarios de la vida cotidiana.”

Las diferentes formas en que la organización de información se hace presente en nuestra sociedad digitalizada tiende a hacerse invisible. La información es ubicua y va pegada a casi cualquier interacción, y, por ello, se da por descontada o se naturaliza, como si fuera un resultado espontáneo de los procesos. Millones de personas crean y navegan sitios web, hacen blogs, etiquetan, twittean y cargan y descargan contenidos de todo tipo de medios sin pensar “estoy organizando información” o “estoy recopilando datos”. En el fondo generan productos dotados de una arquitectura de la información. Para los autores del libro una de las ideas centrales es que la distinción entre la organización de la información y la recuperación de la información que suele manifestarse en las disciplinas académicas y los planes de estudio es mucho menos importante de lo que era antes. El “TDO book”, por tanto

“tiene pocas divisiones marcadas entre los temas de organización de la información (IO) y de recuperación en formación (IR). En cambio, explica los conceptos clave y los desafíos en el diseño y el despliegue de los sistemas de organización de manera que enfatiza continuamente las relaciones y las compensaciones entre IO e IR. El concepto de Sistema de Organización destaca las dimensiones y las decisiones de diseño que determinan colectivamente el alcance y la naturaleza de la organización de los recursos y las capacidades de los procesos que comparan, combinan, transforman e interactúan con los recursos organizados.”

La disponibilidad en español de esta obra puede ayudar a mejorar la formación en organización de información, que es un elemento constitutivo de diversos estudios universitarios y campos de especialización académica y profesional, que proceden de tradiciones separadas. Esta traducción es parcial, puesto que se opta por abrir un camino en las abundantes lagunas en traducciones de manuales actualizados y reputados vinculados a la información digital. El libro TDO es una obra extensa – y por ello sugerente, potente e integradora – pero hemos preferido mostrar resultados tangibles en un plazo corto.



Hemos elegido tres capítulos que delimitan inicialmente los conceptos esenciales para entender transversalmente la disciplina de organizar cualquier tipo de recurso, construyendo una abstracción para los Sistemas de Organización que sea independiente de dominios de aplicación clásicos, en donde se proponen una serie de componentes, y en el que tienen especial importancia las actividades de descripción y representación de objetos que son, esencialmente, contenidos informativos, en donde se diferencia entre recursos primarios y recursos derivados o metadatos. La delimitación del resbaladizo concepto de recurso de información, simple, agregado, dinámico, multiforme, como objeto tangible sobre el que desarrollar las tareas de organización, ocupa otro de los capítulos que hemos recogido. La descripción de recursos pretende responder a las cuestiones complejas de identificar, denominar y enriquecer la información que hace a un recurso valioso dentro de una colección, lo conecta con otros en función de su contenido informativo y permite contextualizarlo como objeto social para la producción de valor y conocimiento.

Estos tres capítulos esperamos que sirvan de estímulo para próximas ediciones ampliadas, en la que se combinen los recursos e intereses de diferentes entidades vinculadas al sector de la información digital, los servicios de información, el patrimonio cultural y la formación especializada. La edición completa del libro, disponible en acceso abierto para su descarga completa y por capítulos en la web de la red de iSchools, se estructura en 12 capítulos, conforme al siguiente esquema:

FUNDAMENTOS DE LOS SISTEMAS DE ORGANIZACIÓN

.....



Este capítulo inicial corresponde con el capítulo 1 de la traducción incluida en esta primera edición parcial en español, en el que se enmarca la actividad de organización información y se definen sus componentes esenciales.

DECISIONES DE DISEÑO EN SISTEMAS DE ORGANIZACIÓN

.....

Se abordan seis cuestiones o dimensiones de amplio alcance que se entrelazan al definir un Sistema de Organización. Qué se organiza, por qué se organiza, cuánto se organiza, cuándo se organiza, cómo se organiza y dónde se organiza. Este marco para describir y comparar sistemas de organización permite entender desde patrones generales de diseño los sistemas usados en bibliotecas, archivos y museos, y generalizar, evitando así sus sesgos anclados en sus tradiciones específicas. Estos patrones pueden dar soporte a la colaboración



interdisciplinar, facilitando que se apliquen categorías y conocimientos de campos más afines en otros más lejanos del discurso sobre organización de recursos de información.

LAS ACTIVIDADES EN LOS SISTEMAS DE ORGANIZACIÓN

.....

Para tener una visión que reúna la forma en que organizamos cosas individualmente con la forma en la que lo hacemos desde bibliotecas, museos, administraciones, centros de investigación y empresas para crear Sistemas de Organización, requiere plantear unas generalizaciones de los conceptos y métodos de la organización tomados de cada uno de estos dominios. A partir de la revisión de una amplia variedad de Sistemas de Organización, se describen cuatro actividades y funciones que comparten todos ellos: seleccionar recursos, organizar recursos, diseñar interacciones y servicios basados en los recursos y mantener esos recursos a lo largo del tiempo.

LOS RECURSOS EN LA ORGANIZACIÓN DE SISTEMAS

.....



Este capítulo corresponde al segundo capítulo incluido en esta traducción, y aborda la pregunta ¿Qué es lo que se organiza en un Sistema de Organización? Se desarrolla el concepto de recurso, que supone una generalización de los diferentes tipos de contenidos que pueden integrarse en un sistema, tanto si son objetos físicos como digitales, estáticos o dinámicos. Se presentan en detalle las complejidades inherentes a identificar y delimitar recursos, dado que son aspectos que influyen en la capacidad de un sistema para dar soporte adecuado a las interacciones establecidas en su diseño.

DESCRIPCIÓN DE RECURSOS Y METADATOS

.....



Incorporado en este libro como capítulo 3, se abordan los principios por los que se organizan recursos y el tipo de interacciones que se podrán habilitar según las opciones de descripción asumidas explícitamente. El asunto central de la profundidad y extensión de la descripción se presentan en el marco de un proceso genérico para diseñar y realizar la descripción, combinando la intervención humana directa y los procesos automatizados.

DESCRIPCIÓN DE RELACIONES Y ESTRUCTURAS

.....

Un aspecto clave de la organización de colecciones de recursos es describir las relaciones que se establecen entre ellos. Se introduce un vocabulario específico para representar relaciones semánticas



tanto entre recursos entre sí, como entre recursos y los conceptos y términos usados en recursos descriptivos. También se debate sobre las relaciones estructurales en los recursos compuestos de partes, y las relaciones entre recursos y los recursos que se citan y con los que se establecen enlaces hipertextuales.

CATEGORIZACIÓN: DESCRIPCIÓN DE TIPOS Y CLASES

.....

Agrupaciones o conjuntos de recursos con descripciones similares o idénticas pueden tratarse como equivalentes, haciéndolos formar parte de una clase o categoría. Identificar y usar categorías es una de las actividades humanas esenciales que sucede automáticamente para las categorías perceptivas, tales como “cosas rojas” o “cosas redondas”. La categorización está profundamente integrada en el lenguaje y la cultura, y usamos categorías lingüísticas y culturales sin apenas darnos cuenta, aunque la categorización también puede ser un proceso cognitivo analítico muy profundo. En este capítulo se repasan las teorías de la categorización desde el punto de vista de cómo se usan y crean en los Sistemas de Organización.

CLASIFICACIÓN: ASIGNACIÓN DE RECURSOS A CATEGORÍAS

.....

A menudo se usa indistintamente categorización y clasificación, pero no se trata de lo mismo. La clasificación la entendemos como categorización aplicada – la asignación a los recursos dentro de un sistema de categorías, llamadas clases, usando una serie predeterminada de principios. Este capítulo debate el amplio abanico de formas en que las clasificaciones se usan en Sistemas de Organización. Se tratan las clasificaciones enumerativas, facetadas, basadas en actividades y la clasificación computacional. Dado que la clasificación y la normalización están estrechamente relacionadas, también se analizan los estándares existentes y cómo se construyen como parte de los Sistemas de Organización.

FORMATOS DE LA DESCRIPCIÓN DE RECURSOS

.....

Este capítulo complementa la perspectiva conceptual y metodológica de la creación de descripciones de recursos con la perspectiva de la implementación. Se revisan varios metamodelos para estructurar las descripciones, con particular énfasis en XML, JSON y RDF. Se cierra con la comparación entre los “mundos de la descripción” – procesamiento de documentos, la web y la web semántica— en los cada uno de estos tres metamodelos encajan mejor.



LA INTERACCIÓN CON LOS RECURSOS

Cuando varios Sistemas de Organización se solapan, intersectan o combinan, bien sea temporal o permanentemente, las diferencias en la descripción de los recursos puede hacer que resulte difícil o imposible localizarlos, accederlos o se impida usarlos. En este capítulo se revisan algunos conceptos y técnicas que desde diferentes dominios se aplican cuando se interactúa con recursos en Sistemas de Organización – integración, interoperabilidad, mapeo de datos, crosswalks, mash-ups, etc. Las interacciones se caracterizan por las capas de propiedades de los recursos que usan: instancias, colecciones, derivados o propiedades combinadas desde diferentes fuentes. Se desarrolla la idea de que un continuo entre organización de información y recuperación de información, y se describen las interacciones de recuperación de información (y otras) en términos de requisitos de organización de información (es decir, descripciones).

HOJA DE RUTA PARA SISTEMAS DE ORGANIZACIÓN

En este capítulo se complementa la perspectiva descriptiva expuesta en los capítulos 2 a 10, con un enfoque más prescriptivo que analiza las opciones de diseño y decisiones entre opciones enfrentadas (tradeoffs) que deben hacerse en las diferentes fases del ciclo de vida de un Sistema de Organización. Los modelos de ciclo de vida de sistemas presentan una gran variedad, pero optaremos por un modelo género en cuatro fases que diferencia la fase de identificación del dominio y alcance, la fase de requisitos, una fase de toma de requisitos y diseño, y una fase operacional.

ESTUDIOS DE CASO

Este último capítulo parte del modelo presentado en el capítulo previo para orientar el análisis de estudios de caso que expanden el alcance de los Sistemas de Organización, cuando se interpretan desde los principios, pautas, términos y modelos expuestos a lo largo del libro.



Finalizamos la presentación de esta obra recordando el principio formulado por Elaine Svenonius de que **“la efectividad de un sistema de acceso a la información es una función directa de la inteligencia puesta en organizarlo”**. Cuando observamos desde una perspectiva amplia los fenómenos involucrados en la información en sociedad,



podemos constatar que una vez que es puesto en movimiento un flujo de información, implica una importante cantidad de recursos y agentes, para su producción, para conseguir un impacto de difusión suficiente, y para mantenerse accesible a lo largo del tiempo. El ciclo de vida de la información supone también grandes ineficiencias, suplidas para ciertos usos hoy en día por monstruosos sistemas de procesamiento de contenidos a gran escala. La falta de una mejor “inteligencia incorporada”, es decir, metadatos y estructuras de información óptimas para los ecosistemas y plataformas digitales, supone un coste social, materializado en sobrecostes, escasez de acceso intelectual al contenido de recursos de información publicados, perduración de restricciones técnicas evitables, infrautilización de recursos y todo tipo de ineficiencias que sugieren la importancia de entender la organización de información como un pegamento para aportar valor tanto a la industria de los contenidos, el negocio digital, la difusión de la cultura y los recursos educativos a todos los niveles. El reto colectivo de que la información sea fuente de valor para sociedades complejas y el conocimiento un recurso distribuido con equidad y sostenibilidad afecta a muchas disciplinas académicas, perfiles profesionales tradicionales y de nuevo cuño, y actores políticos e institucionales. Tender puentes entre dominios creando un lenguaje común que permita trabajar juntos a las ingenierías tecnológicas, las ciencias métricas, las ciencias sociales y las ciencias humanas, es el objetivo de este libro cuando sea leído y puesto en práctica por aquellos actores de la sociedad de la información que están interviniendo en que todo tipo de actividades de información multipliquen sus efectos para generar conocimientos, conexiones y sentido.

Febrero, 2021

Tomás Saorín

Grado en Gestión de información y Contenidos Digitales
Universidad de Murcia

Eva Méndez

Grado en Gestión de información y Contenidos Digitales
Universidad Carlos III

Mario Pérez-Montoro

Grado en Gestión de Información y Documentación Digital
Universidad de Barcelona



CAPÍTULO 1

SISTEMAS DE ORGANIZACIÓN, MARCO DE REFERENCIA

Robert J. Glushko

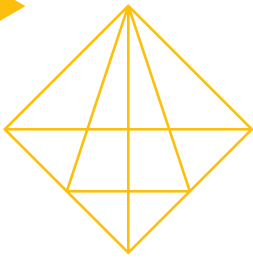
CONTENIDO

- | | | |
|--|----|---|
| | 16 | 1.1 Organizar, una disciplina global |
| | 26 | 1.2 El concepto de “Sistema de Organización” |
| | 29 | 1.3 El concepto de “Recurso” |
| | 32 | 1.4 El concepto de “Colección” |
| | 35 | 1.5 El concepto de la “Ordenación Intencionada” |
| | 39 | 1.6 El concepto de “Principio de Organización” |
| | 46 | 1.7 El concepto de “Agente” |
| | 47 | 1.8 El concepto de “Interacciones” |
| | 49 | 1.9 El concepto de “Recurso de Interacción” |

Organizar es crear capacidades al imponer de forma intencionada un orden y una estructura.



1.1 ORGANIZAR, UNA DISCIPLINA GLOBAL



Organizar es una actividad tan común que a menudo la llevamos a cabo sin ser conscientes de ello. Organizamos zapatos en el armario, libros en las estanterías, especias en la cocina, recibos y registros en carpetas para preparar la declaración de impuestos, y personas en proyectos de negocios y equipos deportivos. Algunos tenemos trabajos que implican hacer unas tareas concretas de organización. Incluso podríamos haber recibido una formación específica para realizarlas siguiendo prácticas disciplinarias especializadas. Puede que dominemos estas tareas, pero, aun así, a menudo no nos paramos a reflexionar sobre la similitud entre las tareas de organización que hacemos nosotros mismos y las que hacen otros, o entre las que hacemos en el trabajo y las que hacemos en casa. Damos por sentados los conceptos y los métodos utilizados en el Sistema de Organización con el que más trabajamos.

El objetivo de este libro es ayudar a los lectores a ser más conscientes de lo que significa organizar cualquier tipo de recursos y de los principios por los que estos se organizan. En concreto, este libro presenta el concepto de “Sistema de Organización”: una colección organizada de forma intencionada que se compone de recursos y de las interacciones que estos apoyan. En este libro también se analizan las decisiones de diseño que se toman en cualquier organización sistemática de recursos y los patrones de diseño para las interacciones que hacen uso de los recursos, como se muestra a continuación:

ORGANIZAMOS COSAS FÍSICAS. Cada uno de nosotros organiza muchos tipos de cosas a lo largo de su vida: libros en estanterías; registros financieros en carpetas y archivadores; ropa en vestidores y armarios; utensilios de cocina y cubiertos en los cajones y los armarios de la cocina. Las bibliotecas públicas organizan libros impresos, periódicos, mapas, CDs, DVDs, y tal vez algunos álbumes de viejos discos de vinilo. Las bibliotecas de investigación también organizan manuscritos valiosos, panfletos, partituras y muchos otros tipos de información impresa. Los museos organizan pinturas, esculturas y otros artefactos de valor cultural, histórico o científico. Las tiendas y los proveedores organizan sus bienes para venderlos a los consumidores y a otros proveedores. Las ligas deportivas organizan a los jugadores en equipos y los equipos organizan a los jugadores por posición o rol.



ORGANIZAMOS INFORMACIÓN SOBRE COSAS FÍSICAS. Cada uno de nosotros organiza información sobre cosas: cuando hacemos un inventario de nuestra casa para el seguro, cuando vendemos lo



que ya no necesitamos en eBay, o cuando dejamos nuestra valoración de un restaurante en Yelp. Los catálogos de fichas de una biblioteca, y sus sustitutos digitales, nos dicen qué libros contiene la colección de la biblioteca y dónde encontrarlos. Los sensores y las etiquetas RFID rastrean el movimiento de los bienes (incluso de los libros de las bibliotecas) a través de cadenas de suministro, además del movimiento (o falta de él) de los coches en las autopistas.



ORGANIZAMOS COSAS DIGITALES. Cada uno de nosotros organiza su información digital (correo electrónico, documentos, libros electrónicos, archivos MP3 y de vídeo, citas y contactos) en ordenadores, smartphones, lectores de libros electrónicos o “la nube”, a través de servicios de información que utilizan los protocolos de Internet. Las grandes bibliotecas de investigación organizan revistas y libros digitales, programas de ordenador, conjuntos de datos (datasets) gubernamentales y científicos, bases de datos, y muchos otros tipos de información digital. Las empresas organizan sus registros comerciales y la información de sus clientes en aplicaciones empresariales, sistemas de almacenamiento Data Warehouse y bases de datos. Los hospitales y las clínicas médicas mantienen e intercambian registros electrónicos de salud y radiografías y escaneos digitales.



ORGANIZAMOS INFORMACIÓN SOBRE COSAS DIGITALES. Los catálogos de las bibliotecas digitales, los portales web y los sitios web de agregación organizan enlaces a otros recursos digitales. Los motores de búsqueda web utilizan el contenido y el análisis de los enlaces junto con clasificaciones de relevancia para organizar los miles de millones de páginas web y “feeds” de actualización permanente de datos, que compiten por nuestra atención. Los servicios basados en la web y otros recursos de información pueden ser interconectados y coreografiados para llevar a cabo procesos empresariales que manejan grandes cantidades de información, o agregados y analizados para predecir y personalizar servicios de información.

Observemos más de cerca estos cuatro tipos o contextos de organización. Contraponemos “organizar cosas” con “organizar información”. A primera vista podría parecer que la organización de cosas físicas como libros, CDs, piezas de máquinas o utensilios de cocina tiene un carácter completamente diferente al de la organización de cosas digitales intangibles. A menudo organizamos las cosas físicas de acuerdo con su forma, tamaño, material de fabricación u otras propiedades intrínsecas y visibles: por ejemplo, podríamos organizar nuestras camisas en el armario por estilo y color, y podríamos organizar nuestra colección de



música separando los viejos álbumes de vinilo de los CDs. Podríamos organizar los libros en las estanterías por su tamaño, poniendo todos los grandes y pesados libros ilustrados en el estante inferior. La organización de ropa y recursos de información en formatos tangibles que se basa en propiedades visibles no se parece mucho a la forma en que se almacenan y organizan los libros digitales en Kindle o la música digital en el reproductor de música. Organizar, almacenar y acceder a radiografías impresas en película puede parecer que tiene poco en común con estas actividades cuando la radiografía tiene formato de imagen digital.

No es de extrañar que la organización de cosas y la organización de información no difieran demasiado cuando la información se presenta de manera tangible. La era de la información digital ubicua de la última o de las últimas dos décadas es sólo algo puntual en el tiempo en comparación con los más de diez mil años de experiencia que tiene el ser humano con la información tallada en piedra, grabada en arcilla o impresa con tinta en papiro, pergamino o papel. Estos recursos de información tangibles han introducido profundamente la noción de la información como algo físico en la cultura, el lenguaje y las formas de diseño y organización de la información. Este enfoque hacia los recursos de información tangibles es especialmente dominante en las colecciones de libros venerados por su distintiva encuadernación, caligrafía y composición tipográfica.

Sin embargo, en otras ocasiones hay diferencias sustanciales entre la forma en que organizamos cosas y la forma en que organizamos información, incluso cuando ésta se presenta en forma física. Tendemos más a organizar nuestros “objetos de información” de acuerdo con **aquello de lo que tratan** y no tanto en base a sus propiedades visibles. En casa clasificamos nuestros CDs por artista o género; mantenemos los libros de cocina separados de los libros de viaje, y los libros de ficción separados de los libros de referencia. Las bibliotecas emplean esquemas de clasificación por temas que tienen unos cientos de miles de categorías distintas.

Asimismo, a veces prestamos poca atención a las propiedades visibles de las cosas tangibles cuando las organizamos y en su lugar las ordenamos de acuerdo con sus propiedades funcionales o prácticas. Guardamos destornilladores, alicates, martillos, sierras, taladros y niveles en una caja de herramientas o juntos en un banco de trabajo, a pesar de que tienen pocas propiedades visuales en común. No los organizamos por lo que vemos en ellos, sino por **lo que sabemos acerca de su uso**. La organización de las herramientas basada en las tareas para las que sirven tiene cierta similitud con la basada en temas o la organización de una biblioteca.

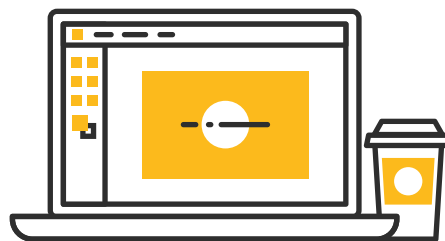


También diferenciamos entre “organizar cosas” y “organizar información sobre cosas”. Esta diferencia parece clara si consideramos el catálogo tradicional de la biblioteca, cuyas tarjetas impresas describen los libros en las estanterías. Cuando las cosas y la información sobre ellas están ambas en formato físico, es fácil ver que la primera es un recurso primario y la segunda un recurso sustituto o asociado que lo describe o se relaciona con él.

¿QUÉ ES LA INFORMACIÓN?

La mayoría de los cientos de definiciones de información la tratan como una idea que da vueltas alrededor de términos igualmente difíciles de definir, como “dato”, “conocimiento” y “comunicación”. Además, estas perspectivas intelectuales e ideológicas sobre la información coexisten con el más mundano uso del término, como cuando preguntamos en la estación “¿Me podría facilitar información sobre el horario de trenes?”.

Un punto de vista abstracto de la información como algo intangible está en los fundamentos intelectuales tanto de la moderna ciencia de la información como de la economía y sociedad de la información. Sin embargo, la perspectiva abstracta a menudo entra en conflicto con la mucho más antigua idea de la información como algo tangible que de forma natural aflora cuando la información se codifica inextricablemente en formatos físicos. A menudo mezclamos confusamente la “información como contenido” con el sentido de “información como contenedor” y también, con demasiada facilidad, tratamos la cantidad de bits almacenados en la memoria de un ordenador o en la nube como una medida de contenido o valor.



Cuando se trata de “organizar información sobre cosas digitales” el contraste es mucho menos claro. Cuando se busca un libro mediante un motor de búsqueda primero se obtiene la descripción del libro que hay en el catálogo y a menudo el libro en sí está a sólo un clic de distancia. Cuando las cosas y la información sobre ellas son ambas digitales



la oposición que planteamos no es tan evidente como cuando una o ambas están en un formato físico. Y aunque usamos las radiografías (en película o en formato digital) como ejemplos de cosas que podríamos organizar, cuando un médico estudia una radiografía, ¿no la está usando como si fuera información sobre el tema, es decir, el paciente de la radiografía? Y cuando el gerente de una empresa toma decisiones de marketing y precios a través del análisis de la información digital sobre qué y cuándo compra la gente, podemos pensar en esto como organizar a los clientes en categorías, o como una forma de organizar la información del cliente.

Estas diferencias y relaciones entre “cosas físicas” y “cosas digitales” han sido discutidas y debatidas durante mucho tiempo por filósofos, lingüistas, psicólogos y otros. (Ver los recuadros: *¿Qué es la información?* y *La diferencia entre datos e información*)

Las distinciones entre la organización de cosas físicas, cosas digitales o información sobre cosas físicas o digitales son difíciles de describir porque muchas de las palabras que usamos están tan sobrecargadas de significados diferentes como la propia “información”. Por ejemplo, la perspectiva de la biblioteconomía suele incluir las propiedades de presentación o implementación en las definiciones de “documento”. Utiliza el término para referirse únicamente a las formas físicas tradicionales. En cambio, la perspectiva de la informática adopta una visión abstracta de “documento” para referirse a cualquier unidad de información autónoma. Separa el contenido de un documento de su presentación o contenedor.

LA DIFERENCIA ENTRE DATOS E INFORMACIÓN

El lector astuto habrá notado que incluimos los datos de los sensores como “información sobre cosas físicas” y los feeds como “información sobre cosas digitales”. Muchos libros de texto en los campos de la ciencia de la información y la gestión del conocimiento distinguen los datos y la información de manera más precisa. Para ellos, los datos se encuentran en la parte inferior de una Jerarquía de la Información, Pirámide del Conocimiento, o Jerarquía DIKW en la que los datos se transforman en Información, que se transforma en Conocimiento, que luego se transforma en Sabiduría.

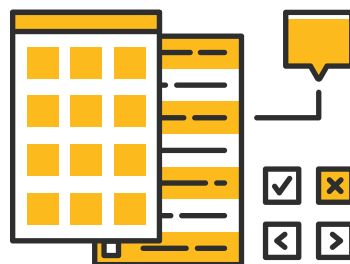
En este ámbito, los datos son observaciones brutas o elementales sobre propiedades de los objetos, los eventos y su entorno. Los datos se convierten en información cuando son agregados,



procesados, analizados, formateados y organizados para añadir significado y contexto de manera que puedan ser utilizados para responder a preguntas. Este proceso puede incluir operaciones de cálculo, inferencia o refinamiento de los datos. Por ejemplo, las mediciones de temperatura, precipitación y velocidad del viento son datos. Cuando se combinan y se resumen, un conjunto de datos se convierte en información estadística sobre el tiempo en un día concreto. Cuando se recogen durante un período de meses o años, estos conjuntos de datos se convierten en información sobre el clima del lugar donde se han recogido.

Este libro sobre “la disciplina de organizar” no hace esta fuerte separación entre los datos y la información como en la Jerarquía/Pirámide convencional. Es probable que las personas que lean este libro sean aspirantes o profesionales en ejercicio en sectores donde la información y los datos se tratan a menudo como sinónimos para referirse al contenido de una base de datos o una aplicación de gestión de datos. La diferencia entre datos e información puede ser útil en teoría, pero no en estos contextos concretos.

La diferencia entre datos e información también se está desdibujando debido a la ampliación de la definición de los datos en el nuevo campo profesional de la ciencia de datos. De hecho, un popular texto introductorio elimina por completo la información de la Jerarquía/Pirámide con su título: *Descubriendo el conocimiento en los datos: una introducción a la minería de datos*.



Una variación similar de la definición ocurre con el concepto de “autor” o “creador”. Cuando decimos que “Herman Melville es el autor de Moby Dick” (Melville 1851) el significado de “autor” no depende de si tenemos una copia impresa o un e-book en mente, sino de lo que cuenta como autoría, que varía mucho entre las disciplinas académicas. Además, las diferentes normas para describir los recursos no coinciden en la precisión con la que identifican a la(s) persona(s) u organización(es) responsable(s) principal(es) de crear el contenido intelectual del re-



curso. Las personas que se toman en serio la descripción de la música critican con razón los servicios de *streaming* y las tiendas online que sólo tienen un campo de “artista” porque no distinguen al compositor, al director, a la orquesta y a otras personas con funciones distintas en la creación de la música.

Si permitimos que el concepto de información sea cualquier cosa que podamos estudiar, “cualquier cosa que informe”, el concepto se vuelve ilimitado. Nuestro objetivo en este libro es salvar el abismo intelectual que separa a las muchas disciplinas que comparten el objetivo de organizar, pero que difieren en lo que organizan. Esto requiere que nos enfoquemos en situaciones en las que la información existe debido a actos intencionales para crearla u organizarla. (Ver el recuadro sobre *La disciplina de organizar*).

Muchos de los temas fundacionales de una disciplina sobre la organización se han presentado tradicionalmente desde la perspectiva del sector bibliotecario y se han enseñado como “biblioteconomía y ciencias de la información”. Estos incluyen la descripción bibliográfica, la clasificación, la denominación, el control de autoridades, la conservación y la normalización. En las últimas décadas estos cimientos se han reconstruido y ampliado desde los campos de la informática, las ciencias cognitivas y otros nuevos campos incluyendo cada vez más sectores privados y contextos no bibliográficos, medios multimedia y sociales, y nuevas aplicaciones y sistemas de servicios que manejan una gran cantidad de información, gracias a la computación móvil, omnipresente y científica. Las últimas adiciones a la Disciplina de Organizar provienen de la ciencia de datos y el aprendizaje automático e introducen consideraciones de velocidad y escala que surgen cuando se aprovecha la potencia computacional masiva y las nuevas técnicas estadísticas para organizar y actuar sobre la información.

Estos nuevos métodos y herramientas de la ciencia de datos y el aprendizaje automático nos permiten organizar más información, hacerlo más rápido y hacer predicciones basadas en lo que la gente ha buscado, comprado o dicho. Sin embargo, esta no es la primera vez que las nuevas ideas y tecnologías desafían la forma en que la gente se organiza e interactúa con los recursos. Hace 50 años, los catálogos online con capacidad de búsqueda cambiaron radicalmente la forma en que la gente utilizaba las bibliotecas. La web, inventada hace menos de treinta años para que los científicos pudieran compartir informes técnicos, es ahora una parte esencial de muchas actividades humanas. Es importante no ver las últimas novedades como algo que lo cambia todo, porque seguirán llegando cosas nuevas, y estos avances tecnológicos todavía dependen y complementan el trabajo de organización



realizado por personas. La ciencia de datos no reemplazará a los organizadores humanos, al menos no más de lo que los han reemplazado otras ciencias. (Ver los recuadros *La ciencia de datos* y *La disciplina de la organizar*).

LA DISCIPLINA DE ORGANIZAR

Una disciplina es un campo de estudio integrado en el que existe un acuerdo sobre las cuestiones y los problemas que deben estudiarse, cómo están interrelacionados, cómo deben estudiarse y cómo deben evaluarse las conclusiones o las teorías sobre tales cuestiones y problemas. Un marco es un conjunto de conceptos que proporciona la estructura básica para comprender un dominio, lo que aporta un vocabulario común para diferentes teorías explicativas.

La organización, las actividades de organizar, es una cuestión fundamental en muchas disciplinas, sobre todo en la biblioteconomía y las ciencias de la información, la informática, el análisis de sistemas, las ciencias de la computación, el derecho, la economía y los negocios. Sin embargo, estas disciplinas sólo llegan a un acuerdo limitado en cuanto a la forma en que abordan y resuelven los problemas sobre organización. Por ejemplo, tradicionalmente la biblioteconomía y las ciencias de la información han estudiado la organización desde una perspectiva bibliográfica dentro del sector público, prestando gran atención a las necesidades de los usuarios en materia de acceso y conservación, y ofreciendo métodos y soluciones prescriptivos. En cambio, la informática y las ciencias de la computación tienden a estudiar la organización en el contexto de las aplicaciones comerciales que gestionan gran cantidad de información, centrándose en la eficiencia, la implementación y la arquitectura de estos sistemas. Las disciplinas de la gestión y la organización industrial se ocupan de la organización de los recursos humanos, materiales y de información en contextos constituidos por fuerzas comerciales, competitivas y reguladoras.

Este libro presenta un marco más abstracto para las cuestiones y los problemas de organización, y enfatiza los conceptos y los objetivos comunes de las disciplinas que las estudian. A su vez, sostiene que cada sistema de organización implica una colección de recursos y podemos tratar las cosas físicas, las digitales y la información sobre tales cosas como recursos. Cada sistema de organización implica una elección de propiedades o principios usados para describir y organizar los recursos, y formas de respaldar las interacciones con los recursos. Al comparar y contrastar cómo se desarrollan estas ac-



tividades en diferentes contextos y dominios, podemos identificar patrones de organización y ver que los sistemas de organización a menudo siguen un ciclo de vida común. Podemos crear una disciplina de la organización de manera disciplinada.

Por esto es necesario adoptar un punto de vista transdisciplinario que nos permita destacar lo que las diferentes disciplinas tienen en común y que se centre en cómo encajan entre sí, en lugar de en qué se distinguen. La selección de recursos, la organización, el diseño de interacción y el mantenimiento se estudian en todas las disciplinas, pero estos conceptos poseen nombres diferentes. Un vocabulario para discutir los retos organizativos comunes y las demandas que de otro modo podrían quedar oscurecidas por perspectivas disciplinarias estrechas, nos ayudará a comprender mejor los sistemas de organización existentes y a la vez sugerirá cómo inventar otros nuevos a través de diferentes elecciones de diseño.

LA CIENCIA DE DATOS Y LA DISCIPLINA DE ORGANIZAR

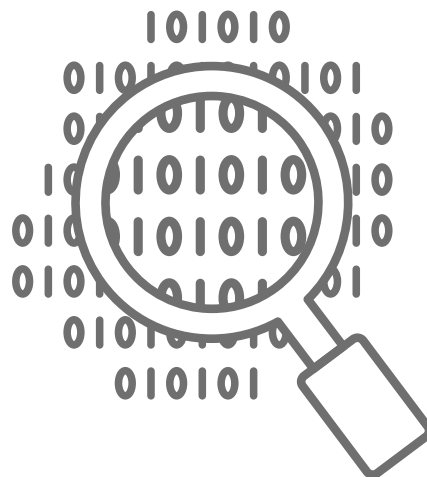
Los avances en la potencia de los ordenadores y las técnicas estadísticas están haciendo posible identificar patrones en los datos y la extracción de información significativa a un nivel sin precedentes. Muchos libros y artículos sobre ciencia de datos, aprendizaje automático y análisis predictivo pronostican audazmente que estos campos emergentes cambiarán el mundo de forma radical. Estas afirmaciones son a la vez provocativas y prometedoras, pero en el fondo, la ciencia de datos trata sobre la forma en que los recursos se seleccionan, describen y organizan; conceptos con una gran tradición en la informática y la biblioteconomía. En lugar de organizar y describir los libros de una biblioteca o los productos de un almacén, un científico de datos podría organizar información sobre los libros o los productos en tablas de datos masivos, de forma que cada recurso aparece como una fila y sus propiedades descriptivas como las columnas. Después de organizar los libros o los productos en categorías, las técnicas de aprendizaje automático pueden clasificar los nuevos libros o productos mediante esas categorías, o tal vez descubrir nuevas categorías basadas en los comportamientos de acceso o compra. Así pues, aunque las técnicas de la ciencia de datos son nuevas, muchos de sus desafíos no lo son; los científicos de datos necesitan seleccionar los recursos con sabiduría y decidir la mejor manera de describirlos; necesitan comprender que la descripción y la categorización de los recursos pueden ser sesgadas; necesitan comprender los acuerdos que llevan al equilibrio entre personas y ordenadores; y, necesitan com-

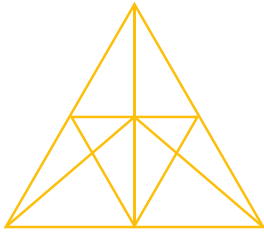


probar los descubrimientos que se logran con los algoritmos en experimentos controlados.

Para dar sentido a las cuestiones sobre la ciencia de datos, hay que entender la diferencia entre “tipo” y “grado”. Hace cien años, la velocidad media a la que un coche podía circular por una autopista era de unos 65 kilómetros por hora. Los coches de hoy en día viajan el doble de rápido. Se trata solo de un cambio de grado. Sin embargo, un aumento de la velocidad a unos 17.500 kilómetros por hora logra una “velocidad orbital” que nos permite entrar en la órbita de la Tierra en el espacio, un viaje que es diferente en su tipo.

Pero, ¿cómo se aplica esto a las ciencias relacionadas con datos? Algunas ciencias de datos implican colecciones de datos que son “altas”, es decir, que contienen muchos millones o incluso miles de millones de registros, cada uno con un número relativamente pequeño de variables. Ser capaz de analizar los datos “altos” con una rapidez nunca antes vista es sobre todo un cambio de grado en comparación con las técnicas tradicionales de las bases de datos. No obstante, en el caso de las colecciones de datos que son “anchas”, en las que cada registro puede contener cientos o miles de variables, las técnicas de la ciencia de los datos podrían permitirnos ver patrones que no se podrían ver en absoluto, o que no se podrían ver de forma asequible y en cantidad suficiente. En este caso, la ciencia de datos podría estar produciendo cambios de tipo.





1.2 EL CONCEPTO DE

“SISTEMA DE ORGANIZACIÓN”

Proponemos unificar las variadas perspectivas sobre la organización y la información con el concepto de “Sistema de Organización”: **una colección de recursos organizada intencionalmente y las interacciones que respaldan**. Esta definición reúne varias ideas esenciales que presentaremos brevemente en este capítulo y que se desarrollan en detalle a lo largo de la edición completa del libro “The discipline of organizing”¹ y los capítulos recogidos en esta edición parcial traducida.



En la imagen 1.1 se representa el modelo conceptual de un Sistema de Organización que muestra recursos dispuestos de forma intencional, sus interacciones (marcadas por diferentes tipos de flechas) y los agentes humanos e informáticos que interactúan con los recursos en diferentes contextos.

Un Sistema de Organización es una caracterización abstracta de la forma en que se describe y organiza un conjunto de recursos para permitir que los agentes humanos o informáticos interactúen con ellos. El Sistema de Organización es una visión arquitectónica y conceptual que

¹ Edición completa disponible online en inglés, a la que nos referiremos como “TDO, core concepts edition”, descargable en la web de la red de iSchools <https://ischools.org/Discipline-of-Organizing>



se distingue de la disposición física de los recursos que lo pueden encarnar, y también se distingue de la persona, empresa o institución que lo implementa y opera. Estas distinciones son a veces difíciles de transmitir en el lenguaje ordinario; por ejemplo, podríamos considerar algún conjunto de descripciones de recursos, principios de organización e interacciones de apoyo como un “Sistema de Organización Biblioteca”. Sin embargo, a veces también necesitamos referirnos a una “biblioteca” como la institución en la que opera este Sistema de Organización y por supuesto la idea de una “biblioteca” como instalación física está profundamente arraigada en el lenguaje y la cultura.

Nuestro concepto de Sistema de Organización se inspira en parte en los conceptos propuestos por **Elaine Svenonius** en el año 2000 para el dominio bibliográfico y que se publicaron bajo el título de *The Intellectual Foundation of Information Organization*. Ella reconoció que las actividades tradicionales de organización de la información para la descripción bibliográfica y la catalogación se complementaban y, en parte, se compensaban con el procesamiento automatizado de textos y la indización, que solían tratarse como parte de una disciplina diferente denominada recuperación de la información. Svenonius sugirió que **las decisiones sobre la organización de la información y la recuperación de la información estaban intrínsecamente vinculadas por un principio de compensación** (tradeoff) y, por lo tanto, debían considerarse como un sistema interconectado: “La eficacia de un sistema para acceder a la información es una función directa de la inteligencia puesta en la organización de la misma” (p.ix)².

Una visión sistémica de la organización y la recuperación de la información capta y proporciona una estructura para las compensaciones inherentes que quedan ocultas en los “silos” derivados de las perspectivas tradicionales de cada disciplina y diferentes categorías: cuanto más esfuerzo se ponga en la organización de la información “a la entrada” de un sistema, cuando se crea o se añade información a una colección, más fácil se podrá recuperar de forma eficaz, y cuanto más esfuerzo se ponga en la recuperación de información “a la salida” de ese sistema, ya no será tan necesaria tanta organización previa. A veces una colección de recursos está muy organizada, pero debido a que fue organizada para propósitos diferentes a los que ahora se contemplan, necesitaremos reorganizarla “a la entrada”. Esto es muy común en el caso de los textos o conjuntos de datos digitales, en los que los recursos previamente organizados o sus descripciones pueden ser clasificados, convertidos a otro formato o idioma, combinados, resumidos

² Con el fin de rendirle homenaje y construir sobre sus ideas, cada una de las subpartes del capítulo 2 “Design Decisions in Organizing Systems” del libro “TDO core concept edition” comienza con una cita de su libro.



o transformados de cualquier otra manera para que encajen en un nuevo Sistema de Organización. Por ejemplo, para comprender las pautas de compra estacionales, un minorista podría combinar los datos de las compras con los datos meteorológicos y los datos del calendario sobre los acontecimientos deportivos más comunes (porque el mal tiempo y los deportes de radiodifusión hacen que la gente se quede en casa), y los tres conjuntos de datos tendrían que representar los datos de “hora” y la “ubicación” de la misma manera.

La visión de sistemas ya no contrapone la organización de la información como actividad humana y la recuperación de información como actividad de las máquinas, o la organización de la información como un asunto para las ciencias de la información y las bibliotecas, mientras que la recuperación de información corresponde a la informática. En su lugar, apreciamos de manera natural que hoy día los ordenadores ayudan a la gente a organizarse y que la gente contribuye con gran parte de la información que utilizan los ordenadores cuando analizan y organizan los recursos. Por ejemplo, muchos algoritmos para la clasificación de las comunicaciones utilizan enfoques de aprendizaje supervisado que comienzan con elementos clasificados por personas.

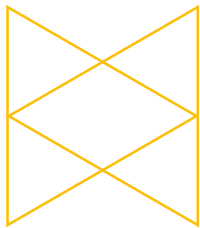


Por último, se puede aplicar una visión sistémica a Sistemas de Organización para cualquier tipo de recurso, lo que permite un debate más matizado sobre la forma en que se reparten los costes y beneficios económicos, sociales y cognitivos de las actividades de organización entre los diferentes agentes implicados, interesados o afectados (*stakeholders*). Todo Sistema de Organización está sesgado por las perspectivas y experiencias de las personas que lo crean. Algunos de esos sesgos son ineludibles, como una especie de automatismo organizativo, porque reflejan las capacidades perceptivas y cognitivas innatas de los seres humanos. Nuestras mentes imponen estructuras y encuentran patrones, incluso cuando no hay ninguno, y no somos capaces de actuar de forma perfectamente racional, así que simplifi-



camos sin ser del todo conscientes de ello. Tampoco se nos da muy bien pensar en las posibilidades futuras y revisar las expectativas ante nuevas evidencias, y esta inercia mental nos hace preservar recursos e interacciones en los Sistemas de Organización aun cuando ya no son necesarios. Otros sesgos en los Sistemas Organizativos reflejan elecciones más intencionales que, ya sea de forma implícita o explícita, suponen ganadores o perdedores, porque tratan algunas interacciones como preferidas mientras desapruaban otras, o de otra manera imponen o superponen un conjunto de valores a las partes interesadas del sistema. Por ejemplo, muchos Sistemas de Organización distribuyen a las personas en grupos o colas para que las interacciones sean más eficientes, pero cuando una aerolínea da prioridad de embarque a los clientes que pagaron más por sus billetes, otros pasajeros pueden pensar que no es justo que tengan que embarcar después.

1.3 EL CONCEPTO DE “RECURSO”



El significado más extendido del término “Recurso” es el de cualquier cosa de valor que pueda apoyar una actividad orientada a un objetivo. Esta definición significa que un recurso puede ser una cosa física, una cosa no física, información sobre cosas físicas, información sobre cosas no físicas, o cualquier cosa que se quiera organizar. Otras palabras que pretenden englobar lo mismo son “entidad”, “objeto”, “artículo” o “instancia”. La palabra “documento” se utiliza a menudo para un recurso de información en formato digital o físico; “artefacto” se refiere a recursos creados por personas, y “activo” a recursos con valor económico.

“Recurso” tiene un significado especializado en la arquitectura de Internet. Es habitual describir las páginas web, las imágenes, los vídeos, etc. como recursos, y el protocolo para acceder a ellos, el Protocolo de Transferencia de Hipertexto (HTTP), utiliza el Identificador Uniforme de Recursos (URI).

Tratar como un **recurso primario cualquier cosa que pueda ser identificada** supone una generalización importante del concepto porque permite que los servicios basados en la web, las fuentes de datos, los objetos con etiquetas RFID, los sensores u otros dispositivos inteligentes, o los agentes computacionales puedan formar parte de sistemas de organización.

En lugar de enfatizar las diferencias entre los recursos tangibles e intangibles, consideramos que es esencial **determinar si el contenido**



Entradas de conciertos

Las entradas de un concierto son artefactos que expresan metadatos relacionados con eventos; incluida la hora, lugar y el número de butaca; los artistas que actúan. Para quienes han asistido a un concierto, la entrada ofrece la promesa de todo ello, así como un recuerdo de otras muchas más cosas inefables. (Foto Murray Malone)

Entradas de conciertos

Una entrada de concierto es un vehículo para transmitir un paquete de declaraciones relativas a un evento, por tanto, es un recurso descriptivo, como una ficha del catálogo de una biblioteca. Una entrada de concierto también es un recurso por derecho propio con un valor intrínseco; puede ser comprada y vendida, a veces a mayor precio de lo que indica su recurso descriptivo. Una entrada es un permiso para usar una butaca en un lugar para un propósito especial a una hora determinada; tras el evento, la entrada pierde su valor intrínseco pero puede adquirir un valor extrínseco como un artefacto en una colección como la del recuadro anterior.

del recurso tangible es información (si se necesita entender y procesar de lo que trata o si representa o se refiere a algún recurso de otro tipo, en lugar de ser tratado como una cosa en sí misma). Tanto en un libro impreso como digital, nos fijamos en su contenido informativo, de qué trata; sus propiedades tangibles pasan a ser secundarias. Por el contrario, en el caso de las perchas del armario y las jarras de cocina que usamos para medir cantidades no son más que sus obvias características utilitarias las que hacen que sus propiedades tangibles sean más importantes. (Por supuesto, no hay un límite claro aquí; puedes comprar “las perchas de moda” como un artículo de decoración, y la vieja jarra de medir podría ser un recuerdo familiar porque perteneció a tu abuela).



Muchos de los recursos de los Sistemas de Organización son **recursos descriptivos o recursos sustitutivos** que describen los recursos primarios; las entradas del catálogo de una biblioteca o la lista de resultados de búsqueda en la web son ejemplos conocidos. En los museos, la información sobre la producción, la cobertura o la historia de la propiedad de un recurso puede ser más importante que el propio recurso en sí; unos pocos fragmentos de cerámica tienen poco valor sin estos recursos de información asociados. Del mismo modo, los datos comerciales o científicos a menudo no pueden ser subestimados o analizados sin información adicional sobre la manera en que fueron recopilados. La mayoría de las empresas basadas en la web explotan los datos sobre la forma en que los usuarios interactúan con los recursos, como los archivos de log que registran cada búsqueda en la web, cada enlace en el que se hace clic y cada página web que se visita.

Los recursos que describen, o que están asociados con otros recursos, se denominan a veces “metadatos”. Sin embargo, cuando se examinan más ampliamente los Sistemas de Organización, suele ser difícil distinguir entre el recurso que se describe y cualquier descripción del mismo o asociada a él. Cuando las descripciones se incorporan a los recursos, como con frecuencia sucede con los metadatos (en la portada de un libro, en la cabecera de un periódico o en el código fuente de las páginas web), decidir qué recursos consideramos primarios es un desafío de diseño y suele tener algo de arbitrario.

También se nos presenta otro reto cuando lo que sirve como metadatos para una persona o proceso puede funcionar como un recurso primario o como datos para otra. En lugar de ser una distinción inherente, la diferencia entre los recursos primarios y los asociados suele ser sólo una decisión sobre en qué recurso nos estamos centrando en una determinada situación. Un espécimen de animal en un museo de historia natural puede ser un recurso primario para los visitantes del museo y los científicos interesados en la anatomía, pero la información sobre el lugar donde se recogió el espécimen es el recurso primario para los investigadores interesados en la ecología o la migración.

Los Sistemas de Organización pueden referirse a las personas como recursos y a menudo utilizamos ese término para evitar especificar el género o el papel específico de un empleado o trabajador, dentro del ámbito de gestión del departamento de “recursos humanos” en un entorno laboral. Una empresa se define por su disposición intencional de recursos humanos, y hay tanto variedad como regularidad en estas disposiciones³.

3 Puede consultarse el recuadro complementario “Business Structures” en el



Los recursos humanos en los Sistemas de Organización pueden entenderse de manera muy parecida a los recursos físicos o digitales inanimados: se seleccionan, se organizan y se gestionan, y pueden crear valor individualmente o mediante sus interacciones con otros dentro y fuera del sistema. Sin embargo, los seres humanos son recursos singularmente complicados, y cualquier Sistema de Organización que los utilice debe tener en cuenta sus derechos, motivaciones y relaciones⁴.

1.4 EL CONCEPTO DE “COLECCIÓN”

.....



Una colección es un grupo de recursos que han sido seleccionados con algún propósito. Algunos términos similares son conjunto (matemáticas), agregación (modelado de datos), conjunto de datos (ciencia y negocios) y corpus (lingüística y análisis literario).

Preferimos el término colección porque tiene menos significados especializados. La palabra “colección” suele usarse para describir conjuntos personales de recursos físicos (mi colección de sellos o de discos de vinilo) así como de los digitales (mi colección de música digital). Distinguimos las bibliotecas de derecho de las bibliotecas de software, los sistemas de gestión del conocimiento de los sistemas de almacenamiento de datos (Datawarehouse) y una colección particular de sellos de una colección de monedas, principalmente, porque contienen **diferentes tipos de recursos**. Del mismo modo, distinguimos las colecciones de documentos por el tipo del recurso, diferenciando los documentos de tipo narrativo como novelas y biografías de otros documentos transaccionales como catálogos y facturas, y también con otras formas intermedias entre ambos extremos, como pueden ser libros de texto y enciclopedias.

Una colección puede contener **identificadores** para sus recursos junto con o en lugar de los propios recursos, lo que permite que un recurso forme parte de más de una colección, como las canciones en las listas de reproducción.

Una colección por sí misma es también un recurso. Al igual que otros recursos, una colección puede tener recursos descriptivos asociados a ella. Un índice es un recurso de descripción que contiene información

.....
apartado §6.5 del capítulo 6 “The Structural Perspective”, de libro “TDO core concepts edition.”

4 Véase el recuadro complementario “Las personas como recursos”, en el apartado §2.2.3. del capítulo 2 “Los recursos en los Sistemas de Organización”, en este mismo libro.



sobre la ubicación y la frecuencia de los términos de una colección de documentos para que se pueda buscar de manera eficiente.

Ya que las colecciones son un tipo de recurso importante y de uso frecuente, es importante distinguirlas como un concepto separado. En particular, el concepto de colección está profundamente arraigado en las bibliotecas, museos y otras instituciones que se dedican a seleccionar, reunir, organizar y mantener recursos. Los sistemas de organización en estos dominios pueden describirse a menudo como colecciones de colecciones que se organizan de forma variada según el tipo de recurso, el autor, el creador o el recopilador de los recursos de la colección, o por cualquier otro conjunto de principios o propiedades. En un contexto empresarial, el uso de “colección” para describir un conjunto de recursos es mucho menos común, pero las empresas organizan muchos tipos de recursos, incluidos sus empleados, proveedores, clientes, productos y los activos tangibles e intangibles utilizados para crear los productos y dirigir el negocio. De hecho, una empresa en sí misma puede describirse a veces de forma abstracta como una colección de recursos, especialmente cuando los recursos son componentes o servicios de software).

Un tipo de recurso y su sistema de organización convencional suelen ser el foco de una disciplina. Las etiquetas de categorías, como biblioteca, museo, zoológico y repositorio de datos poseen ya unos significados esenciales, junto a amplias experiencias y prácticas asociadas. Los conceptos y vocabularios especializados a menudo evolucionan con el fin de describirlas. La riqueza que se desprende de esta compleja construcción social y cultural dificulta definir con precisión los límites de las categorías.

Siempre podemos crear nuevas categorías ampliando las definiciones convencionales de “biblioteca” u otros sistemas de organización conocidos y añadiendo modificadores, como cuando Flickr se describe como una biblioteca de intercambio de fotografías basada en la web. Pero siempre que se define un Sistema de Organización con respecto a una categoría familiar, se refuerzan las instancias y características típicas o principales de esa categoría que están profundamente arraigadas en el idioma y la cultura, y se marginan las que son atípicas. En el caso de Flickr, esto significa que sugerimos características que no existen (como la clasificación controlada) u omitimos las características distintivas (como el etiquetado por parte de los usuarios).



¿QUÉ ES UNA BIBLIOTECA?

Casi todos los pájaros vuelan, pero no todos lo hacen. ¿Qué características son las más importantes para clasificar algo como un ave? ¿Qué características son más importantes cuando pensamos en algo como una biblioteca?

Podríamos tratar la circulación, el préstamo y la devolución del mismo artículo, como una de las interacciones con los recursos que define a una biblioteca. En ese caso, una institución que presta artículos de su colección con la esperanza de que los prestatarios devuelvan algo mejor no parece una biblioteca. Pero si pensáramos en recursos como las semillas de plantas, sí se esperaría que los prestatarios devuelvan semillas de las plantas que cultivaron a partir de las semillas prestadas, tal vez “biblioteca de semillas” sea un nombre adecuado para este nuevo Sistema de Organización. Del mismo modo, aunque los recursos de su colección son artículos de enciclopedia en lugar de seres vivos, la enciclopedia de contenido libre Wikipedia se asemeja a la Biblioteca de Semillas al alentar a sus usuarios a “devolver” artículos que mejoren a los actuales.

El sitio web de intercambio de fotografías Flickr funciona para la mayoría de sus usuarios como un archivo personal de fotografías. Los miles de millones de fotos subidas por los usuarios de Flickr y su elección de compartirlas públicamente lo transforman en una colección compartida con capacidad de búsqueda, y muchas personas también piensan en Flickr como una biblioteca de fotos. Pero Flickr carece de la descripción autorizada y la clasificación normalizada que caracterizan a una biblioteca.

Un desafío de categorización parecido surge con el proyecto de digitalización de Google Books.

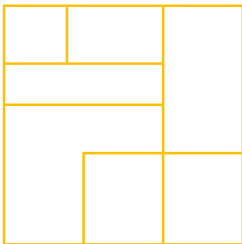




En términos más generales, una visión sistematizadora de los Sistemas de Organización hace que sea muy importante qué categoría se utiliza para establecer definiciones o plantear comparaciones. El proyecto Google Books hace que las obras agotadas y académicas sean mucho más accesibles, pero cuando el cofundador de Google, Sergei Brin, lo describió como “una biblioteca indiferente al paso del tiempo”, molestó a muchas personas con un sentido más tradicional de lo que implica la categoría de biblioteca. Podemos identificar fácilmente las opciones de diseño de Google Books que son más características de los Sistemas de Organización usados en el dominio comercial, y de esta manera el proyecto podría haber tenido una opinión más favorable si se hubiera descrito como una librería online que ofrecía muchos beneficios de forma gratuita.

1.5 EL CONCEPTO DE

“ORDENACIÓN INTENCIONADA”



La ordenación intencionada enfatiza los actos de organización explícitos o implícitos realizados por personas, o por procesos computacionales que actúan como representantes de la intencionalidad humana o como implementación de ésta. La ordenación intencionada es más fácil de ver en los sistemas de organización creados por individuos que pueden tomar todas las decisiones necesarias sobre la organización de sus propios recursos. También es fácil verla en los Sistemas de Organización creados por instituciones como bibliotecas, museos, empresas y gobiernos donde la responsabilidad y la autoridad para organizar está centralizada y se explicita en políticas, normas o reglamentos.

Sin embargo, esta intencionalidad descendente (*top-down*) no siempre es necesaria para crear un Sistema de Organización. La organización puede surgir con el tiempo a través del comportamiento colectivo en situaciones donde no hay un control centralizado, cuando las decisiones tomadas por individuos, cada uno actuando intencionalmente, crean rastros, registros u otra información que se acumula con el tiempo. Los Sistemas de Organización que utilizan mecanismos ascendentes (*bottom-up*) en lugar de descendentes se denominan a veces autoorganización, porque emergen de las interacciones agregadas de los agentes, con los recursos o entre sí. Los sistemas autoorganizados pueden cambiar su estructura interna o su función en respuesta a la retroalimentación o al cambio de circunstancias.



Esta definición es lo suficientemente amplia como para incluir a los ecosistemas comerciales y a los biológicos, las pautas de tráfico y los proyectos de software de código abierto. Otro buen ejemplo de organización emergente se refiere a los sistemas de rutas, en los que las personas (igual que las hormigas y otros animales) pueden seguir y, por lo tanto, reforzar las rutas tomadas por sus predecesores. Cuando de las interacciones locales entre hormigas, abejas, pájaros, peces y otras especies de animales surgen disposiciones muy ordenadas y óptimas, se suele denominar “inteligencia de enjambre” (*swarm intelligence*). Cuando esto sucede con las valoraciones humanas de noticias, videos de YouTube, restaurantes y otros tipos de fuentes digitales y físicas lo llamamos “inteligencia colectiva” (*crowdsourcing*). Lo que las situaciones animales y humanas tienen en común es que la información se comunica entre individuos. A veces esta comunicación es directa, como cuando Amazon te muestra la valoración media de un libro o qué libros han sido comprados por personas como tú. En otras ocasiones la comunicación es indirecta, se logra cuando los agentes modifican su entorno (como lo hacen cuando crean rutas) y otros pueden responder a estas modificaciones. La “mano invisible” de Adam Smith es otro ejemplo en el que los individuos generan colectivamente un resultado que no pretendían directamente, pero que surge de sus acciones separadas e interesadas al responder a las señales de precios en el mercado. Asimismo, aunque no existe una organización descendente, la red en su conjunto, con sus más de un billón de páginas únicas, es un sistema autoorganizado que en su núcleo sigue unos principios organizativos claros.

LA WEB COMO SISTEMA DE ORGANIZACIÓN

La web de hoy en día apenas se parece al sistema de distribución de informes científicos y técnicos para el que fue diseñada cuando el físico e informático Tim Berners-Lee la ideó en 1990 en el laboratorio de la Organización Europea de Investigación Nuclear (CERN) cerca de Ginebra. Sin embargo, como sistema de organización, la web sigue siguiendo los principios que Berners-Lee definió en su creación. Entre ellos figuran los formatos de datos y los protocolos de interacción estándar; la falta de necesidad de un control centralizado de la creación o el enlace de páginas; el acceso remoto a través de la red desde cualquier lugar; y la capacidad de funcionar con una gran variedad de ordenadores y sistemas operativos. Esta arquitectura hace que la web sea abierta y extensible, pero no le da mecanismos incorporados de autoridad o confianza.



Ya que la web funciona sin ninguna autoridad central o control de la autoría, cualquier persona u organización puede añadir contenido a ella. Como resultado, aunque la web en su conjunto no exhibe la disposición intencional centralizada de recursos que caracteriza a muchos Sistemas de Organización, podemos verla como una combinación de millones de Sistemas de Organización y cada uno encarna una disposición intencional separada de las páginas web. Además, la mayoría de las veces interactuamos con la web de forma indirecta utilizando un motor de búsqueda, que se ajusta a la definición de Sistema de Organización porque sus algoritmos de indexación y recuperación siguen unos principios.

Muchos Sistemas de Organización se implementan como colecciones de páginas web. Algunas de estas colecciones se crean en la web como nuevas páginas, otras se crean transformando las colecciones de recursos existentes y otras combinan recursos nuevos y existentes.

El requisito para una ordenación intencionada excluye a los patrones naturales creados por procesos físicos o geológicos, estos no se consideran sistemas organizadores. Hay información en las pilas de escombros dejadas después de un tornado o tsunami y en los estratos del Gran Cañón. Pero no son Sistemas de Organización porque los patrones de disposición fueron creados por fuerzas naturales deterministas en lugar de por agentes que siguieron uno o más principios de organización. Por otro lado, las colecciones de datos geológicos como las mediciones de la composición química de los diferentes estratos y lugares del Gran Cañón sí son Sistemas de Organización. Las decisiones sobre qué medir, cómo combinar y analizar las mediciones, y cualquier teoría que se ponga a prueba o se aplique, reflejan la disposición intencional de los datos por parte del geólogo.

Otras pautas de ordenación de los recursos son ilusiones o percepciones que requieren un punto de vista concreto. Los mejores ejemplos son los patrones de las estrellas tal como aparecen para un observador en la Tierra. Las tres estrellas alineadas con precisión, son conocidas como “Cinturón de Orión”, están a cientos de años luz de la Tierra, y también entre ellas. La disposición que se percibe de las estrellas es innegable, pero las estrellas no están alineadas en el universo. Las constelaciones astronómicas como la de Orión son formas intencionales impuestas desde nuestra percepción de la ubicación de las estrellas, y estas ordenaciones percibidas y las explicaciones que nos ofrece el



uso del concepto de constelaciones conforman un Sistema de Organización que está profundamente arraigado en la cultura humana y en la práctica de la navegación marítima por las estrellas.



Una disposición no intencional

La composición y disposición de las capas de roca (“estratos”) en el Gran Cañón, en el suroeste de los Estados Unidos, ha sido estudiada extensamente por los geólogos. La composición de la roca sugiere el entorno en el que se formó, y la disposición absoluta y relativa de las capas de roca revela la cronología de importantes eventos geológicos. (Foto B. Rosen, Licencia Creative Commons CC-BY-ND 2.0)

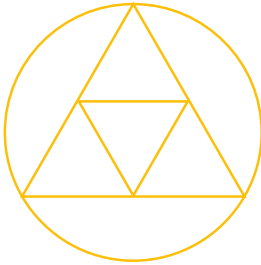
En conjunto, **la organización intencionada de los recursos en un Sistema de Organización es el resultado de decisiones sobre lo que se organiza, por qué se organiza, cuánto se organiza, cuándo se organiza y cómo o por quién se organiza**⁵. Un Sistema de Organización se define por el impacto compuesto de las decisiones tomadas sobre todas estas dimensiones de diseño. Debido a que estas preguntas están interrelacionadas, sus respuestas se aplican de manera integrada al definir un Sistema de Organización.

⁵ Cada uno de estos aspectos se examina con más detalle en el capítulo 2 “Design Decisions in Organizing Systems” del libro “TDO, core concepts edition”.



1.6 EL CONCEPTO DE

“PRINCIPIO DE ORGANIZACIÓN”



La disposición de los recursos en un Sistema de Organización sigue o encarna uno o más principios de organización que permiten al sistema lograr sus propósitos. Los principios de organización son directrices para el diseño o la disposición de un conjunto de recursos que se expresan idealmente de una manera que no asuma ninguna implementación o realización concreta. Llamamos a esta filosofía de diseño “Pensamiento arquitectónico”⁶.

Cuando organizamos una estantería, la oficina en casa, la cocina, o los archivos MP3 de nuestro reproductor de música, los recursos en sí mismos pueden ser nuevos y modernos, pero muchos de los principios que rigen su organización son los que han influido en el diseño de los Sistemas de Organización durante cientos de años. Por ejemplo, organizamos muchas colecciones de recursos utilizando las propiedades que son más fáciles de percibir, o cuyos valores varían más entre los artículos de la colección, porque estos principios hacen que sea fácil localizar un recurso en concreto. También agrupamos los recursos que utilizamos con más asiduidad, hacemos que los recursos que utilizamos a menudo sean más accesibles que los que utilizamos con poca frecuencia y ponemos recursos raros o únicos donde podemos protegerlos. Los principios organizativos muy generales y abstractos a veces se denominan “diseño heurístico” (por ejemplo, “hacer que las cosas sean más fáciles de encontrar”). Los principios de organización más específicos y comúnmente utilizados incluyen el orden alfabético (ordenar los recursos según sus nombres) y el orden cronológico (ordenar los recursos según la fecha de su creación u otro evento importante en la vida del recurso). Algunos principios de organización clasifican los recursos en categorías predefinidas y otros principios de organización se basan en combinaciones novedosas de las propiedades de los recursos para crear nuevas categorías.

Ya que la razón de escribir este libro es ampliar el estudio de la organización de la información más allá de sus raíces en la biblioteconomía y las ciencias de la información, en él se enfatizan los principios de organización que tienen un propósito funcional específico como la identificación, la selección, la recuperación o la preservación de los recursos. Sin embargo, durante miles de años la gente ha recopilado sistemáti-

⁶ Puede consultarse el apartado §11.5.2. del capítulo 11 “The Organizing System Roadmap” del libro “TDO, core concepts edition”.



camente cosas, información sobre esas cosas y observaciones de todo tipo, organizándolas en un esfuerzo por comprender cómo funciona su mundo; los babilonios crearon inventarios y mapas estelares; los egipcios rastrearon las inundaciones anuales del Nilo; y los mesoamericanos crearon calendarios astronómicos. La expresión “dar sentido” (sensemaking) se utiliza a menudo para describir este propósito genérico y menos específico de la organización como el acto de derivar significado de la experiencia haciendo encajar nuevos acontecimientos u observaciones en los que ya se conocen.



Organización de especias por tipo de cocina

Una forma de organización alternativa a la alfabética es organizar las especias según el tipo de cocina o perfiles de sabor, que se puede definir a partir de los ingredientes y especias que tienen a usarse conjuntamente. Patricia Glushko organiza sus especias en tres grupos: Cocina india (incluye guindilla, comino, cúrcuma, cilantro), Cocina mediterránea / Oriente media (incluye albahaca, orégano, paprika, tomillo, eneldo) y semillas. Cada grupo de especias está en un amplio contenedor diferente, lo que resulta cómodo al cocinar.

(Foto R. Glushko)

La expresión de principios organizativos de manera que se separen el diseño y la implementación se ajusta bien a la **arquitectura de tres niveles** que es familiar para arquitectos y diseñadores de software: interfaz de usuario (implementación de las interacciones), lógica de negocio (disposición intencional) y datos (recursos).



La separación lógica entre los principios de organización y su aplicación es fácil de ver con los recursos digitales. En una biblioteca digital, al usuario no le importa si los recursos se almacenan localmente o se recuperan a través de una red. La esencia de un Sistema de Organización de una biblioteca surge de los recursos que organiza y de las interacciones que permite con estos recursos. Por lo general, los usuarios se preocupan mucho por las interacciones que pueden realizar, como los tipos de búsqueda y clasificación que permite el catálogo de la biblioteca en línea. La forma en que se implementan los recursos y las interacciones no suele ser motivo de preocupación, a menos que sea claramente ineficiente.

La separación de los principios de organización y su aplicación es más difícil de reconocer en un Sistema de Organización que sólo contiene recursos físicos, como la cocina o el armario, en los que parece haber interacciones no mediadas con los recursos en lugar de acceder a ellos a través de algún tipo de interfaz de usuario o “capa de presentación” que apoya los principios especificados en el “capa intermedia” y materializados en la “capa de almacenamiento”. Como resultado, la gente puede fácilmente dejarse atrapar solo por las preocupaciones de la presentación. Con demasiada frecuencia perdemos el tiempo organizando por colores las carpetas de archivos y poniendo etiquetas en los contenedores de almacenamiento, cuando sería mejor pensar con más cuidado en la organización lógica de la carpeta y el contenido del contenedor. No sirve de nada utilizar colores y etiquetas para que la organización lógica se resalte con nitidez si no está previamente bien diseñada.

Un lugar donde se pueden apreciar fácilmente estos diferentes niveles de recursos físicos es en la organización de las especias en una cocina. Diferentes cocinas pueden incorporar un principio de organización por orden alfabético para organizar una colección de especias, pero la ubicación exacta y la disposición de las especias en una cocina determinada depende de la configuración de los estantes y cajones, de si se utiliza un armario para especias o una bandeja giratoria, y de otras consideraciones relativas al nivel de almacenamiento. Análogamente, las especias podrían organizarse lógicamente por tipo de cocina, separando las especias indias de las mexicanas, pero este principio de organización no implica nada acerca de dónde podrán encontrarse en la cocina.

La imagen 1.2. titulada “Niveles o capas de presentación, lógica y almacenamiento”, ilustra la separación de los niveles de presentación, lógica y almacenamiento para cuatro tipos diferentes de Sistemas de Organización de bibliotecas y para Google Books. Ninguno de ellos son iguales en todos los niveles o capas. Se puede observar cómo en el



caso una biblioteca que utiliza robots de inventario para gestionar el almacenamiento de libros, esto no afecta al resto sus niveles superiores.



Imagen 1.2. Capas o niveles de Presentación, Lógico y de Almacenamiento.

Es altamente recomendable que en el diseño e implementación de un Sistema de Organización se separe el almacenamiento de los recursos de la lógica de su ordenación y de los métodos para interactuar con ellos. Esta arquitectura a tres niveles es familiar para los diseñadores de Sistemas de Organización informáticos, pero también es útil al pensar con este enfoque incluso en Sistemas de Organización de recursos físicos.

Dado que las cosas tangibles sólo pueden estar en un lugar a la vez, muchos Sistemas de Organización, como los de las bibliotecas modernas con catálogos en línea y colecciones físicas, resuelven esta limitación creando “proxies” o sustitutos digitales de sus recursos tangibles, o bien creando recursos digitales paralelos (por ejemplo, libros digitalizados). Las implicaciones de organizar, encontrar, utilizar y reutilizar



recursos en cualquier Sistema de Organización reflejan directamente la mezcla de estas dos materializaciones de la información; de esta manera podemos pensar en la biblioteca moderna como un Sistema de Organización digital que se basa principalmente en recursos digitales para organizar una mezcla de recursos físicos y digitales.

El Sistema de Organización de una colección pequeña puede a veces utilizar sólo como principio de organización mínimo o por defecto la colocación, es decir, poner todos los recursos en el mismo lugar; por ejemplo, en el mismo contenedor, en el mismo estante o en el mismo buzón de correo electrónico. Si no cocinas mucho y sólo tienes un pequeño número de especias en tu cocina, no necesitas alfabetizarlas porque es fácil encontrar la que quieres.

LOS TRES NIVELES DE SISTEMAS DE ORGANIZACIÓN

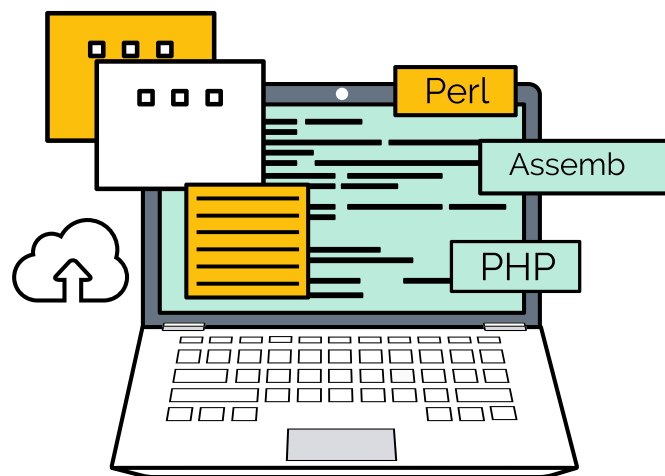
Los arquitectos y los diseñadores de software están de acuerdo en que es conveniente construir aplicaciones que separen el almacenamiento de los datos, la lógica de negocio o las funciones que utilizan los datos, y la interfaz de usuario o los componentes de presentación a través de los cuales los usuarios u otras aplicaciones interactúan con los datos. Esta arquitectura modular permite que cada uno de los tres niveles se mejore, rediseñe o reimplemente de manera independiente para satisfacer requisitos cambiantes o para aprovechar nuevas tecnologías. Una distinción análoga es la que existe entre un algoritmo como descripción lógica de un método para resolver un problema de programación y su implementación en un lenguaje de programación particular como Java o Python.

Estas distinciones de arquitectura son también importantes para los bibliotecarios y especialistas en información. Nuestra nueva forma de ver los Sistemas de Organización enfatiza la importancia de identificar las interacciones deseadas con los recursos, determinando qué principios de organización pueden permitir ciertas interacciones, y luego decidir cómo almacenar y gestionar los recursos de acuerdo con esos principios. Aplicar el pensamiento arquitectónico a los Sistemas de Organización hace más fácil comparar y contrastar los existentes y diseñar otros nuevos. La separación de los principios de organización en el “nivel medio” de sus implicaciones en los niveles de “datos” y “presentación” permite a menudo aplicar el mismo Sistema de Organización lógico en diferentes entornos que soportan las mismas o equivalentes interacciones con los recursos. Por ejemplo, un nuevo requisito de mejora de la búsqueda en un catálogo de biblioteca en un móvil sólo afectaría al nivel de presentación.



Del principio de frecuencia de uso emerge implícitamente cierta forma de organización. En la cocina o en el armario de la ropa, los recursos que se utilizan con mayor frecuencia se sitúan al frente porque es el lugar más accesible. Sin embargo, a medida que una colección crece en tamaño, el tiempo empleado para organizar, localizar y recuperar un recurso en particular se vuelve más relevante. La colección debe organizarse explícitamente para que estas interacciones sean eficientes y la organización debe preservarse después de que la interacción tenga lugar; es decir, los recursos se devuelven al lugar donde fueron encontrados. En consecuencia, la mayoría de los Sistemas de Organización emplean principios de organización que aprovechan las propiedades de los recursos que se organizan (por ejemplo, el nombre, el color, la forma, la fecha de creación, la categoría semántica o biológica), y a menudo se utilizan simultáneamente múltiples propiedades. Por ejemplo, en la cocina se pueden ordenar las ollas y sartenes por tamaño y forma para poder agruparlas y almacenarlas de forma compacta, pero también se pueden organizar las cosas por estilo de cocina y separar los utensilios de la barbacoa del wok y otros instrumentos para hacer comida oriental.

A diferencia de los recursos físicos, las propiedades organizativas más útiles para los recursos de información son las que reflejan su contenido y significado, y éstas no son directamente evidentes cuando se mira un libro, un documento o una colección de datos. Es necesario un esfuerzo intelectual significativo o un cálculo estadístico para revelar estas propiedades cuando se asignan términos de materia, se elabora un índice o se utilizan características como inputs en programas de aprendizaje automático y análisis de datos.





Separación de los principios de organización de su implementación

Tanto si las especias están organizadas alfabéticamente, por tipo de cocina, por temporada, por frecuencia de uso o por cualquier otro principio, esta decisión es lógicamente distinta de la disposición física de las especias. Hay muchos tipos de estantes para especias, armaritos, estanterías giratorias y otros dispositivos diseñados para ordenarlas.

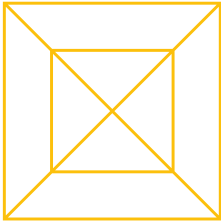
(Collage creado por R. Glushko de varios catálogos web.)

Los Sistemas de Organización más eficaces para los recursos de información suelen basarse en las propiedades estadísticas que surgen del análisis de la colección como un todo. Por ejemplo, la pertinencia de los documentos para una consulta de búsqueda es mayor cuando contienen una frecuencia superior a la media de los términos de consulta comparados con otros documentos de la colección, o cuando están vinculados a documentos pertinentes. Asimismo, los algoritmos de clasificación de los mensajes de correo electrónico recalculan continuamente la probabilidad de que palabras como “beneficiario” o “Viagra” indiquen si un mensaje es “spam” o “no spam” en la colección de mensajes procesados.



1.7 EL CONCEPTO DE

“AGENTE”



Muchas disciplinas tienen puestos de trabajo especializados para distinguir a las personas que organizan los recursos (por ejemplo: catalogadores, archiveros, indexadores, conservadores, gestores de colecciones...). Utilizamos el término más general, agente, para cualquier entidad capaz de organizar de forma autónoma e intencionada, porque considera que el trabajo de organización realizado por las personas y el trabajo de organización realizado por ordenadores tienen objetivos comunes, a pesar de las obvias diferencias en los métodos.

Podemos analizar los agentes en los Sistemas de Organización para entender cómo los esfuerzos humanos y computacionales se complementan y sustituyen los unos a otros con el fin de organizar recursos. Podemos determinar los contextos económicos, sociales y tecnológicos en los que cada tipo de agente puede ser empleado de mejor manera. Podemos determinar cómo el Sistema de Organización distribuye costes y esfuerzos entre sus creadores, usuarios, mantenedores y otros colectivos implicados (*stakeholders*).

Un grupo de personas puede ser un agente organizador, como cuando un grupo se reúne en un club o en un comité técnico del organismo de normalización en el que los miembros del grupo subordinan su individualidad para lograr un bien común.

También usamos el término agente cuando discutimos las interacciones con los Sistemas de Organización. Las entidades que más suelen acceder al contenido de las bibliotecas, museos u otras colecciones de recursos físicos son los agentes humanos, es decir, las personas. En otros Sistemas de Organización, como los sistemas de información empresarial o los repositorios de datos, las interacciones con los recursos se llevan a cabo mediante procesos computacionales, dispositivos robóticos u otras entidades que actúan de forma autónoma en nombre de una persona o grupo.

En algunos Sistemas de Organización, los propios recursos son capaces de iniciar interacciones con otros recursos o con agentes externos. Esto es más evidente con los recursos humanos u otros recursos vivos, en los que una parte crítica del diseño de cualquier Sistema de Organización de este tipo es determinar qué tipos de interacciones se deben alentar o permitir que se inicien. Volveremos a este tema después de



discutir el diseño de las interacciones con los recursos ordinarios que son pasivos, situación que se da en la mayoría de los Sistemas de Organización que involucran recursos físicos.

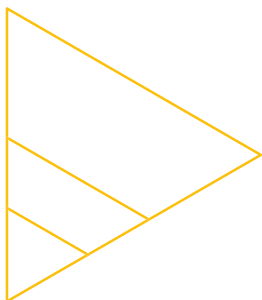
Otros recursos que pueden iniciar interacciones son los recursos aumentados con capacidades sensoriales, computacionales o de comunicación que les permiten obtener información de su entorno y luego hacer algo útil con ella. Probablemente esté familiarizado con las etiquetas RFID, que permiten la identificación y la localización precisa de los recursos físicos a medida que se desplazan a través de las cadenas de suministro y las tiendas, y con los dispositivos “inteligentes” como los termostatos Nest que aprenden a programarse a sí mismos.



1.8 EL CONCEPTO DE

“INTERACCIONES”

Una interacción es una acción, función, servicio o capacidad que hace uso de los recursos de una colección o de la colección en su conjunto. La interacción de acceso es fundamental en cualquier colección de recursos, pero muchos Sistemas de Organización proveen funciones adicionales para hacer el acceso más eficiente y apoyar otras interacciones adicionales con los recursos a los que se accede. Por ejemplo, las bibliotecas y los Sistemas de Organización similares utilizan catálogos para permitir interacciones para encontrar un recurso conocido, identificar cualquier recurso de la colección y discriminar o seleccionar entre recursos similares.





Algunas de las interacciones con los recursos en un Sistema de Organización están intrínsecamente determinadas por las características del recurso. Dado que los objetos de los museos son únicos o extremadamente valiosos, los visitantes pueden verlos, pero no pueden tomarlos prestados, a diferencia de la mayoría de los recursos de las bibliotecas. Una biblioteca puede tener múltiples copias impresas de Moby Dick, pero nunca puede prestar más de las que posee. Después de que un ejemplar de un libro impreso sea prestado en una biblioteca, hay muchos tipos de interacciones que pueden tener lugar: lectura, traducción, resumen, anotación, etc., pero éstas no están apoyadas directamente por el Sistema de Organización de la biblioteca y son invisibles para él.

En el caso de las obras que no son de dominio público, la ley de derechos de autor otorga al titular de estos el derecho a impedir algunos usos, pero al mismo tiempo la doctrina jurídica del “uso legítimo” y otras similares sobre derechos de copia permiten ciertos usos limitados incluso para las obras protegidas por derechos de autor.

Los recursos digitales permiten una mayor gama de interacciones que los físicos. Cualquier número de personas o procesos pueden solicitar un pronóstico del tiempo a un servicio meteorológico en línea porque el pronóstico no se agota con la solicitud y el coste marginal de permitir otro acceso es casi cero. Además, con los recursos digitales se pueden habilitar muchos nuevos tipos de interacciones mediante aplicaciones, servicios web o interfaces de programación de aplicaciones (API) en el Sistema de Organización. En particular, la traducción, el resumen, la anotación y la sugerencia de palabras clave son servicios de gran utilidad que suelen estar respaldados por los motores de búsqueda de la web y otras aplicaciones web. De manera similar, un Sistema de Organización con recursos digitales puede implementar una interacción de “mantener todo actualizado” que coloca (push) automáticamente el contenido nuevo en el navegador del usuario.

Pero, así como la tecnología puede permitir interacciones, puede impedir las o limitarlas. Si una colección de recursos digitales (libros electrónicos o música, por ejemplo) no está almacenada en el propio ordenador personal o dispositivo del usuario, una conexión continua a Internet es un requisito para el acceso. Además, las políticas de control de acceso y la tecnología de gestión de derechos digitales (DRM) pueden limitar los dispositivos que pueden acceder a la colección e impedir la realización de copias, anotaciones y otras acciones que de otro modo podrían estar autorizadas por la doctrina de “uso legítimo”.

El diseño de la interacción es especialmente crucial para la gestión de los recursos que tienen la capacidad de iniciar interacciones entre sí o

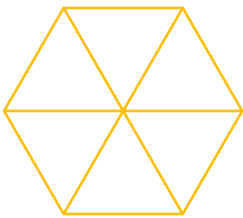


con agentes externos. Considérense las grandes diferencias en el comportamiento de los trabajadores en empresas organizadas según los principios de la gestión científica y las que encarnan los principios Kai-zen de mejora continua. En los primeros, el trabajo es muy estandarizado y burocrático, lo que da a los trabajadores poca autonomía. En el segundo, el trabajo está siempre estandarizado, pero los trabajadores están motivados para analizar y mejorar los procesos de trabajo siempre que sea posible, y se les da una gran discrecionalidad en la forma de hacerlo.

Al igual que con los principios de organización, es útil pensar en las interacciones de una manera abstracta o lógica que no implique una implementación concreta, porque puede propiciar la aparición de diseños innovadores para los Sistemas de Organización.

1.9 EL CONCEPTO DE

“RECURSO DE INTERACCIÓN”



Las interacciones con los recursos físicos a veces dejan rastros u otras pruebas. Muchos de estos rastros son involuntarios, como las huellas dactilares, la mancha de una taza de café en un periódico, o la erosión al tomar un atajo a través de un césped. Los fans de Sherlock Holmes y CSI saben que los investigadores forenses pueden usar estos residuos de interacciones para identificar o confirmar la culpabilidad de los sospechosos. Otros rastros de interacción son intencionales, como cuando un estudiante subraya los apuntes o las notas en un libro de texto o el graffiti pintado con spray en un edificio. Sin embargo, no toda interacción deja un rastro, ya que se desvanecen con el tiempo y los diferentes rastros asociados con un mismo recurso carecen de consistencia. Esto significa que la mayoría no son de mucha utilidad.

Sin embargo, cuando los Sistemas de Organización contienen recursos digitales o recursos físicos que tienen capacidad de detección, registro o comunicación, los rastros o trazas de la interacción pueden hacerse predecibles, persistentes y consistentes. Cada registro de una elección del usuario en cuanto al acceso, la navegación, la compra, el resaltado, la vinculación y otros aspectos se convierte entonces en un “recurso de interacción” que puede analizarse para reorganizar la colección de recursos o influir de algún otro modo en las interacciones posteriores con los recursos primarios.



Los recursos de interacción son a menudo piezas esenciales de información que hacen funcionar a los Sistemas de Organización. La mayoría de las personas que cobran peaje han sido reemplazadas por “etiquetas de peaje” inteligentes que difunden su identidad cuando el coche en el que están pasa por un receptor de radiofrecuencia en una zona de peaje. Cada recurso de interacción creado identifica una cuenta bancaria y una tarjeta de crédito con la que se paga el peaje; en conjunto, todos estos recursos de interacción pueden utilizarse como recursos primarios en otros Sistemas de Organización que gestionan la congestión del tráfico o que apoyan el diseño de carreteras. De manera similar, los recursos de interacción creados por los motores de búsqueda pueden utilizarse para ajustar el orden de los resultados de la búsqueda, seleccionar anuncios o personalizar el contenido de las páginas web.





CAPÍTULO 2

LOS RECURSOS EN LOS SISTEMAS DE ORGANIZACIÓN

Robert J. Glushko / Daniel D. Turner / Kimra McPherson / Jess Hemerly

CONTENIDO



53

2.1 Introducción

53

2.1.1 ¿Qué es un recurso?

2.1.1.1 Los recursos compuestos por partes

2.1.1.2 Los recursos bibliográficos, los componentes informativos y las cosas “inteligentes” como recursos

57

2.1.2 La identidad, los identificadores y los nombres



59

2.2 Cuatro distinciones sobre los recursos

59

2.2.1 El dominio de los recursos

62

2.2.2 El formato de un recurso

65

2.2.3 La agencia de un recurso

2.2.3.1 Recursos pasivos o recursos operandos

2.2.3.2 Recursos activos, o recursos operantes

72

2.2.4 El foco de un recurso

74

2.2.5 Recursos según formato y foco

2.2.5.1 Descripción física de un recurso primario físico

2.2.5.2 Descripción digital de un recurso primario físico

2.2.5.3 Descripción digital de un recurso primario digital

2.2.5.4 Descripción física de un recurso primario digital



78

2.3 La identidad de los recursos

78

2.3.1 Identidad en los recursos físicos

79

2.3.2 Identidad y recursos bibliográficos

81

2.3.3 Identidad y componentes informativos

83

2.3.4 Identidad y recursos activos



85

2.4. El nombre de los recursos

85

2.4.1. ¿Qué es un nombre?

85

2.4.2. Los problemas de la denominación

2.4.2.1 El problema del vocabulario



96

2.5. Los recursos y el paso del tiempo

96

2.5.1. Persistencia

2.5.1.1 Identificadores Persistentes

2.5.1.2 Recursos permanentes o persistentes

98

2.5.2 Efectividad

100

2.5.3 Autenticidad

101

2.5.4 Procedencia

- 2.4.2.2 Homonimia, polisemia y los falsos amigos
- 2.4.2.3 Nombres con asociaciones no deseadas
- 2.4.2.4 Nombres que asumen atributos no permanentes
- 2.4.2.5 La brecha semántica

2.4.3 La elección de buenos nombres e identificadores

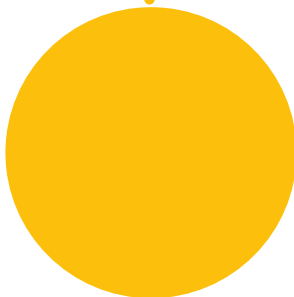
2.4.3.1 Producción de nombres informativos

2.4.3.2 Uso de vocabularios controlados

2.4.3.3 Permitir los alias

2.4.3.4 Construcción de identificadores únicos o cualificados

2.4.3.5 La diferencia entre identificar y resolver





2.1. INTRODUCCIÓN

Este capítulo se construye sobre los fundamentos conceptuales introducidos en el capítulo 1 y se detiene en explicar en detalle qué es lo que se puede entender por recurso. En concreto, ponemos el foco en los aspectos relativos a la identidad – es decir, las decisiones sobre qué es lo que será tratado como un recurso independiente – y se debaten también las consideraciones y principios que hay que tener en cuenta cuando asignamos una denominación (nombre, título) o un identificador a un recurso.

2.1.1. ¿QUÉ ES UN RECURSO?

Los recursos son aquello que organizamos

En el primer capítulo introdujimos el concepto de recurso (§1.3 El concepto de “Recurso”) en su sentido más común: “cualquier cosa de valor que pueda apoyar una actividad orientada a un objetivo”, y resaltamos que un grupo de recursos puede tratarse como una colección en un Sistema de Organización. Y, ¿qué queremos decir realmente con “cualquier cosa de valor”? A priori, puede parecer que la cuestión de la identidad, de aquello que define a un recurso en concreto, no es difícil de responder. Después de todo, vivimos en un mundo lleno de recursos y encontrar, seleccionar, describir, organizar y referirse a ellos son actividades cotidianas. Y, aunque los recursos humanos no son el foco principal de este libro, estaríamos cometiendo una omisión si no explicamos por qué tiene sentido entender también a las personas como recursos. (Véase la barra lateral titulada “Las personas como recurso”).

Sin embargo, incluso cuando los recursos con los que tratamos son tangibles, la forma en que los organizamos no siempre es obvia o, al menos, no lo es en la misma medida y en el mismo momento para todos nosotros. No todos pensamos en ellos de la misma manera. El primer paso para poder organizar un recurso, así como otros similares, es reconocerlo, en el sentido de percibirlo como algo tangible. Aquellas de sus propiedades que atraen nuestra atención y que utilizamos para organizarlos dependen de nuestras experiencias, nuestros propósitos y nuestro contexto.

Cuando nombramos o describimos un recurso, estamos añadiéndole información más allá de su propia entidad, y haciéndolo algo más que ese “algo” que era antes. Podemos describir el mismo recurso de



muchas maneras diferentes. Con frecuencia podemos considerar que un recurso determinado es uno de los muchos componentes de una categoría amplia o uno de los pocos componentes de una categoría reducida, o un ejemplar único de una categoría compuesta por él exclusivamente. Por ejemplo, podemos reconocer algo como un artículo de ropa, como un calcetín o como ese calcetín sucio con el agujero en el talón que usamos en la larga caminata de ayer. Sin embargo, incluso después de categorizar algo, podríamos no ser tan rigurosos al hablar de ello; a menudo nos referimos a dos objetos como “la misma cosa” cuando lo que queremos decir es que son “el mismo tipo de cosa”. De hecho, podríamos debatir sobre si una categoría con un solo componente es realmente una categoría, porque difumina una importante distinción entre determinados artículos o ejemplares y la clase o tipo a la que pertenecen.

Las cuestiones que importan y las decisiones que debemos tomar sobre los ejemplares, las clases y los tipos de recursos no son completamente estancas. No obstante, intentaremos centrarnos en este capítulo en los recursos concretos (los ejemplares) dejando fuera lo relativo a clases y tipos¹.

2.1.1.1 LOS RECURSOS COMPUESTOS POR PARTES

Por muy difícil que sea decidir qué es un recurso cuando se trata de objetos individuales, es aún más difícil cuando los recursos son objetos o sistemas compuestos por otras partes. En estos casos, debemos centrarnos en la totalidad del objeto o sistema y tratarlo como un recurso, tratar sus partes como recursos y tratar las relaciones entre las partes y el todo, como lo hacemos con los dibujos de ingeniería y las instrucciones de montaje.

¿Cuántas cosas componen un coche? Si intentas imaginar cómo se ensambla un coche, podrías pensar en varias docenas de componentes grandes como el bastidor, la suspensión, la cadena de tracción, el depósito de gasolina, los frenos, el motor, el tubo de escape, el compartimento de los pasajeros, las puertas y otros componentes preensamblados. Por supuesto, cada uno de ellos está compuesto de muchas partes, como el motor o, incluso, la radio. Hay quienes se han dedicado a la ardua tarea de contar hasta diez o quince mil partes que componen, de media, un coche. Pero, incluso con tal grado de precisión, muchas partes siguen siendo cosas compuestas de otras. Hay tornillos, cables, cinturones y mucho más; en realidad, son demasiadas para contarlas.

¹ Puede consultarse capítulo 7 “Categorización: Describir clases y tipos de recursos”, en la edición online completa en inglés del libro “TDO core concepts edition”.



La ambigüedad sobre el número de partes del conjunto se mantiene también para los recursos informativos; un periódico puede ser considerado como un único recurso, pero también puede constar de múltiples secciones, cada una de las cuales contiene historias separadas, cada una de las cuales tiene a su vez muchos párrafos y así sucesivamente. De la misma manera, si bien una página web puede tratarse como un único recurso, también puede considerarse como una colección de partes más granulares, cada una de las cuales puede identificarse por separado como el código fuente o el texto de ancla de un enlace. Asimismo, cuando un cliente solicita en el banco una tarjeta de crédito, se puede recabar información acerca de los préstamos pendientes, el historial de pagos, los ingresos actuales y otro tipo de información, o el banco puede simplemente acceder a la puntuación crediticia, el indicador agregado que combina todos estos datos financieros en un solo número.



¿Cuántos componentes definen el juego del ajedrez?

El juego del ajedrez ejemplifica las diferentes maneras de decidir qué considerar como único recurso. ¿Lo podemos considerar como dos juegos de piezas de ajedrez? ¿seis tipos de piezas de ajedrez (1 rey, 1 reina, 2 torres, 2 alfiles, 2 caballeros, 8 peones)? ¿33 cosas separadas (las 32 piezas y el tablero)?

(Fotografía de Emma Jane Hogbin Westby).

Licencia de Creative Commons CC-BY-2.0



2.1.1.2 LOS RECURSOS BIBLIOGRÁFICOS, LOS COMPONENTES INFORMATIVOS Y LAS COSAS “INTELIGENTES” COMO RECURSOS

Los recursos de información suelen plantear problemas adicionales en su identificación y descripción porque su propiedad más importante suele ser su contenido, que no se reconoce fácil y sistemáticamente. Los Sistemas de Organización de recursos de información en forma física, como los de las bibliotecas, tienen que hacer malabarismos con la dualidad entre sus propiedades tangibles, y con su carácter inherente de recurso de información abstracto; es decir, el libro impreso versus el conocimiento que contiene. Aquí, el Sistema de Organización hace hincapié en los recursos descriptivos o “recursos sustitutos”, como los registros bibliográficos, que describen el contenido informativo, en lugar de sus propiedades físicas.

Otra cuestión especialmente complicada sobre los recursos que componen una biblioteca es la siguiente: **¿Qué conjunto de recursos debería ser tratado como la misma obra porque contienen esencialmente un contenido intelectual o artístico similar?** Podemos hablar de *Macbeth* como la obra de Shakespeare, pero ¿qué es esta cosa que llamamos “*Macbeth*”? ¿Es una sucesión de palabras guardada en un archivo digital o está escrita a mano? ¿Es la colección de palabras impresas con alguna tipografía específica y una paginación concreta? ¿Todas las ediciones e impresiones de estas palabras son el mismo *Macbeth*? ¿Cómo deberíamos organizar las numerosas representaciones en vivo y grabadas de obras de teatro y películas que comparten el nombre de *Macbeth*? ¿Qué hay de las creaciones basadas o inspiradas en *Macbeth* que no comparten el título “*Macbeth*”, como la película de Kurosawa “*Kumonosu-jo*” (Trono de Sangre) que traslada la trama al Japón feudal? Patrick Wilson propuso una analogía genealógica, caracterizando una obra como “un grupo o familia de textos”, con la idea de que una creación como *Macbeth* (la de Shakespeare) es el “ancestro de los miembros posteriores de la familia”.

Los diseñadores y arquitectos de sistemas de información se enfrentan a retos de diseño análogos cuando describen los componentes informativos en Sistemas de Organización empresariales o científicos. **El contenido de la información se fusiona o confunde intrínsecamente con la estructura y la presentación siempre que se utiliza en un momento y contexto específicos.**

Por un lado, parece evidente que un formulario de pedido contiene componentes de información como ARTÍCULO, NOMBRE DEL CLIEN-



TE, DIRECCIÓN e INFORMACIÓN DE PAGO, pero la disposición de estos componentes, su tipo de letra y tamaño, y otras propiedades no semánticas pueden variar mucho en orden e incluso dentro de un único sistema de información que reutiliza estos componentes para cartas, avisos de entrega, etiquetas de correo y entradas en bases de datos.

Cuestiones similares sobre la identidad de los recursos se plantean con la aparición de la computación ubicua, en la que la capacidad de procesamiento de información y la conectividad están incorporadas en los objetos físicos, en dispositivos como los teléfonos inteligentes y todas las infraestructuras de telecomunicaciones que los circundan. Estas “cosas inteligentes” están equipadas con sensores, etiquetas de identificación por radiofrecuencia (RFID), datos de GPS y metadatos aportados por los usuarios, y que van creando un torrente desordenado de información sobre la ubicación y otras propiedades que deben ser clasificadas en flujos identificados y luego cotejadas o asociadas con el recurso original.

Las consideraciones y los métodos para determinar «¿qué es un recurso?» se abordan en el apartado §2,3 de este capítulo sobre “La identidad de los recursos”, tanto para recursos físicos, como para recursos bibliográficos, componentes informativos y “cosas inteligentes” sobre las que ya hemos empezado a tratar en este apartado.

2.1.2 LA IDENTIDAD, LOS IDENTIFICADORES Y LOS NOMBRES

La respuesta a la pregunta planteada sobre “¿Qué es un recurso?” tiene dos partes.

► **La primera es la identidad: ¿qué cosa estamos tratando como el recurso?**

► **La segunda es la identificación: diferenciar entre un determinado recurso concreto y otros recursos similares.**

Estos problemas están estrechamente relacionados. Una vez que se ha decidido qué tratar como recurso, se crea un nombre o un identificador para poder referirse a él de forma fiable. Un nombre es una etiqueta para un recurso que se utiliza para distinguirlo de otro. Un identificador es un tipo especial de nombre asignado de forma controlada y regido por unas reglas que definen posibles valores y convenciones de deno-



minación. En el caso de un recurso digital, su identificador sirve como variable de entrada (input) en el sistema o la función que determinará su ubicación para que pueda ser recuperado, un proceso que se denomina “resolver” un identificador o resolución.

La elección de nombres e identificadores -ya sea para una persona, un servicio, un lugar, una tendencia, una obra, un documento o un concepto- suele ser difícil y muy polémica. Esta tarea se ve dificultada por innumerables factores, entre ellos el público que necesitará acceder, compartir y utilizar los nombres empleados, las limitaciones de idioma, la política institucional y los sesgos individuales y culturales.

Un problema frecuente es el hecho de que un recurso tenga más de un nombre o identificador. Cuando algo tiene más de un nombre, cada uno de los múltiples nombres es un sinónimo o alias. Un ejemplar físico de un libro concreto puede llamarse tapa dura o rústica o simplemente un texto. George Furnas y sus compañeros de investigación se refirieron a este inconveniente como el “problema del vocabulario”.

Lo llamemos libro o texto, el recurso suele tener un número de catálogo de la Biblioteca del Congreso de los Estados Unidos, así como un ISBN como identificador. Cuando el libro está en una caja de libros que se envía desde la editorial a una librería o una biblioteca, esa caja tendrá un número de seguimiento con código de barras asignado por el servicio de entrega y un formulario de recepción creado por la editorial, cuyo identificador asocia el envío con el cliente. Cada uno de estos identificadores es único con respecto a algún ámbito o contexto establecido.

Una solución parcial para el problema del vocabulario es utilizar un vocabulario controlado. Podemos imponer reglas que estandaricen la forma en que se asignan los nombres y las etiquetas de los recursos. Alternativamente, podemos establecer mapeo entre los términos utilizados en nuestro lenguaje natural y los términos autorizados o controlados. Sin embargo, el control del vocabulario no erradica la ambigüedad. Incluso si el pasaporte o un documento nacional de identidad requieren nombres completos autorizados en lugar de apodos, podría haber fácilmente más de un Robert John Smith en el sistema.

El control del lenguaje utilizado para un fin determinado plantea otras cuestiones: ¿Quién elabora y hace cumplir estas reglas? ¿Qué sucede cuando Sistemas de Organización establecidos sobre normas diferentes se comparan, combinan o usan juntos en contextos distintos de aquellos para los que fueron concebidos originalmente?



2.2 CUATRO DISTINCIONES SOBRE

LOS RECURSOS

La naturaleza de un recurso es fundamental para la creación y el mantenimiento de Sistemas de Organización de calidad. Hay cuatro distinciones que planteamos en relación con los recursos: el dominio, el formato, la agencia y el foco. En la imagen 2.1, Dominio, Formato, Foco y Agencia de los recursos, se representan estas cuatro distinciones, perspectivas o puntos de vista sobre los recursos; al no ser independientes, no podemos presentar estas distinciones como categorías de recursos.



2.2.1 EL DOMINIO DE LOS RECURSOS

El dominio de un recurso es una noción intuitiva que agrupa los recursos según el conjunto de características naturales o intuitivas que los distinguen de otros recursos. El término "dominio" se opone a la idea de agrupaciones de recursos ad hoc o arbitrarias, motivadas por su presencia en algún momento en un mismo espacio.



En el caso de los recursos físicos, los dominios pueden distinguirse grosso modo según el tipo de material del que están hechos, utilizando propiedades fácilmente perceptibles. El nivel más alto de la clasificación realizada por Carl Linneo en 1735, lo clasifica todo en los reinos animal, vegetal y mineral. Dicha clasificación está profundamente arraigada en la mayoría de los idiomas y culturas para crear un sistema jerárquico de categorías de dominios. Muchos aspectos de este sistema de categorías de dominios están determinados por las limitaciones naturales que determinan la pertenencia a una categoría u otra, y que existen como patrones de propiedades compartidas y correlacionadas; un recurso identificado como miembro de una categoría debe ser también miembro de otra con la que comparte algunas propiedades, pero no todas. Por ejemplo, una estatua de mármol en un museo debe ser también un tipo de material y un pez en un acuario debe ser también un tipo de animal.

EL ESPECTRO DE TIPOS DE DOCUMENTOS

Se pueden distinguir diferentes dominios o tipos de documentos según la medida en que su contenido esté prescrito semánticamente, por la cantidad de estructura interna, así como por las correlaciones entre su presentación/formato y su contenido/estructura. Estas tres características de contenido, estructura y presentación varían sistemáticamente desde los tipos de documentos narrativos o discursivos, como las novelas, hasta los tipos de documentos transaccionales, como las facturas.

Los tipos narrativos son escritos por personas y son heterogéneos en estructura y contenido, que suele ser esencialmente prosa y elementos gráficos. Sus características de presentación refuerzan cuidadosamente su estructura y su semántica; por ejemplo, el texto de los títulos o encabezamientos principales es de mayor tamaño porque anuncian un contenido más importante que el que se encuentra en las notas a pie de página, de menor tamaño. Los documentos transaccionales suelen crearse mecánicamente y, en consecuencia, son homogéneos en cuanto a su estructura y su contenido; Este último suele consistir en “datos”: un contenido fuertemente tipificado con una semántica precisa que puede ser procesada por ordenadores.

En la parte media del espectro se encuentran tipos de documentos híbridos como los libros de texto, las enciclopedias y los manuales técnicos que contienen a la vez texto narrativo y contenido estructurado con imágenes, tablas de datos, ejemplos de códigos, etc.



En el caso de los recursos de información, las propiedades que se perciben fácilmente, como el color o el tamaño de un libro, se correlacionan de manera menos fiable con el dominio del recurso, por lo que con mayor frecuencia distinguimos los dominios en función de las propiedades semánticas; las definiciones de los tipos de recursos, como “enciclopedia”, “novela” y “factura”, no los distinguen tanto según la gran variedad de formas físicas en las que podríamos encontrarlos, sino más bien según su temática típica o su tipo de contenido. Organizar libros por color o tamaño podría ser viable para colecciones muy pequeñas o en un estudio fotográfico, pero agruparlos según las propiedades físicas haría que fuese prácticamente imposible encontrar libros en una gran biblioteca.

Podemos organizar los tipos de recursos de información en una jerarquía. Sin embargo, como los límites entre las categorías son un tanto difusos, **es más útil ver los dominios de los recursos de información como un continuo que se extiende desde el contenido narrativo débilmente estructurado hasta el contenido transaccional fuertemente estructurado**. Este marco, llamado el espectro de los tipos de documentos por Glushko y McGrath, encierra la idea de que los límites entre los dominios de recursos, como los que hay entre los colores del arco iris, son fáciles de ver para los colores que están muy separados en el espectro, pero difíciles de ver para los adyacentes. (Ver el recuadro lateral, “El espectro de los tipos de documentos” y su correspondiente representación, Imagen 2.2).

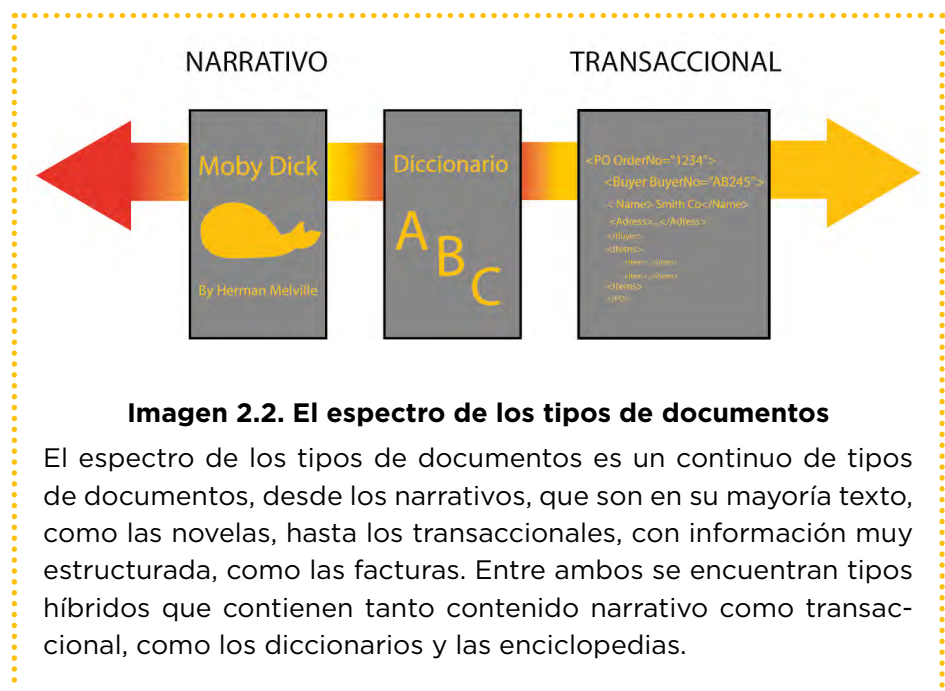


Imagen 2.2. El espectro de los tipos de documentos

El espectro de los tipos de documentos es un continuo de tipos de documentos, desde los narrativos, que son en su mayoría texto, como las novelas, hasta los transaccionales, con información muy estructurada, como las facturas. Entre ambos se encuentran tipos híbridos que contienen tanto contenido narrativo como transaccional, como los diccionarios y las enciclopedias.



2.2.2 EL FORMATO DE UN RECURSO

Los recursos de información pueden existir en numerosos formatos, partiendo desde la más básica distinción entre los físicos y los digitales. Esta distinción es muy importante en la implementación de un sistema de almacenamiento o preservación de recursos porque es ahí donde, por lo general, las propiedades físicas tienen que ser tenidas en consideración y muchas veces suponen restricciones. Esta distinción es menos importante a nivel lógico cuando diseñamos interacciones con recursos porque los sustitutos digitales de los recursos físicos pueden superar las limitaciones que plantean sus propiedades físicas. Cuando buscamos coches o electrodomésticos en una tienda en línea no importa dónde se encuentran los coches o electrodomésticos reales o cómo están organizados. (Ver el recuadro “Los tres niveles de los Sistemas de Organización” del capítulo 1).

Muchas representaciones digitales pueden asociarse con recursos físicos o digitales, pero es importante saber cuál es el recurso original o primario, especialmente para aquellos únicos o valiosos.

Hoy en día muchos recursos sobre los que aplicamos Sistemas de Organización se crean nativamente de forma digital, creados en procesadores de texto, cámaras digitales, grabadoras de audio y vídeo. Otros recursos son elaborados por sensores de dispositivos “inteligentes” y por otros sistemas que crean recursos digitales cuando interactúan con códigos de barras, códigos QR, etiquetas RFID u otros mecanismos para rastrear la identidad y la ubicación.

Otros recursos digitales son creados por digitalización, el proceso de transformación de un artefacto cuyo formato original es físico para que pueda ser almacenado y manipulado por un ordenador. Digitalizamos las palabras impresas, las fotografías, los planos y los álbumes de fotos. El texto impreso, por ejemplo, puede digitalizarse escaneando las páginas y utilizando programas de reconocimiento de caracteres o simplemente volviéndolo a escribir.

Hay un gran número de formatos digitales que difieren en muchos aspectos, pero simplificando podemos diferenciarlos aplicando dos dimensiones: **el grado en que distinguen el contenido de la información de su presentación o representación, y el nivel de explicitación con que se representan las distinciones de contenido.** En conjunto, estas dos dimensiones nos permiten comparar los formatos según su digamos “Coeficiente de inteligencia en información” global, siendo uno de sus principios genéricos que los formatos “más inteligentes” contienen más información procesable por ordenador, como se ilustra en la imagen 2.3.



Los formatos digitales simples para documentos de “texto simple” contienen sólo los caracteres que se ven en el teclado del ordenador. El ASCII es el formato simple más utilizado, pero es inadecuado para la mayoría de los idiomas, que tienen conjuntos de caracteres más amplios, y tampoco puede manejar caracteres matemáticos. El estándar Unicode fue diseñado para superar estas limitaciones².

La mayoría de los formatos de documentos también codifican explícitamente una jerarquía de componentes estructurales, como capítulos, secciones o componentes semánticos como descripciones o pasos de un procedimiento, y a veces el aspecto de la forma “renderizada” o impresa. Otra distinción importante que hay que tener en cuenta es si la información se codifica como una secuencia de caracteres de texto, de manera que sea legible tanto para el ser humano como para el ordenador. La codificación del contenido de los caracteres con XML, por ejemplo, permite la estratificación de la codificación intencional o el marcado entrelazado con el contenido de “texto simple”. Dado que los procesadores XML deben ser compatibles con Unicode, cualquier carácter puede aparecer en un documento XML. Los formatos digitales más complejos son los de los recursos multimedia y los datos multidimensionales, en los que el formato de datos está altamente optimizado para análisis o aplicaciones especializadas.

La digitalización de recursos no textuales como negativos fotográficos, dibujos y grabaciones analógicas visuales y de audio plantea un complicado conjunto de opciones sobre la densidad de píxeles, la profundidad de color, la tasa de muestreo, el filtrado de frecuencias, la compresión y otras muchas cuestiones técnicas que determinan la representación digital.

Puede haber múltiples usos y dispositivos previstos para un recurso digitalizado que podrían requerir diferentes enfoques y formatos de digitalización. Los usuarios intermedios de los recursos digitalizados necesitan conocer el formato en el que se ha creado un artefacto digital, para poder reutilizarlo tal como está o procesarlo de otras formas.

Algunos formatos digitales admiten interacciones cualitativamente diferentes y más potentes que las posibles con los recursos físicos. Los museos están utilizando las tecnologías del mundo de la realidad virtual para crear exposiciones interactivas en las que los visitantes pueden volar a través del sistema solar, escanear sus propios cuerpos y

² Ambos formatos son tratados con más detalle en el capítulo 9 “Los formatos de los recursos descriptivos”, disponible en la edición online completa del libro “TDO core concepts edition”.

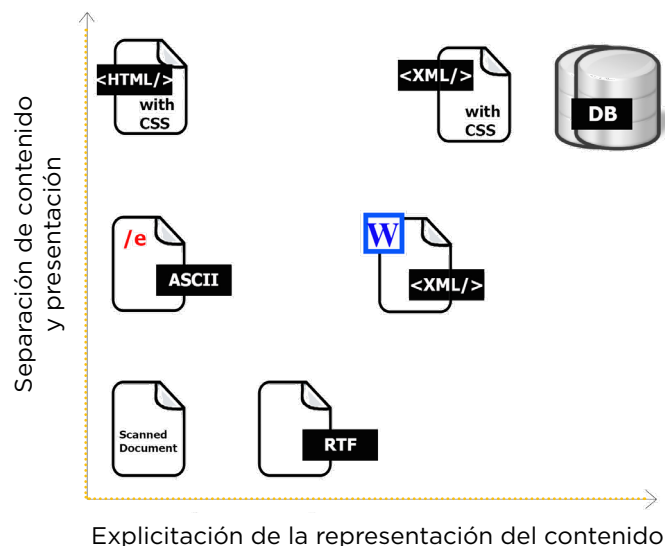


modificar la gravedad para que puedan, por ejemplo, rebotar contra las paredes. Sofisticados formatos de documentos digitales pueden permitir la interacción con imágenes digitales o videos, gráficos en 3-D o conjuntos de datos incrustados. El “Google Art Project” contiene fotografías de altísima resolución de pinturas famosas que permiten ver detalles que son indetectables en las condiciones normales de visualización en los museos.

No obstante, conviene recordar que las representaciones digitales de los recursos físicos también pueden perder información y capacidades importantes. Los sonidos distintivos de la música hip hop producidos al “rascar” los discos de vinilo en los tocadiscos no pueden producirse a partir de archivos de música MP3 en formato digital.

Imagen 2.3. El Coeficiente de inteligencia en información (Information IQ)

La noción de “Coeficiente de inteligencia” o CI de la información capta la idea de que los formatos de los documentos difieren en dos dimensiones: la explicitación de la representación del contenido y la separación del contenido y la presentación. Un documento escaneado es sólo una imagen de un documento con ninguna de estas distinciones, por lo que es bajo en ambas dimensiones. Una base de datos o un documento XML distingue explícitamente entre los tipos de contenido y la presentación se asigna por separado, por lo que tiene valores altos en ambas dimensiones y tiene el mayor “coeficiente intelectual” en información. Las distinciones de contenido de un documento HTML suelen ser de presentación y, por lo tanto, tiene un CI más bajo. Los formatos con alto CI de la información facilitan el procesamiento informático.





Los derechos de autor a menudo presentan una barrera a la digitalización, tanto por los impedimentos legales a la reproducción o distribución, como porque la propia digitalización permite la aplicación del derecho de autor en un grado inviable antes de la llegada de la digitalización, eliminando algunas formas comunes de acceso e interacciones que son inherentemente posibles con los libros impresos físicos, como la capacidad de darlos o venderlos a otra persona.

2.2.3 LA AGENCIA DE UN RECURSO

La agencia es la medida en que un recurso puede **iniciar acciones por sí mismo**. Podemos definir un continuo entre recursos completamente pasivos que no pueden iniciar ninguna acción y recursos activos que pueden iniciar acciones basadas en la información que perciben de su entorno o que obtienen mediante interacciones con otros recursos. Un libro que se lee en la playa se calentará al absorber la energía del sol, pero no tiene forma de medir su temperatura y es un recurso completamente pasivo. Un termómetro de mercurio ordinario detecta y muestra la temperatura, pero no es capaz de comunicarla, mientras que un termómetro digital inalámbrico o una estación meteorológica pueden hacerlo.

Los recursos pasivos sirven como sustantivos u operandos sobre los que se actúa. Por el contrario, los recursos activos sirven como verbos u operantes que causan y realizan acciones. Se necesita el concepto de agencia para incorporar en el marco conceptual de los Sistemas de Organización a los recursos que son fuentes activas de información, o que tienen un carácter computacional. Este concepto también nos permite incluir los recursos vivos, o más específicamente, a los recursos humanos, en el debate sobre los Sistemas de Organización de una manera más general, lo que resalta su agencia.

LAS PERSONAS COMO RECURSOS

Las ediciones anteriores de este libro eludieron la cuestión de considerar a las personas como recursos para evitar complicar la disciplina de la organización. La gente se organiza de innumerables maneras para coexistir, compartir conocimientos y lograr más de lo que podrían como individuos. Conductas como la confianza y la reciprocidad podrían considerarse “principios de organización” para la sociedad humana. Pero estas relaciones e interacciones orgánicas usualmente carecen de la organización intencional nece-



saría para ser consideradas verdaderos Sistemas de Organización, excepto cuando la gente se enmarca en “comunidades intencionales”.

Sin embargo, las personas sí se califican como recursos en los Sistemas de Organización según nuestra definición: al igual que los recursos digitales y físicos, los recursos humanos pueden ser identificados, categorizados, descritos en términos de sus atributos y relaciones, y participan en interacciones para crear valor. En las empresas, las personas se organizan para amplificar sus habilidades, conocimientos y agencia. El organigrama de una empresa suele ser una jerarquía formal en la que el papel de cada trabajador se define a través de sus responsabilidades y sus relaciones con los demás miembros de la empresa. Tratar a un empleado de forma abstracta como un recurso con funciones, entradas y salidas específicas y predecibles permite que los empleados o los procesos dependan unos de otros sin que se distraigan con los detalles del trabajo de los demás. Este enfoque de “caja negra” (black-boxing) puede fomentar la especialización y permitir que una organización funcione de manera más eficiente.

2.2.3.1 RECURSOS PASIVOS O RECURSOS OPERANDOS

Los Sistemas de Organización que contienen recursos pasivos o “recursos operandos” son omnipresentes por la sencilla razón de que vivimos en un mundo de recursos físicos que identificamos y nombramos para interactuar con ellos. Los recursos pasivos suelen ser tangibles y estáticos y, por lo tanto, se vuelven valiosos sólo como resultado de alguna acción o interacción con ellos.

La mayoría de los Sistemas de Organización con recursos físicos o los que contienen recursos que son sus equivalentes digitalizados tratan a esos recursos como pasivos. Un libro impreso en una estantería de una biblioteca, un libro digital en un lector de libros electrónicos, una estatua en una sala de un museo o una caja de cerveza en un refrigerador de un supermercado sólo crean valor cuando se sacan, se ven o se consumen. Ninguno de estos recursos muestra ninguna “agencia” y no puede iniciar ninguna acción para crear valor por sí mismo.



2.2.3.2 RECURSOS ACTIVOS, O RECURSOS OPERANTES

Los recursos activos crean efectos o valor por sí mismos, a veces cuando interactúan con los recursos pasivos. Los recursos activos pueden ser personas, otros recursos vivos, agentes informáticos, fuentes activas de información, servicios basados en la web, coches que se conducen solos, robots, aparatos, máquinas u otros objetos ordinarios como bombillas, paraguas y zapatos que se han hecho “más inteligentes”. Podemos explotar la capacidad de computación, la capacidad de almacenamiento y el ancho de banda de comunicaciones para crear recursos activos que puedan hacer cosas y apoyar interacciones que son imposibles para los recursos pasivos físicos ordinarios.

Podemos analizar los recursos activos de acuerdo con cinco capacidades que aumentan progresivamente su “agencia”: sensibilidad, activación, conectividad, programabilidad, componibilidad. Estas capacidades se complementan entre sí para dar a los recursos y a los Sistemas de Organización en los que participan más formas de crear valor a través de interacciones e intercambios de información.

SENSIBILIDAD (AWARENESS)

- La capacidad mínima para que un recurso tenga un cierto grado de agencia es que sea capaz de percibir o ser consciente de algún aspecto de su entorno o de sus interacciones con otros recursos. Un termómetro mide la temperatura, un fotodetector mide la luz, un medidor mide el combustible que queda en el depósito de gasolina de un coche, un dispositivo GPS calcula su ubicación después de detectar y analizar las señales de los satélites, un sensor *wearable* de parámetros físicos rastrea los latidos del corazón y la distancia que se camina. Pero captar algo del entorno, en sí mismo, no crea ningún valor en un Sistema de Organización: es necesario que además haga algo.

ACTIVACIÓN (ACTUATION)

- Un recurso tiene la capacidad de activación cuando puede crear efectos o valor iniciando alguna acción como resultado de la información que percibe; “activador” (actuador) se utiliza a menudo para describir un recurso que puede mover o controlar un mecanismo o sistema físico, mientras que “efector” se utiliza cuando el recurso es biológico. Los recursos pueden actuar encendiendo luces, altavoces, cámaras, motores, interruptores, enviando un mensaje sobre el estado o el valor de un sensor, o moviéndose por sí mismos (como en el caso de los robots).



Se crea una actuación potencial o latente cuando un recurso puede mostrar o emitir algún aspecto de su estado, pero sólo se crea valor si otra fuente (posiblemente humana) ve la pantalla o escucha la emisión y luego actúa sobre ella.

Por ejemplo, los chips de RFID, que son esencialmente códigos de barras con transpondedores de radio incorporados, pueden ser adjuntados a recursos que de otra manera serían pasivos para hacerlos activos. Los chips de RFID comienzan a transmitir cuando detectan la presencia de un dispositivo de lectura de RFID. Esto permite el seguimiento automatizado de la ubicación y la detección del contexto. Los receptores de RFID están incorporados en líneas de montaje, muelles de carga, aparcamientos, cabinas de peaje o estanterías de tiendas para detectar cuando algún recurso marcado con RFID se encuentra en algún lugar significativo. Las etiquetas RFID pueden ser más útiles si además se les hace registrar y transmitir la información de los sensores que detectan la temperatura, la humedad, la aceleración, e incluso la contaminación biológica.

CONECTIVIDAD (CONNECTIVITY)

■ Para que un recurso activo realice un trabajo útil debe estar conectado de alguna manera a un mecanismo de actuación que manipula o controla algún otro recurso. Esta conexión puede ser directa y permanente entre el recurso y la cosa que acciona, como la de un termostato cuya capacidad de detección de temperatura tiene una conexión fija con un sistema de calefacción o refrigeración que apaga o enciende dependiendo de la temperatura.

Una innovación importante en el diseño de los recursos activos es “envolver” los recursos físicos con software para que se les pueda dar direcciones IP y hacer conexiones con protocolos de Internet, lo que les permite enviar información a una aplicación externa con mayores capacidades de actuación. Se dice que esos recursos forman parte de la “Internet de las cosas”.

Los teléfonos inteligentes son recursos activos que pueden identificar y compartir su propia ubicación, orientación, aceleración y un número creciente de otros parámetros contextuales para permitir la personalización de los servicios de información. Los teléfonos inteligentes también pueden ejecutar aplicaciones que reciben y envían mensajes a otros recursos inteligentes para supervisar y optimizar su funcionamiento.



COMPUTABILIDAD O PROGRAMABILIDAD (COMPUTATION OR PROGRAMMABILITY)



Los recursos activos simples operan de manera determinista: dada la lectura del sensor, haz esto. Otros recursos activos tienen capacidades computacionales que les permiten analizar la información actual e histórica de sus sensores, identificar valores o patrones de datos significativos en estos recursos de interacción, y luego adaptar su comportamiento.

Muchos termostatos son programables, pero la mayoría de las personas no se molestan en programarlos, por lo que pierden el potencial de ahorro de energía. Nest Labs fabrica un termostato de aprendizaje que se programa a sí mismo. El termostato Nest utiliza sensores de temperatura, humedad, movimiento y luz para saber si la gente está en casa, y una conexión Wi-fi para obtener datos meteorológicos locales.

El robot aspirador Roomba navega alrededor de los muebles, cables de alimentación, escaleras, y optimiza sus rutas de limpieza para pasar por lugares particularmente sucios. Pero aspirar es todo lo que hace. Los robots más sofisticados están diseñados para ser versátiles y adaptables, de modo que pueden realizar repetidamente cualquier tarea necesaria en un proceso de fabricación y sus capacidades pueden actualizarse continuamente mediante actualizaciones de software, al igual que las aplicaciones de su teléfono inteligente. Una nueva generación de robots “tipo Baxter” (nombre tomado de uno de sus modelos) puede ser entrenada a través de observar un ejemplo; una persona mueve los brazos y las manos de Baxter para mostrarle lo que debe hacer, y cuando Baxter se ha programado para repetirlo, asiente con la cabeza.

COMPONIBILIDAD Y COOPERATIVIDAD (COMPOSABILITY AND COOPERATION)



Los recursos activos más “inteligentes” pueden hacer más que analizar la información que recogen y adaptar lo que hacen. Además, exponen a otros recursos lo que conocen y pueden activar a otros recursos, haciendo que hagan, utilizando protocolos y formatos estándar no-propietarios. Esto significa que los recursos activos que se diseñaron e implementaron de manera independiente pueden trabajar juntos para crear valor.

Muchos Sistemas de Organización en la web consisten en colecciones o composiciones de recursos digitales activos. Las interacciones entre estos recursos activos a menudo implementan modelos



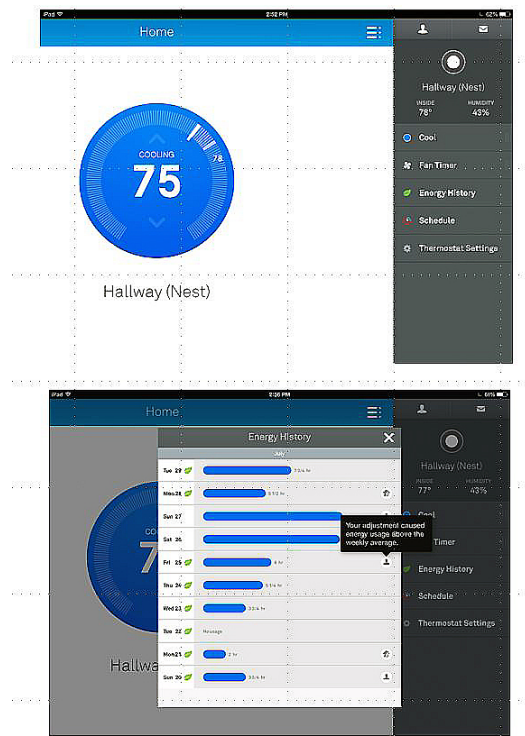
de negocio que hacen uso intensivo de información, en los que el valor se crea mediante el intercambio, la manipulación, la transformación o el procesamiento de información, en lugar de la manipulación, la transformación o el procesamiento de recursos físicos.

Estamos empezando a ver los mismos principios de modularidad y compatibilidad aplicados a los recursos físicos, con bibliotecas de software de código abierto para el uso de sensores y microcontroladores y API fáciles de usar. En esencia, estamos utilizando el software y los recursos físicos a la manera de bloques de construcción funcionales y los estándares serán de vital importancia.

Recursos activos: El “ecosistema” del termostato Nest

Estas dos capturas de pantalla de la aplicación del iPad Nest muestran el panel de control del termostato y un informe del historial de energía con una nota emergente que explica que el reajuste de la temperatura dio lugar a un uso de energía superior a la media en ese día. El termostato sirve como un dispositivo central, comunicándose con las luces, los electrodomésticos, las alarmas de humo, tu coche, tu ropa de deporte y otros recursos activos (<https://nest.com/works-with-nest/>).

(Capturas de pantalla de Andrea Angquist. Usadas con permiso).





Los patrones de diseño para componer Sistemas de Organización a partir de recursos físicos “inteligentes” están progresivamente aplicándose para conseguir una “casa inteligente”, un “edificio de oficinas inteligente” o una “ciudad inteligente”. Se están realizando muchos experimentos y están surgiendo nuevos productos que están probando diferentes combinaciones de hardware y software para entender los equilibrios o compromisos de diseño entre los elementos implicados, para determinar mejor dónde deberían situarse los dispositivos “inteligentes”. Por ejemplo, podemos comparar una “casa inteligente” construida alrededor de un dispositivo central súper inteligente que se comunica y coordina con muchos otros dispositivos “no tan inteligentes” del mismo fabricante, con una en la que todos los dispositivos son igualmente inteligentes y provienen de diferentes fabricantes.

A escalas más complejas, un edificio verdaderamente inteligente no sólo tendrá termostatos programables para controlar los sistemas de calefacción y refrigeración. Tomará en cuenta los pronósticos meteorológicos, calendarios de viaje, información sobre el costo de la electricidad de diferentes fuentes y otra información relevante que será usada como input en un modelado informático de cómo el edificio se calienta y se enfría para optimizar el uso y el costo de la energía mientras se mantienen las habitaciones a temperaturas adecuadas.

Las interfaces de aplicación (APIs) estandarizadas permiten que los recursos activos interactúen con personas para obtener información que de otro modo podría provenir de los sensores o que aumenta el valor de la información de los sensores. Un termostato programable que puede registrar las preferencias temporales de las personas que utilizan el espacio controlado por el termostato es más capaz que uno con un solo umbral de temperatura. Un protocolo estándar de Internet para comunicarse con el termostato permitiría controlarlo a distancia.

Los formatos de datos y los protocolos de comunicación abiertos y normalizados permiten la agregación y el análisis de información de muchos casos del mismo tipo de recurso activo. Por ejemplo, los teléfonos inteligentes que funcionan con la aplicación Google Maps transmiten información sobre su velocidad y ubicación. El aprendizaje automático y sofisticadas técnicas de optimización de este conjunto de datos pueden dar lugar a una inteligencia colectiva que se puede devolver a los recursos de los que se ha obtenido. En este caso, Google puede identificar atascos y generar rutas alternativas para los conductores atascados en el tráfico.

Pero no todo se puede hacer mejor con los ordenadores. La web ha permitido el uso de personas como recursos activos para llevar a cabo



tareas de corta duración que pueden ser descritas con precisión pero que no pueden ser realizadas de forma fiable por las computadoras. Estas tareas a menudo requieren de un juicio estético o subjetivo. Las personas que realizan estas tareas basadas en la web son a menudo llamadas “Turcos Mecánicos” (Mechanical Turks) por analogía con una falsa máquina de jugar al ajedrez del siglo XVIII que tenía un humano escondido en su interior que movía las piezas con destreza.

SI OCURRE ESTO, ENTONCES HAZ ESO

IFTTT (“If this, then that” en inglés) es un sistema de programación visual que permite a personas que no son programadoras conectar y controlar recursos activos en los mundos físico y digital. Los programas de IFTTT, llamados recetas, pueden coger información de una creciente variedad de servicios de Internet (fecha/hora, calendario, tiempo, noticias, correo electrónico, medios sociales y muchos otros) y usar esta información con una lógica sencilla de control para activar acciones en otros servicios o recursos. Las recetas de ejemplo pueden copiar una foto de Instagram a Google Drive, añadir datos de Fitbit diarios a una hoja de cálculo o controlar las luces en función de la hora, el tiempo o la puesta de sol.

“Si se pone el sol, enciende el enchufe.” El icono de la izquierda es el disparador; el icono de la derecha es la acción.

(Fotografía de R. Glushko)



2.2.4 EL FOCO DEL RECURSO

Un cuarto aspecto para diferenciar entre tipos de recursos distingue los recursos originales o primarios de los recursos que los describen. Cualquier recurso primario puede tener uno o más recursos descriptivos asociados para facilitar su búsqueda, la interacción o su interpretación. Los recursos que son descripciones son esenciales para organizar sistemas en los que los recursos primarios no están bajo su control y sólo se puede acceder a ellos o interactuar con ellos a través de la descripción. Los recursos de descripción suelen denominarse **metadatos**.



La distinción entre recursos primarios y recursos descriptivos, o metadatos, está profundamente arraigada en la biblioteconomía y en los Sistemas de Organización tradicionales cuyas colecciones son predominantemente recursos de texto como libros, artículos u otros documentos. En estos contextos, los recursos de descripción se denominan comúnmente recursos bibliográficos o catálogos, y cada recurso primario suele estar asociado a uno o más recursos que los describen de forma compacta.

En las empresas, los Sistemas de Organización de los recursos de información digital, como los documentos comerciales o los registros de datos creados por transacciones o procesos automatizados, casi siempre emplean recursos que describen o están asociados a grandes conjuntos o clases de recursos primarios.

El contraste entre los recursos primarios y los recursos descriptivos es muy útil en muchos contextos, pero cuando se examinan más ampliamente los Sistemas de Organización, a menudo es difícil distinguirlos y determinar qué recursos son primarios y cuáles son metadatos. A menudo optar por reconocerlos como de uno u otro tipo es una decisión sobre qué aspecto centra nuestra atención en un determinado momento o situación.

Por ejemplo, muchos usuarios de Twitter tratan el cuerpo del mensaje de 140 caracteres como el recurso primario, mientras que los metadatos asociados con el mensaje y el remitente (si es un reenvío, una respuesta, un enlace, etc.) son menos importantes. Sin embargo, en el caso de las empresas que utilizan los metadatos de Twitter para medir el impacto del emisor y de la marca, o para identificar redes sociales y las tendencias, el centro de atención son los metadatos, no el contenido.

Otro ejemplo sería el de los jugadores de equipos deportivos profesionales: se trata de recursos humanos, pero millones de personas participan en ligas de deportes de fantasía en las que los equipos consisten en recursos basados en las estadísticas generadas por los jugadores humanos reales. Dicho de otro modo, los recursos asociados o secundarios en los deportes reales se tratan como recursos primarios en las ligas virtuales.

PÁRATE A PENSAR: LA “INTERNET DE LAS COSAS”

Hay mucho bombo y platillo en la Internet de las Cosas (IoT), pero también hay mucha innovación. Si buscas la frase “Internet de las cosas” junto con casi cualquier recurso físico, lo más probable es que encuentres algo, prueba con “bebé”, “perro”, “tenedor”, “lechuga”, “pijama”, “farol” y luego con otros términos de tu propia cosecha.



2.2.5 RECURSOS SEGÚN FORMATO Y FOCO

La aplicación combinada de la contraposición de formato entre los recursos físicos y digitales y la distinción entre recursos primarios y descriptivos produce un marco o framework que puede ser útil, con cuatro categorías de recursos (Imagen 2.4, Formato de recursos x Foco).



Imagen 2.4. Formato de los recursos x Foco.

Las distinciones de formato de los recursos y el foco de los recursos se combinan para obtener cuatro categorías diferentes de recursos: recursos físicos, recursos digitales, descripciones físicas y descripciones digitales.

2.2.5.1 LA DESCRIPCIÓN FÍSICA DE UN RECURSO PRIMARIO FÍSICO

La relación más antigua entre los recursos descriptivos y los recursos físicos es aquella en que las descripciones u otra información sobre los recursos físicos están a su vez codificadas en una forma física. Hace casi diez mil años, en Mesopotamia, pequeñas fichas de arcilla guardadas en recipientes de arcilla servían como información de inventario para contar unidades de bienes o ganado. Se tardó 5000 años en evolucionar hacia la escritura cuneiforme, en la que las marcas de arcilla representaban las gallinas y hacían innecesarios tanto las fichas como los recipientes.



Las fichas impresas sirvieron como recursos de descripción física para los libros en las bibliotecas durante casi dos siglos. Permitían insertar nuevas fichas dentro del orden alfabético de cada catálogo, según puntos de acceso diferentes (autor, título, materia), y distribuir las en cajones según creciera el inventario de recursos.



Un documento cuneiforme en el Pergamon

El Museo de Pérgamo en Berlín contiene una gran colección de artefactos babilónicos, persas y asirios que tienen casi tres mil años de antigüedad, incluyendo numerosas tablillas de arcilla cuneiforme como esta.

(Fotografía de R. Glushko.)

2.2.5.2 DESCRIPCIÓN DIGITAL DE UN RECURSO PRIMARIO FÍSICO

En este caso, el recurso digital describe un recurso físico. El ejemplo más familiar de esta relación es el catálogo en línea de una biblioteca, utilizado para encontrar la ubicación de los recursos físicos de la biblioteca, los cuales, a partir de la década de 1960, fueron reemplazando a las tarjetas físicas por registros de una base de datos. Los catálogos en



línea de los museos suelen contener una fotografía digital de la pintura, la escultura u otro objeto del museo que aparece descrito en una ficha con datos básicos que lo identifican y contextualizan.

Los recursos descriptivos digitales de recursos físicos primarios son esenciales en la gestión de la cadena de suministro, la logística de venta al por menor, el transporte y todos los modelos de negocio que dependen de tener información oportuna y precisa sobre dónde están las cosas o sobre su estado actual. Este recurso de descripción digital se crea como resultado de una interacción con un recurso físico primario como un sensor de temperatura o con algún recurso físico secundario que ya está asociado con el recurso físico primario como una etiqueta RFID o un código de barras.

Los sistemas de realidad aumentada combinan una **capa de información digital en tiempo real** sobre algún objeto físico con una vista digital o una representación del mismo. Las líneas amarillas de “*first down*” superpuestas sobre la imagen real en las retransmisiones de los partidos de fútbol americano son un ejemplo familiar. Las técnicas de realidad aumentada que superponen metadatos identificativos o descriptivos se utilizan en pantallas para apoyar la operación o el mantenimiento de equipos complejos, en guías de navegación de teléfonos móviles y guías turísticas, en la publicidad y en otros ámbitos en los que, de otro modo, los usuarios podrían tener que consultar una fuente de información separada. La tecnología avanzada de las cabinas de los aviones incluye avisos y ayudas proyectadas sobre el propio cristal frontal, que presentan datos críticos basados en la instrumentación disponible, incluidas luces de señalización de la pista aterrizaje generadas por realidad aumentada cuando la visibilidad es escasa por nubosidad o niebla.

Las pantallas de realidad aumentada han sido incorporadas recientemente a una tecnología *wearable* como Google Glass, que se monta en los marcos de las gafas para mostrar información obtenida de Internet después de ser solicitada por comandos de voz. Algunas marcas de coches de lujo han incorporado una tecnología similar para proyectar los datos del tablero de mandos, las condiciones del tráfico y otras indicaciones en la luna del coche.





Compras con código de barras en un supermercado virtual

Woolworth's Australia creó un “supermercado virtual” con fotos de los productos y códigos de barras. El escaneo realiza un pedido, que se entrega desde el supermercado local del cliente.

(Fotografía de R. Glushko)

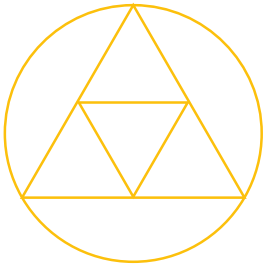
2.2.5.3 DESCRIPCIÓN DIGITAL DE UN RECURSO PRIMARIO DIGITAL

Un recurso digital que describe otro recurso digital. Se trata de la relación presente en una biblioteca digital o en cualquier Sistema de Organización basado en la web, que permite acceder a un recurso digital primario directamente desde el recurso digital secundario.

2.2.5.4 DESCRIPCIÓN FÍSICA DE UN RECURSO PRIMARIO DIGITAL

Esta es la relación que se implementa cuando encontramos un código de barras QR incrustado en anuncios de periódicos o revistas, en vallas publicitarias, aceras, camisetas o en las estanterías de las tiendas. Escanear el código QR con la cámara de un teléfono móvil puede lanzar un sitio web que contenga información sobre un producto o servicio, hacer un pedido de una unidad del artículo señalado en un catálogo web, marcar un número de teléfono o iniciar otra aplicación o servicio identificado por el código QR.





2.3 LA IDENTIDAD DE LOS RECURSOS

Las tareas esenciales al construir cualquier Sistema de Organización son las siguientes: determinar la identidad de los recursos que pertenecen a un dominio, decidir qué propiedades son importantes o relevantes para las personas o sistemas que operan en ese dominio y luego especificar los principios por los que esas propiedades encapsulan o definen las relaciones entre los recursos. En los Sistemas de Organización utilizados por personas o con un alcance reducido, los métodos para realizar esas tareas suelen ser ad hoc y poco sistemáticos, por lo que los Sistemas de Organización son idiosincrásicos y no escalan bien. En el otro extremo, los Sistemas de Organización diseñados para su uso institucional o en todo un sector, especialmente en los ámbitos de uso intensivo de información, requieren métodos de diseño sistemático para determinar qué recursos tendrán identidades separadas y cómo se relacionan entre sí. Esos recursos y sus relaciones se describen luego en modelos conceptuales que orientan la aplicación de los sistemas que gestionan los recursos y apoyan las interacciones con ellos.

2.3.1 IDENTIDAD EN LOS RECURSOS FÍSICOS

Nuestros sistemas visuales y cognitivos humanos hacen un trabajo notable en la selección de objetos distinguiéndolos de su entorno y entre sí. De hecho, tenemos poca dificultad para reconocer un objeto o una persona incluso si los vemos desde una distancia y un ángulo de visión novedosos o con diferente iluminación, sombras, etc. Cuando vemos un partido de fútbol, no tenemos ningún problema para distinguir a los jugadores que se mueven por el campo y el contraste de los colores de sus uniformes nos permiten ver que hay dos equipos diferentes.

Los mecanismos perceptivos que nos hacen ver las cosas como objetos permanentes con propiedades visibles contrastantes son sólo el requisito previo para las tareas de organización que pasan por la identificación del objeto específico, la determinación de las categorías de objetos a las que pertenece y la decisión de cuál de esas categorías es apropiada destacar. La mayoría de las veces realizamos estas tareas de manera automática e inconsciente; otras veces tomamos decisiones conscientes sobre ellas. Para algunos propósitos consideramos un equipo deportivo como un único recurso, para otros, los tomamos como una colección de jugadores separados, como delanteros y defensas, como titulares y reservas, y así sucesivamente.



Aunque tenemos muchas opciones sobre cómo podemos organizar a los jugadores de fútbol, todas ellas incluirán el concepto de un solo jugador como el recurso identificable más pequeño. Nunca vamos a pensar en un jugador de fútbol como una colección intencional de recursos de piernas, brazos, cabeza y cuerpo identificados por separado, porque no hay otras formas de “ensamblar” un humano a partir de partes del cuerpo. En términos generales, existen algunas limitaciones naturales en la organización de la materia en partes o colecciones basadas en tamaños, formas, materiales y otras propiedades que nos hacen identificar algunas cosas como recursos indivisibles en algún dominio.



2.3.2 IDENTIDAD Y RECURSOS BIBLIOGRÁFICOS

Reflexionar sobre la cuestión de la identidad es algo relativamente reciente en el mundo de los bibliotecarios y catalogadores. Las bibliotecas existen desde hace unos 4000 años, pero hasta los últimos cientos de años los bibliotecarios crearon “contenedores” de encabezamientos y temas para organizar los recursos sin molestarse en dar a cada artículo individual un identificador o nombre separado. Esto significaba que quienes buscaban primero tenían que hacer una conjetura según su nivel de formación y conocimiento del entorno, sobre cuál era el recipiente que contenía la información deseada: ¿“Historia”? “Filosofía Médica y Química”... y luego repasar toda la categoría en busca del artículo deseado. Las elecciones eran ad hoc y siempre locales, es decir, cada registrador de datos decidía los contenedores y las agrupaciones de cada catálogo.



Las distinciones establecidas por Panizzi, Lubetzky, Svenonius y otros teóricos de la biblioteconomía han evolucionado hoy en día en un modelo abstracto a cuatro niveles (véase la imagen 2.5, El modelo FRBR) entre la obra abstracta, una expresión en múltiples formatos o géneros, una manifestación particular en uno de esos formatos o géneros y un elemento físico específico.

Si volvemos a la pregunta “¿Qué es eso a lo que llamamos *Macbeth*?” podemos ver cómo las diferentes formas de respuesta encajan en esta jerarquía de abstracción. La respuesta más específica es que “*Macbeth*” es un elemento específico, un recurso muy particular e individual, como ese libro en rústica con páginas marcadas en amarillo que poseías cuando leías *Macbeth* en el instituto. Una respuesta más abstracta es que *Macbeth* es una idealización llamada obra, una categoría que incluye todas las obras, películas, ballets u otras creaciones intelectuales que comparten una cantidad reconocible de la trama y el significado de la obra original de Shakespeare.

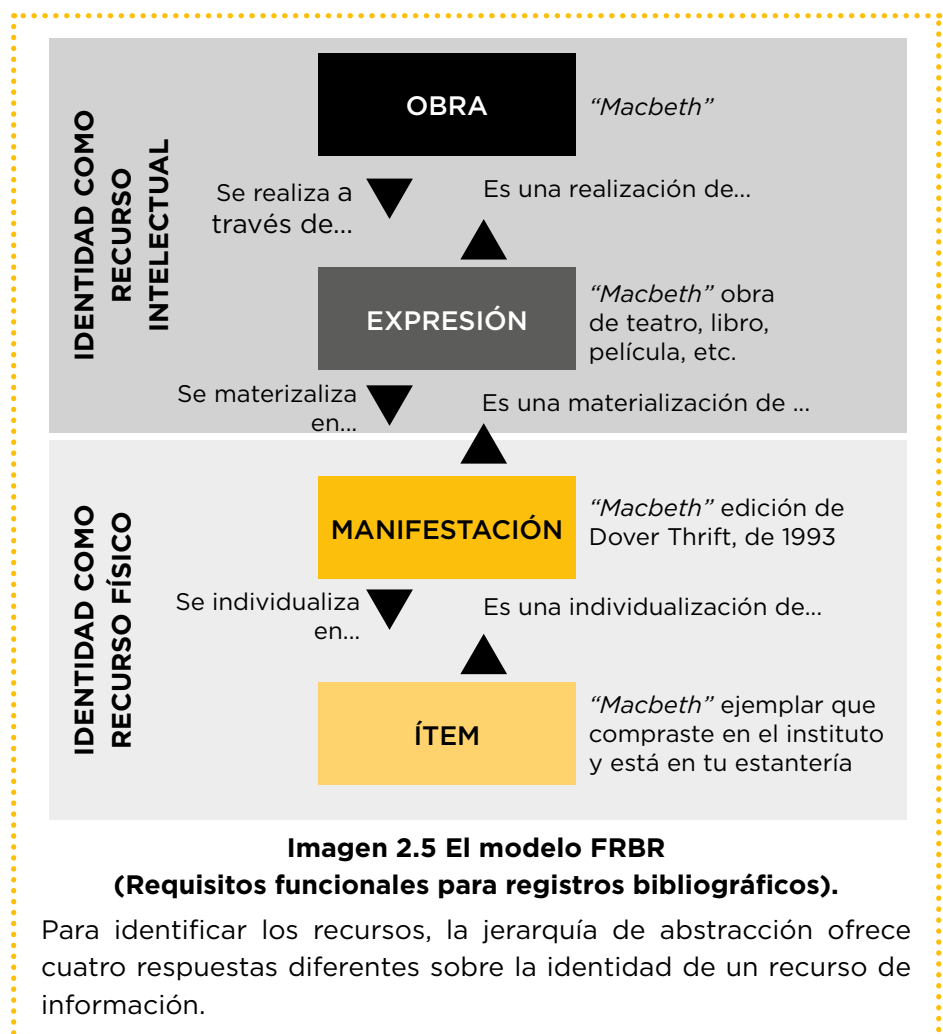


Imagen 2.5 El modelo FRBR

(Requisitos funcionales para registros bibliográficos).

Para identificar los recursos, la jerarquía de abstracción ofrece cuatro respuestas diferentes sobre la identidad de un recurso de información.



2.3.3. IDENTIDAD Y COMPONENTES INFORMATIVOS

En los ámbitos de uso intensivo de información, los documentos, las bases de datos, las aplicaciones informáticas u otros depósitos o fuentes de información explícitas son omnipresentes y esenciales para la creación de valor para el usuario, el lector, el consumidor o el cliente. El valor se crea mediante la comparación, la compilación, la coordinación o la transformación de la información en alguna secuencia o articulación de procesos que operan con la información que fluye de una fuente o proceso de información a otro. Estos procesos se emplean en contabilidad, los servicios financieros, las adquisiciones, la logística, la gestión de la cadena de suministro, la contratación de seguros y la tramitación de reclamaciones, los servicios jurídicos y profesionales, el soporte al cliente, la programación informática y la gestión de la energía.

Los procesos que crean valor en los dominios intensivos en información están “pegados” por los **componentes de información compartida** que se intercambian en documentos, registros, mensajes o descripciones de recursos de algún tipo. Los componentes de información son los recursos primitivos y abstractos en estos contextos intensivos en información. **Son las unidades de significado que sirven como bloques de construcción de descripciones compuestas y otros artefactos de información.**

Los procesos de creación de valor en los dominios que hacen un uso intensivo de información funcionan mejor cuando sus partes componentes provienen de un vocabulario común controlado para los componentes, o cuando cada uno utiliza un vocabulario con una granularidad y precisión semántica compatible con los demás. Por ejemplo, el valor creado por un historial clínico personal surge cuando la información de los médicos, las clínicas, los hospitales y las compañías de seguros puede combinarse porque todos ellos comparten el mismo componente “paciente” como una pieza lógica de información.

Esta definición abstracta de los componentes de la información no ayuda a identificarlos, por lo que introduciremos algunos criterios heurísticos: Un componente de información puede ser:

1 // Cualquier pieza de información que tenga una etiqueta o identificador único o

2 // Cualquier pieza de información que sea autónoma y comprensible por sí misma.



Décadas de esfuerzos prácticos y teóricos en la elaboración de modelos conceptuales, la teoría relacional y el diseño de bases de datos han dado lugar a métodos rigurosos para identificar los componentes de información cuando los requisitos y las reglas de negocio para la información pueden especificarse con precisión. Por ejemplo, en el ámbito de las transacciones comerciales, la información requerida, como códigos de artículo, cantidades, precios, pagos, etc., debe codificarse como un tipo particular de datos -número entero, decimal, cadena Unicode, etc.- con valores posibles claramente definidos y que siga reglas de ocurrencia claras.

Identificar componentes puede parecer aparentemente fácil en el extremo transaccional del espectro de tipos de documentos (véase el recuadro en §2.2.1 Dominio de los recursos), como es el caso de los pedidos o facturas, formularios que requieren la introducción de datos u otros tipos de documentos muy estructurados como los catálogos de productos, en los que las piezas de información suelen estar etiquetadas y delimitadas por recuadros, líneas, espacios en blanco u otras características de presentación que codifican las distinciones entre los tipos de contenido. Por ejemplo, la presencia de etiquetas para ARTÍCULO, NOMBRE DEL CLIENTE, DIRECCIÓN y INFORMACIÓN DE PAGO en los campos de un formulario de pedido en línea sugiere que estos elementos de información son componentes semánticamente distintos dentro de una aplicación de comercio online. Además, estas etiquetas pueden tener análogos en nombres de variables en el código fuente que implementa el formulario de pedido, o como etiquetas en un documento XML creado por la aplicación de pedidos; `<NombreCliente>John Smith</NombreCliente>` y `<Ítem>A-19</Item>` en el documento de pedido puede ser fácilmente identificado cuando se envía a otros servicios por la aplicación de gestión de pedidos.

Pero los métodos teóricos para identificar componentes, como los mencionados de la teoría relacional y la normalización, que funcionan para los datos estructurados no se aplican estrictamente cuando los requisitos de información son más cualitativos y menos precisos, en el extremo narrativo del espectro de tipos de documentos. Estos requisitos de información son típicos de los tipos de documentos discursivos, no estructurados y semiestructurados, y de las fuentes de información que son típicas en el campo del derecho, la educación y los servicios profesionales. Los documentos narrativos incluyen publicaciones técnicas, informes, políticas, procedimientos y otra información menos estructurada, en los que los componentes semánticos rara vez se etiquetan de forma explícita y a menudo están rodeados de texto más genérico. A diferencia de los documentos transaccionales que dependen de una semántica precisa porque son utilizados por los ordenado-



res, los documentos narrativos son utilizados por personas que pueden preguntar si no están seguras de lo que significa algo, por lo que hay menos necesidad de definir explícitamente el significado de los componentes de la información. Las excepciones ocasionales, como cuando los componentes de los documentos narrativos se identifican con etiquetas explícitas como NOTA y ADVERTENCIA, sólo confirman la regla general.

2.3.4. IDENTIDAD Y RECURSOS ACTIVOS

Los recursos activos (que trataremos en detalle en el punto §2.4.3.2 del siguiente apartado) inician efectos o crean valor por sí mismos. En muchos casos, un recurso físico inherentemente pasivo, como un paquete de productos o un palé de envío, se transforma en uno activo cuando se asocia con una etiqueta RFID o un código de barras. Los teléfonos móviles contienen identificaciones de dispositivos o de abonados, de modo que cualquier información que comuniquen puede asociarse tanto al teléfono como, a menudo, por referencia indirecta, a una persona concreta. Si el recurso tiene una dirección IP, se dice que forma parte de la “Internet de las cosas”.

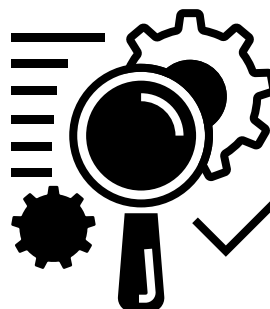
Los Sistemas de Organización que crean valor a partir de recursos activos suelen complementarse o coexistir con Sistemas de Organización que tratan sus recursos como recursos pasivos. En una biblioteca tradicional, los libros reposaban pasivamente en estanterías y requerían que los usuarios leyeran sus lomos para identificarlos. Hoy en día, algunos libros de la biblioteca contienen etiquetas RFID activas que los convierten en fuentes de información dinámicas que se autoidentifican publicando sus propias ubicaciones. Del mismo modo, un supermercado o unos grandes almacenes pueden organizar sus bienes como recursos físicos en los estantes, tratándolos como recursos pasivos; a ese Sistema de Organización tradicional se le superpone uno que utiliza la información de las transacciones en el punto de venta, creada cuando los artículos se escanean en los mostradores de salida para reordenar automáticamente los bienes y reponer el inventario en la tienda donde se vendieron. En algunas tiendas, los estantes contienen sensores que continuamente “hablan con la mercancía” y la información que recogen puede mantener los niveles de inventario e incluso ayudar a prevenir el robo de mercancías valiosas mediante el seguimiento de las mercancías a través de una tienda o almacén. El inventario se convierte en una colección de recursos activos; cada artículo ansioso por anunciar su propia ubicación y listo para realizar su propia venta.



La medida en la que un recurso activo es “inteligente” depende de la capacidad de computación de que disponga para perfeccionar los datos que recoge y comunica. Una gran colección de sensores puede transmitir un torrente de datos capturados que requiere un procesamiento considerable para distinguir los eventos significativos de los que reflejan un funcionamiento normal, y también de los que son valores estadísticos atípicos de valores extraños causados por el ruido aleatorio. Este desafío se hace cualitativamente más difícil a medida que la cantidad de datos crece y alcanza un gran tamaño, porque un evento de uno en un millón podría ser un valor estadístico atípico que puede ser ignorado, pero si hay mil valores atípicos similares en mil millones de lecturas de sensores, este grupo de datos probablemente revele algo importante. Por otro lado, dar a cada sensor la capacidad de refinar sus datos para que sólo comunique información significativa podría hacer que los sensores sean demasiado costosos de instalar.

EL BIG DATA HACE A LOS JUGADORES DE FUTBOL “INTELIGENTES”

El equipo de fútbol de la selección de Alemania, que ganó el Mundial de 2014, aprovechó la sofisticada recopilación y análisis de datos para optimizar la habilidad de los jugadores y el entrenamiento de estrategias. La empresa alemana de software SAP analizó los datos de vídeo de las cámaras del campo que captaron miles de puntos de datos por segundo sobre la posición y el movimiento de los jugadores para identificar las mejoras en los pases y el manejo del balón de los jugadores alemanes y detectar los puntos débiles de los oponentes. La firma alemana de equipamiento deportivo Adidas diseñó tacos con sensores que registran el kilometraje, la posición en el campo y los movimientos. (Norton 2014) y (Reynolds 2014).





2.4. EL NOMBRE DE LOS RECURSOS

Determinar la identidad de una cosa, documento, componente de información, o dato que necesitamos no siempre es suficiente. A menudo necesitamos darle un nombre a ese recurso, una etiqueta que nos ayude a entenderlo y hablar sobre lo que es. Pero nombrar no es sólo una tarea simple que consiste en asignar una secuencia de caracteres. En esta sección, discutiremos sobre por qué ponemos nombres, algunos de los problemas con los nombres y los principios que nos ayudan a nombrar las cosas de manera útil.

2.4.1 ¿QUÉ ES UN NOMBRE?

Cuando un niño nace, sus padres le dan un nombre, a menudo una decisión muy estresante y polémica. Los nombres sirven para distinguir a una persona de otra, aunque los nombres pueden no ser únicos - hay miles de personas llamadas James Smith y María García. Los nombres también, intencionalmente o no, sugieren características o aspiraciones. El nombre que se nos dio al nacer es sólo uno de los nombres con los que se nos identificará durante nuestras vidas. Tenemos apodos, nombres que usamos profesionalmente, nombres que usamos con amigos y nombres que usamos en línea. Nuestros bancos, nuestras escuelas y nuestros gobiernos sabrán quiénes somos gracias a los números que asocian a nuestros nombres. Siempre y cuando sirva para identificarte, tu nombre podría ser cualquier cosa.

Los recursos que no son personas necesitan nombres para que podamos encontrarlos, describirlos, reutilizarlos, referirnos a ellos o enlazarlos, registrar quién es su propietario o interactuar con ellos de otras muchas maneras. En muchos dominios los nombres asignados a los recursos también están influidos o limitados por normas, prácticas de la industria o consideraciones tecnológicas.

2.4.2. LOS PROBLEMAS DE LA DENOMINACIÓN

Darle nombre a cualquier cosa, desde un negocio hasta un concepto o una acción, puede ser un proceso difícil y es posible hacerlo bien o mal. En la siguiente sección se detallan algunos de los principales problemas que plantea la asignación de un nombre a un recurso.



2.4.2.1 EL PROBLEMA DEL VOCABULARIO

Todo lenguaje natural ofrece más de una forma de expresar cualquier pensamiento y, en particular, suele haber muchas palabras que pueden utilizarse para referirse a la misma cosa o concepto. Las palabras que las personas eligen para nombrar o describir las cosas están afectadas por sus experiencias y contexto, por lo que habrá gente que a menudo no estará de acuerdo con las palabras que utilicen otros. Además, las personas suelen sorprenderse un poco cuando esto sucede, porque lo que parece el nombre natural u obvio para una persona no es natural u obvio para otra. Una forma de evitar sorpresas es que las personas cooperen en la elección de los nombres de los recursos, y los arquitectos de información suelen utilizar técnicas de diseño participativo como clasificación de tarjetas (*card sorting*) o de listado libre (*free listing*) para este fin.

En los años 80, en los primeros tiempos del diseño de interfaces de usuario, George Furnas y sus colegas de los Laboratorios Bell realizaron una serie de experimentos para medir en qué medida las personas se ponían de acuerdo al nombrar recursos o funciones. La respuesta corta es “muy poco”. Dejados a nuestro propio criterio, terminamos obteniendo una sorprendente larga lista de nombres para una misma cosa.

En un experimento, se pidió a mil pares de personas que “escribieran el nombre que le darían a un programa que cuenta las actividades interesantes que ocurren en algunas de las principales áreas metropolitanas”. Menos de 12 pares de personas se pusieron de acuerdo en un nombre. Furnas llamó a este fenómeno el problema de vocabulario, concluyendo que ninguna palabra podría ser considerada el “mejor” nombre.

2.4.2.2 HOMONIMIA, POLISEMIA Y LOS FALSOS AMIGOS

A veces la misma palabra puede referirse a diferentes recursos: un “banco” puede ser una institución financiera o conjunto de peces. Cuando dos palabras se escriben igual, pero tienen diferentes significados son homógrafos; si también se pronuncian igual son homónimos. Si los diferentes significados de los homógrafos están relacionados, son polisémicos. También eso está presente entre palabras de apariencia similar entre diferentes idiomas.

Los recursos con nombres homónimos y polisémicos a veces se identifican incorrectamente, especialmente cuando es a través de un procedimiento automatizado que no siempre puede utilizar el contexto o el sentido común para determinar el referente correcto. La polise-



NOMBRES POCO FIABLES: KNOCKIN' ON HEAVEN'S DOOR

Figure 1: Top 25 Representations of "Knockin' On Heaven's Door" [35]

```
Guns N' Roses - Knockin' On Heaven's Door
Guns N' Roses - Knocking On Heavens Door
Guns 'N' Roses - Knockin' On Heaven's Door
Guns N' Roses - Knockin On Heavens Door
Guns N' Roses - Knockin' On Heavens Door
Guns N' roses - knockin on heavens door
Guns N' Roses - Knocking On Heaven's Door
Guns N Roses - Knockin' On Heaven's Door
Guns N Roses - Knockin On Heavens Door
Guns And Roses - Knocking On Heavens Door
Guns Nroses - Knockin On Heavens Door
Guns 'n' Roses - Knockin' On Heaven's Door
Guns N Roses - Knocking On Heavens Door
Guns 'n'Roses - Knockin' On Heaven's Door
Guns 'N Roses - Knockin' On Heaven's Door
Guns & Roses - Knockin' on Heaven's Door
Guns N'roses - Knockin' On Heaven's Door
Guns and Roses - Knockin' On Heaven's Door
Guns'n Roses - Knocking On Heavens Door
Guns 'n' Roses - Knockin' On Heavens Door
Aerosmith - Knocking On Heaven's Door
Guns 'n Roses - Knocking On Heaven's Door
Guns 'n' Roses - Knocking On Heavens Door
Guns N Roses - Knocking On Heaven's Door
```

En 2008, el empleado de recomendaciones musicales de la emisora Last.fm, Richard Jones, recopiló una lista con los 100 descriptores más usados para la versión de Guns N' Roses de la canción de Bob Dylan "Knockin' on Heavens Door". Entre los primeros 21 más usados, la canción se le atribuye, incorrectamente, a la banda Aerosmith.

La homografía puede causar más problemas que la simple homografía, porque los significados que se superponen pueden dificultar apreciar cuando se realizan malas interpretaciones de lo que quiere decir un término. Si una empresa piensa que un "contenedor" es un equipamiento público para depositar la basura y encarga varios de ellos, mientras que otra empresa piensa que un "contenedor" es un depósito de gran volumen que llevan los semirremolques y que se apila en los buques de carga, es posible que su desacuerdo no se descubra hasta que reciba los tipos de contenedores equivocados.

2.4.2.3 NOMBRES CON ASOCIACIONES NO DESEADAS

Los falsos cognados son una categoría especial de palabras cuyo uso produce nombres de poca calidad, y hay muchas historias que relatan errores de comercialización de productos, en las que el nombre o la descripción de un producto se traduce mal, a otros idiomas o culturas, con asociaciones indeseables.



Si bien puede ser tentador quitarle importancia a aquellos sesgos y creencias que desconocemos sobre nombres e identificadores y tomarlos como supersticiones y prácticas inofensivas, sus implicaciones son ubicuas y están lejos de ser benignas. El orden alfabético puede parecer una disposición justa y no discriminatoria de los recursos, pero como es fácil elegir el nombre en la parte superior de una lista alfabética, muchas empresas de servicios seleccionan nombres que empiezan por “A”, “AA” o incluso “AAA” (busque en cualquier directorio de servicios impreso). Una consecuencia de este sesgo es que las personas o recursos con nombres que comienzan con letras al final del alfabeto son discriminados sistemáticamente porque a menudo no llegan a ser considerados al recorrer un listado, o porque se evalúan en el contexto de los recursos que ha aparecido antes alfabéticamente, en lugar de por sus propios méritos.

2.4.2.4 NOMBRES QUE ASUMEN ATRIBUTOS NO PERMANENTES

Muchos recursos reciben nombres basados en atributos que pueden resultar problemáticos más adelante si el atributo cambia de valor o de interpretación.

A menudo se hace referencia a los recursos web utilizando una URL que contiene el nombre de dominio del servidor en el que se encuentra, seguido de la ruta del directorio y el nombre del archivo en el ordenador que ejecuta el servidor. Esto trata la ubicación actual del recurso como si fuera su nombre, por lo que el nombre cambiará si el recurso se mueve de sitio. Esto significa que los recursos que son idénticos en contenido, como los de una copia de archivo o de un sitio web “mirror”, tendrán nombres diferentes al original aunque sean copias exactas. Un problema análogo es al que enfrentan los restaurantes o negocios con nombres de calles o números en sus nombres si pierden sus contratos de arrendamiento o quieren expandirse a otros lugares.

Algunos recursos web dinámicos que son generados por programación, tienen URIs que contienen información sobre la tecnología del servidor utilizada para crearlos. Cuando la tecnología cambia, las URIs ya no funcionan.

Algunos recursos tienen nombres que contienen fechas, años u otros indicadores de tiempo, la mayoría de las veces para señalar el futuro. El estudio cinematográfico llamado “20th Century Fox” tomó ese nombre en la década de 1930 para darle una identidad progresista y orientada al futuro, pero hoy en día un nombre que contenga “20th Century” sugiere lo contrario.



Cuando George Orwell puso el título de “1984” a una novela que escribió en 1949, lo pensó como una advertencia sobre un futuro totalitario a medida que la Guerra Fría se afianzaba en una Europa dividida, pero hoy en día 1984 es el pasado y el título no tiene el mismo impacto.



DE ‘KENTUCKY FRIED CHICKEN’ A ‘KFC’

“Kentucky Fried Chicken” fue fundado en 1930 por Harland Sanders como un pequeño restaurante en el almacén de una gasolinera en Corbin, Kentucky. Fue una de las primeras cadenas de comida rápida en entrar en el mercado internacional, y en 1987 fue la primera cadena de restaurantes occidental en abrirse en China. Cambió su nombre a “KFC” unos años más tarde, sin duda en parte porque en Beijing, Moscú, Londres y otros lugares que no están cerca de Kentucky mucha gente probablemente nunca ha oído hablar del lugar.

(Fotografía de Kyle Taylor. Lente CC-BY-2.0.)



2.4.2.5 LA BRECHA SEMÁNTICA

La brecha semántica es la diferencia de perspectiva en la denominación y descripción cuando los recursos son descritos por procesos automatizados en lugar de por personas.

La brecha semántica es mayor cuando los programas informáticos o los sensores obtienen y nombran alguna información en un formato optimizado para la captura, el almacenamiento, la decodificación u otros criterios técnicos eficientes. Los nombres -como IMG20268.jpg en una foto digital- pueden tener sentido para la cámara ya que almacena fotos tomadas consecutivamente pero no son buenos nombres cuando los usa una persona. Seguro que preferiríamos nombres que describan el contenido de la foto, como GoldenGateBridge.jpg.

Cuando tratamos de examinar el contenido de los recursos creados por ordenadores o capturados por sensores, como un video o un programa de software compilado, una renderización del contenido simplemente como texto parece una insensatez.

LA BRECHA SEMÁNTICA ¿CÓMO SE LLAMA ESTA CANCIÓN?

```
.dzY9KpS^M4i0_x{\nSPñ'-00oty0i2g0yyVcGBch %0axc#"ÉiNcdS0ae' |N
0yNY# ,kw.Y{0yvyAEf>0 R0p±0"§{00S±PAt±á1}00-UP"i'É3 0c:
0jUt: ^"0hAc'f1q0p0cE "GgiAxet»0"n0aß'E0xy"OKB0S-°0ZF..0q0( 0E2IN0
|A>9Byú pF70Y0±0Z0UAAé0000er0!0S"0rx4a00Dah+0<0i@ Z00
JN 6KACK-i"0f#p0 4eLueY3Sc°3=0éYHg0y0«D& 00xy'°0SA0-0&...Y|5i0{?
0j0-feByó,dí01±P{<ev0AN|0AE0f0?ysfe' i-y'y0aNNéA00J' L">íeZAgT0
C050ES,0<±' a,##R00%±A«E'A!0yó,d'0ncZ0<"Z'U"06yD0e0}g3Z0kY0F0-
Azyy0iú' 00iY' ADE±±A00,/T0c0e4YJ9Y0E0000{°00}¿ ThA»-yó,d00á-
fXí00»RmZ0ÚV0S0"000[005A0",~0á0°*u0S] "rí00 HT!f00"0"t08µ00°,|
qjix<<'E°Np0ivv~|izp0-98'0(d0;P0c0qv°0ue-G00nHAM' Iaa-íE{>|0h~0é-,
0'0J3TA0E. |éCakEYus"0Yug0iwj-µ0eyyú»=00S{wC. 00JQ0ñ' 5E %0}
0°0 0[!b'00i0ayf5|j0~)Y:50±AzZN1y0y;J000U0Ej00°Uk0B·YFYTIS-IZÉ
0="yy00pz'0y)00°FE00rQ0i0SBA°0µ<Iyc0S000>0?~a0íy°0(m'00u0i0j¿
a'50,KZQ0E0I0E0U|í5'f~E10|04-->}'n°H3x»0<0A0f+H0f0 00' Swo0RE0E0
0v0cXG0'r<9L-è 000=p{ñ0K0Y0ÁV[0rHo<0S|ye0mE'0U'0gI0G/*"0'c-
ú-0°Yab0yó,d|0=0V0a#00a0"0y|0m>-J0=0RZ00%Yé'000AS'íS00i¿...0.#g|
$>XEV-0c=-m0[ñE[5;0c1-9f0{0>yó,d000R0=cb'0:| E0A0000eµ0'm0j0y
|x00 9n0«0U'J,0MWE 00EH°'è'y.ct00S'0r0w0D00qz;8y0,d0U
P|P'E0A0A0iZ'Ytzi 00iH0Y00L0000µ0Xp05xi-Bo0'xy9-±rgj0e}00+n'0fc
00b|Fr1<?000n00<00000V|0<'3mz0<u0%' méU0001*000éA0}0H0A>.00A0A
000%'00±bK0chY'N0±0N2P0.ú«E0ryy0-yú¿S¿ñ,0_Y°D'-C0S0q|X7i0|Tésú
ú'f3-Y0y0tA0/ 0E-t"0q0t00b04x<"1;\A000[6600Sf'000P'úH0z
|00°00U|X0E0j|í0400~LLk0(0E6U-I-°)%;0¿C#x-m0CN0/|~[a0NR'0;WNS°L
0mk M00500AA }X00'00Z0tAY°V*%+07D)(E000°00W00 v0E'fyc0pxw+-?x0A0
A50 E0J|ú000N3fBn0A?°00}3y'yZm0pG00Y0fX0e|-0070[3&B'0c0?>0wvú0
B-/?0°Pyó,d=0'>N0a0°00J0000IbU&t0E00'00°0g0c0D0owIµ'00#úZ%0mK0V
.+S0cy00?MByyyi°,íC|0íY|000(ké-0yó,d00-1Z0á0ZXA'0[0,µ0ug00/ á
0»09±c0x9-'0A00E00AU0&0S/'00°CTP00i0uay_y?0}0°00u°yó,dx0}¿T0dAZ
7A0m'Ry-5S-éKf?0A0i°".-°x0-c0au0VY00c·%T0T0t0µ?0°00 N0.,00g31A0 á
í0°A0E007°%0TR:00C)00P·[0T0S0000µÁy_¿0í0A0450D7800'0Y"0k0b?>0°0
```

El fichero fue codificado para ser interpretado por un programa de ordenador, no por una persona.



2.4.3 LA ELECCIÓN DE BUENOS NOMBRES E IDENTIFICADORES

Si alguien te dice que va a cenar con su mejor amigo, un primo, con alguien con quien juega al baloncesto, y su mentor profesional en el trabajo ¿para cuántos comensales habrá que reservar mesa? Entre dos y cinco; es posible que todas esas descripciones relacionales se refieran a una sola persona, o a cuatro personas diferentes, y como “amigo”, “primo”, “compañero de baloncesto” y “mentor” no nombran a personas específicas, tendrás que adivinar quién viene a cenar.

Si en lugar de descripciones se te dice que los invitados a la cena son Bob, Carol, Ted y Alice, puedes contar cuatro nombres y sabes cuántas personas van a cenar. Pero aun así no puedes estar seguro de qué cuatro personas concretas están involucradas, porque hay muchas personas con esos nombres.

La incertidumbre se elimina si usamos identificadores en lugar de nombres. Los identificadores son nombres que se refieren inequívocamente a una persona, un lugar o un recurso específico porque se asignan de forma controlada. Los identificadores suelen ser cadenas de números o letras en lugar de palabras para evitar los sesgos y asociaciones que las palabras pueden transmitir. Por ejemplo, un profesor puede calificar los exámenes que se identifican por los códigos de los estudiantes en lugar de los nombres.

Nombres {y, o, vs} Identificadores

Las personas cambian de nombre por muchas razones: cuando se casan o se divorcian, o porque su nombre a menudo se pronuncia o escribe mal, para hacer una declaración política o étnica, o porque quieren destacar. Unos años un jugador de fútbol con un gran ego llamado Chad Johnson, que es el segundo apellido más común en los EE.UU., decidió cambiar su nombre por el número de 85 que llevaba en la camiseta del equipo, convirtiéndose en Chad “Ochocinco”. Usaba la página web ochocinco.com y usaba el nombre ochocinco en Facebook y Twitter. Irónicamente, cuando Ochocinco quiso poner Ocho Cinco en la parte de atrás de su camiseta de fútbol, la liga de fútbol no lo permitió conforme a su reglamento sobre equipaciones. Eso seguramente contribuyó a su decisión de cambiar su nombre de nuevo a Chad Johnson en 2012.



2.4.3.1 PRODUCCIÓN DE NOMBRES INFORMATIVOS

El principio más básico de la denominación es elegir nombres que sean informativos, lo que hace que sean más fáciles de entender y recordar. Es más fácil saber lo que hace un programa de ordenador o un documento XML si utiliza nombres como “ItemCost” y “TotalCost” en lugar de sólo “I” o “T”. La gente introducirá la información sobre su dirección postal de forma más consistente y reutilizable si un formulario pide explícitamente “Calle”, “Ciudad” y “Código Postal” en lugar de “Línea1” y “Línea2”.

Los identificadores pueden diseñarse con una estructura interna y una semántica que transmita información más allá del aspecto básico de apuntar a recurso específico. Un número de ISBN (Número Estándar Internacional para Libros en venta) como “978-0-262-07261-8” identifica un recurso concreto (07261=“Ingeniería de Documentos”) y también revela que el recurso es un libro (978), en inglés (0), y publicado por The MIT Press (262).

Los puntos de navegación que marcan las intersecciones de las señales radioeléctricas de las balizas terrestres o los satélites, y que son cruciales para los pilotos de aeronaves, solían ser códigos de cinco letras sin sentido que se cambiaron para que sugirieran sus ubicaciones; los nombres de los puntos de referencia semánticos hacían que fuera menos probable que los pilotos introdujeran nombres equivocados en los sistemas de navegación. Por ejemplo, algunos de los puntos de navegación cerca de Orlando, Florida -el hogar de Disney World- son MICKI, MINNIE, y GOOFY.

2.4.3.2 USO DE VOCABULARIOS CONTROLADOS

Una forma de fomentar buenos nombres de recursos en un determinado dominio o tarea es establecer un vocabulario controlado. Un vocabulario controlado es como un diccionario fijo o cerrado que incluye los términos que se pueden utilizar en un dominio determinado. Un vocabulario controlado reduce el número de palabras utilizadas, disminuyendo la sinonimia y la homonimia, eliminando las asociaciones indeseables, dejando un conjunto de palabras con significados definidos con precisión y reglas que rigen su uso.

Un vocabulario controlado no es simplemente un conjunto de palabras permitidas; también incluye sus definiciones y a menudo especifica las reglas por las que los términos del vocabulario pueden utilizarse y combinarse. Los diferentes dominios pueden crear vocabularios con-



trolados específicos para sus propios fines, pero lo importante es que el vocabulario se utilice de manera coherente en todo ese dominio.

En cuanto a los recursos bibliográficos, entre los aspectos importantes del control de vocabulario figuran la determinación de las formas autorizadas de los nombres de los autores, los títulos uniformes de las obras y el conjunto de términos por los que se conocerá un determinado tema. En biblioteconomía, el proceso de creación y mantenimiento de estos nombres y términos uniformes se conoce como **control de autoridades**.

En muchos dominios existen registros oficiales de autoridades: un nomenclátor asocia nombres y ubicaciones y nos dice si debemos referirnos a Bombay o a Mumbai; el Sistema de Nombres de Dominio (DNS) asigna nombres de dominio y de host legibles por personas, con sus correspondientes direcciones IP; el Chemical Abstracts Service Registry asigna identificadores únicos a cada sustancia química descrita en la literatura científica publicada en acceso abierto; numerosas instituciones asignan identificadores únicos a diferentes categorías de especies animales.

En algunos casos, los archivos de autoridades son creados o mantenidos por una comunidad, como en el caso de MusicBrainz, una “enciclopedia musical abierta” a la que los usuarios contribuyen con información sobre los artistas, los lanzamientos, las pistas y otros aspectos sobre la música. Los metadatos de música son notoriamente poco fiables; un estudio encontró más de 100 variaciones en la descripción de la canción Knockin’ on Heaven’s Door (escrita por Bob Dylan) tal como fue grabada por Guns N’ Roses.

2.4.3.3 PERMITIR LOS ALIAS

Un vocabulario controlado es sumamente útil para las personas que lo utilizan, pero si se diseña un Sistema de Organización para otras personas que no lo usan con asiduidad o que no pueden utilizarlo por su cuenta, es necesario tener en cuenta la variedad de palabras que realmente emplearán cuando busquen o describan recursos. El nombre autorizado de una cierta especie de pez es *Amphiprion ocellaris*, pero la mayoría de la gente lo buscaría como “pez payaso”, “pez anémona”, o incluso por su nombre más familiar de Nemo, por la familiaridad con la película de Disney.

Furnas sugiere permitir el uso de alias para conectar los vocabularios no controlados o naturales que la gente utiliza, con el vocabulario controlado empleado por el sistema organizador. Con esto quiere decir



que debe haber muchas rutas de acceso alternativas a cada palabra o función que un usuario está tratando de encontrar. Por ejemplo, el nombre de nacimiento del 42º Presidente de los Estados Unidos de América es “William Jefferson Clinton”, pero las páginas web que se refieren a él como “Bill Clinton” son mucho más comunes, y las búsquedas de lo primero se redirigen a lo segundo. Un mecanismo relacionado utilizado por los motores de búsqueda es la corrección ortográfica, que consiste esencialmente en tratar todas las ortografías incorrectas como alias del considerado correcto (“¿Quiso decir California?” cuando tecleó “Claifornia”).

UN ALIAS CON ÉXITO CASI MATA A ESTE PEZ



Un pez que antes se conocía como Pez Diente de la Patagonia por sus grandes y poco atractivos dientes se hizo popular en los restaurantes estadounidenses cuando un mayorista de pescado comenzó a comercializarlo como la Lubina Chilena aunque normalmente se encuentra más al sur en las frías aguas antárticas y no es una lubina. Por desgracia, este alias tuvo tanto éxito que llevó a la sobrepesca, amenazando la supervivencia de la especie. Algunos chefs, restauradores y distribuidores de mariscos con un mayor compromiso ambiental organizaron un boicot para salvar la especie. (Fabricant 2002).

(Foto publicada por el Gobierno de los Estados Unidos. No está protegido por el derecho de autor (17 USC Sec. 105).)

2.4.3.4 CONSTRUCCIÓN DE IDENTIFICADORES ÚNICOS O CUALIFICADOS

Aunque un identificador se refiera a un solo recurso, esto no significa que no haya dos identificadores idénticos. Un sistema de inventario militar podría utilizar el número de inventario 99 000 1111 para identificar un paquete de raciones diarias para climas fríos, mientras que otro sistema de inventario podría utilizar el mismo número para identificar una válvula de radio electrónica. Cada identificador es único en su sistema



de inventario, pero si se envía una solicitud de suministro al almacén equivocado, se podría enviar a los soldados hambrientos válvulas de radio en lugar de raciones.

Podemos evitar o reducir las colisiones entre identificadores añadiendo información sobre su espacio de nombres (namespace), el dominio del que se seleccionan los nombres o identificadores, creando así lo que se suele denominar como nombres cualificados. Hay varias docenas de ciudades de EE.UU. llamadas “Springfield” y “Washington”, pero añadir el código del estado a las direcciones postales los distingue. De la misma manera, podemos añadir prefijos a los nombres de los elementos XML cuando creamos documentos que reutilizan componentes de múltiples tipos de documentos, distinguiendo `<libro:Título>` de `<legal:Título>`.

Dado que no han sido creadas por un algoritmo cuyos resultados sean probadamente únicos, no consideramos que las huellas dactilares, u otra información biométrica, sean identificadores globalmente únicos para las personas, pero a todos los efectos prácticos lo son.

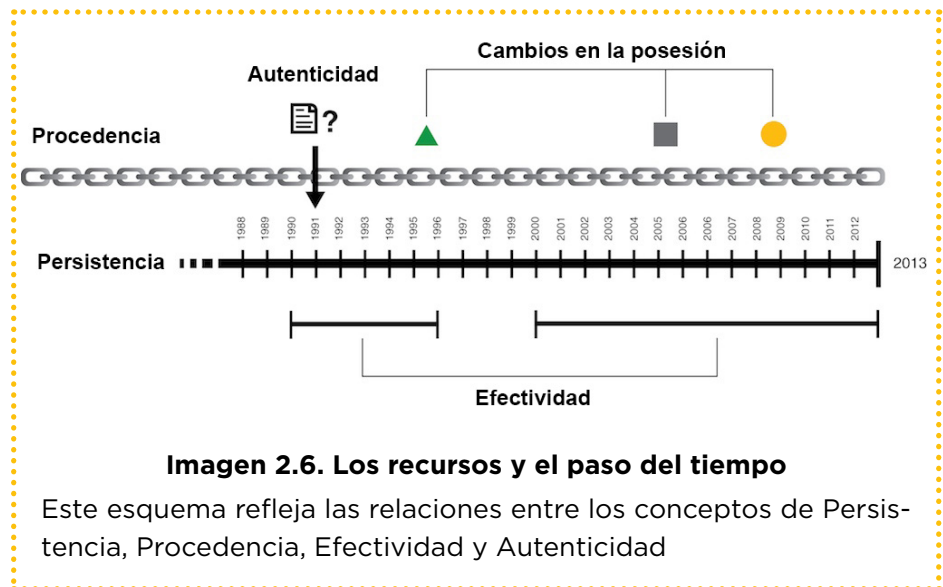
2.4.3.5 LA DIFERENCIA ENTRE IDENTIFICAR Y RESOLVER

Cuando el identificador no contiene información sobre la ubicación de los recursos, debe “resolverse” para determinar su ubicación. En el caso de los recursos físicos, la resolución tiene lugar con la ayuda de señales, mapas u otros recursos asociados que describen la disposición de los recursos en algún entorno físico; por ejemplo, los mapas “Usted está aquí” asocian cada identificador de recurso con una coordenada u otro medio de encontrarlo en el mapa. En el caso de los recursos digitales, el resolvidor es un sistema o servicio de directorio que interpreta un identificador y busca su ubicación o inicia directamente la recuperación del recurso (lo abre en el navegador, lo descarga, lo consulta).



2.5 LOS RECURSOS Y EL PASO DEL TIEMPO

Los problemas de “¿qué es el recurso?” y “¿cómo lo identificamos?” son complejos y a menudo requieren un trabajo continuo para asegurar que se les dé una respuesta adecuada a medida que evoluciona un Sistema de Organización. Tal vez necesitemos saber cómo cambia o no cambia un recurso a lo largo del tiempo (su persistencia), si su estado y contenido entran en juego en un momento determinado (su eficacia), si el recurso es lo que se dice que es (su autenticidad) y, a veces, quién ha certificado su autenticidad a lo largo del tiempo (su procedencia). Un recurso puede tener persistencia, pero sólo la procedencia proporcionada por una cadena de custodia documentada permite responder con autoridad a las preguntas sobre la autenticidad. La eficacia describe los límites de la vida útil de un recurso en la línea temporal.



2.5.1. PERSISTENCIA

Aunque se haya llegado a un acuerdo sobre el significado de “una cosa” en su propio sistema de organización, se sigue planteando la cuestión de cómo puede verse afectada la identidad del recurso a lo largo del tiempo, es decir, su persistencia.

2.5.1.1. IDENTIFICADORES PERSISTENTES

¿Cuánto tiempo debe durar un identificador? Karen Coyle da una respuesta convencional, aunque insatisfactoria: “Tanto como sea necesario”. En algunos casos, el tiempo es relativamente corto. Cuando pides un café especial y el barista te pide el nombre, este identificador sólo



tiene que durar hasta que recojas tu pedido al final del mostrador. Pero otros plazos son mucho más largos. Para las bibliotecas y repositorios de datos científicos, económicos, de censo u otros, el plazo puede ser “para siempre”.

En el diseño de un plan de identificadores persistentes se debe tener en cuenta tanto el plazo necesario como el número de recursos que se deben identificar. Cuando el Protocolo Internet (IP) fue diseñado en 1980, contenía un esquema de direcciones de 32 bits, suficiente para más de 4.000 millones de direcciones únicas. Pero el enorme crecimiento de la red y la aplicación de las direcciones IP a recursos de tipos inesperados han requerido un nuevo esquema de direcciones de 128 bits.



LA GRAN ESFINGE DE GIZA

La gran esfinge lleva en el mundo desde hace miles de años. Ha sobrevivido a actos de vandalismo, a usarse como la diana de la artillería de Napoleón y ser tragada por tormentas de arena hasta que volvió a la superficie tras las excavaciones del siglo xx.

Sus orígenes, igual que los de su nombre “Esfinge” siguen siendo tema de debate entre los expertos, y a pesar de esto, es una figura mundialmente conocida que vivirá eternamente en nuestra memoria.

2.5.1.2. RECURSOS PERMANENTES O PERSISTENTES

Aunque la persistencia suele tener una dimensión tecnológica, es más importante considerarla como el compromiso de una institución u organización de realizar actividades a lo largo del tiempo para garantizar que se disponga de un recurso cuando se necesite. Dicho de otro modo, se realizan de forma planificada las actividades de preservación y de gobernanza para garantizar, como resultado de ellas, la persistencia³.

La sutil relación entre la preservación y la persistencia plantea algunas preguntas interesantes sobre lo que significa que un recurso permanezca igual a lo largo del tiempo. Una forma de pensar en la persisten-

³ Sobre las actividades que componen un Sistema de Organización, puede consultarse el capítulo tercero del libro “TDO core concepts edition”, en concreto los apartados de preservación (§3.5.2) y gobernanza (§3.5.4).



cia es considerar que un recurso persistente nunca cambia. Sin embargo, los recursos físicos a menudo requieren mantenimiento, reparación o restauración para mantenerlos accesibles y utilizables, y podríamos preguntarnos si en algún momento estas actividades los han transformado en recursos diferentes. Asimismo, los recursos digitales requieren copias de seguridad y migraciones periódicas para mantenerlos disponibles y esto podría incluir el cambio de su formato digital.

PÁRATE A PENSAR: LA PARADOJA DE TESEO

Cada día que el barco de Teseo está en el puerto, una sola tabla es reemplazada, hasta que después de unos años el barco es reconstruido por completo: no queda ni una sola tabla original. ¿Sigue siendo el barco de Teseo? Y supongamos que, mientras tanto, los constructores del barco han estado construyendo un nuevo barco con los tablonces reemplazados. ¿Es ese el barco de Teseo?

Muchos recursos, como los periódicos en línea o los feeds de blogs, cambian continuamente su contenido, pero siguen teniendo identificadores persistentes. Esto sugiere que deberíamos pensar en la persistencia de forma más abstracta, y considerarla más bien como recursos persistentes que permanecen funcionalmente iguales para dar soporte a las mismas interacciones en cualquier momento de su ciclo de vida, incluso si sus propiedades físicas o valores informativos cambian.

2.5.2. EFECTIVIDAD

Muchos recursos, o sus propiedades, tienen también una eficacia espacial o temporal, lo que significa que producen efecto en un momento y/o lugar determinados; es casi seguro que dejarán de ser eficaces en alguna fecha futura, y pueden dejar de serlo en diferentes lugares.

La eficacia temporal, a veces conocida como “tiempo de vida” (time-to-live), se expresa normalmente como un rango entre dos fechas. Consiste en una fecha desde la que el recurso es efectivo y, opcionalmente, una fecha en la que el recurso deja de ser efectivo o se vuelve obsoleto. Para algunos tipos de recursos, la fecha de entrada en vigor es el momento en que se crean, pero para otros, la fecha de entrada en vigor puede ser un momento diferente del momento de la creación. Por ejemplo, una ley aprobada en noviembre puede entrar en vigor el primero de enero del año siguiente, y las tarjetas de crédito deben activarse primero, pero ya no pueden utilizarse después de su fecha de caducidad. La “fecha de entrada en vigor” es el reverso del “consumir preferentemente antes de” de los productos perecederos. Esa fecha indica cuando un producto se estropea, mientras que la fecha efectiva de



un artículo es cuando “empieza a estar bueno” y, por tanto, el recurso al que reemplaza debe ser eliminado o archivado.

La efectividad espacial tiene en consideración condicionantes como fronteras, seguridad, carreteras, altitud, profundidad y otra serie de factores geográficos. Ciertos tipos de recursos, incluyendo a las personas, tienen restricciones como a dónde pueden o no poder trasladarse o en dónde usarse, como pueden ser las mercancías peligrosas, explosivos, drogas, medicamentos o alcohol. Los aspectos jurisdiccionales afectan a las fronteras, corredores de transporte de mercancías, estaciones meteorológicas y estudios geográficos. Los globos de medición atmosférica son sensibles a la altitud y las bombonas de buceo lo son a la profundidad.

Concuerne a la efectividad de un recurso, a veces se produce una intersección con el control de autoridades de nombres y lugares. Los cambios de nombre de un recurso suelen estar asociados a ciertas fechas, eventos y ubicaciones. Las leyes y reglamentos difieren según exista una u otra pertenencia organizacional o geopolítica. Además esas fronteras cambian con frecuencia, como en los territorios que han sufrido conflictos civiles, ocupación extranjera u otro tipo de conflictos políticos, y que han tenido diferentes nombres a lo largo del tiempo, o incluso en un mismo periodo.

¿EN QUÉ PAÍS TE CREES QUE VIVES?

Incluso si siempre has vivido en el mismo sitio, la respuesta a “¿En que país vives?” puede depender de en qué momento se te haga la pregunta. Consideremos el caso de una mujer mayor nacida en 1929 en Zemun, un distrito de la ciudad del este de Europa de Belgrado, y que nunca se ha mudado de ciudad. El lugar en el que vive ha sido parte de siete países diferentes durante su vida: el reino de Yugoslavia (1929-1941); estado independiente de Croacia (1941-1945); República Popular Federal de Yougoslavia (1945-1963); República Socialista Federal de Yugoslavia (1963-1992); República Federal de Yugoslavia (1992-2003); Estado Unido de Serbia y Montenegro (2003-2006); Republica de Serbia (2007—presente).

A día de hoy estas fronteras en disputa provocan problemas en Google Maps cuando tiene que presentar ciertas fronteras internacionales. Dado que Google está sujeto a las leyes del país en el que se encuentran sus servidores, entra en conflicto con el punto de vista del lugar desde el que se puede estar solicitando su servicio de mapas.



En la mayoría de los casos la efectividad implica requisitos de persistencia dado que es importante ser capaz de determinar y reconstruir la configuración de recursos tal y como era en un momento anterior en el tiempo. Un nuevo impuesto que entrase en vigor el 1 de enero, tendría que ser interpretado conforme a su anterior redacción si la inspección fiscal realizada por el gobierno es anterior a esa fecha.

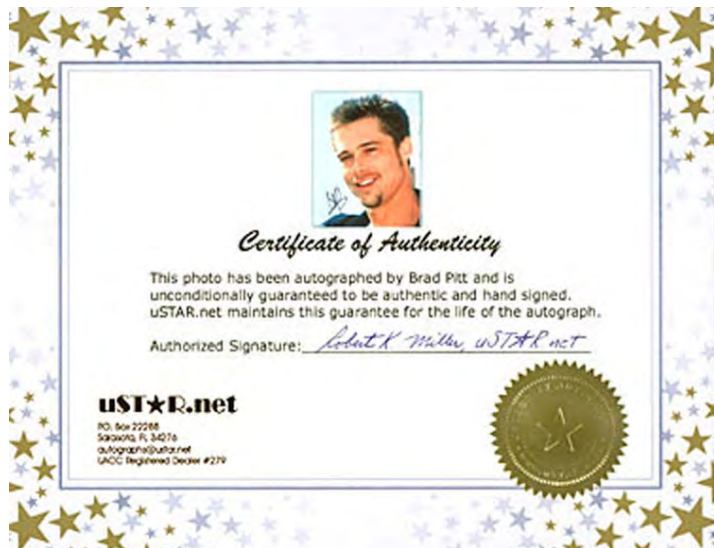
2.5.3. AUTENTICIDAD

En el lenguaje común decimos que algo es auténtico si algo se puede comprobar, o que se acepta, que es lo que dice ser. La importancia y los matices de las preguntas sobre la autenticidad se ven reflejadas en la cantidad de palabras que se emplean para describir la relación entre “la cosa real” (el original) y el resto de cosas: copia, reproducción, réplica, fake, falsificación, recreación, reconstrucción, impostor, suplantador, doble ...y un largo etcétera.

Es fácil pensar en ejemplos en los que la autenticidad de un recurso importa: un contrato legal firmado, una obra de arte, un objeto de valor histórico, o simplemente un autógrafo.

El responsable o usuario de un Sistema de Organización, sea un humano o un programa informático, puede autenticar los nuevos recursos que se van creando. Una tercera parte puede actuar también como garantía de autenticidad. Existen profesiones basadas en investigar y asegurar que un recurso es auténtico. Existe un amplio cuerpo de técnicas para establecer la identidad de una persona o de un recurso físico. A menudo se valora la integridad física de una información registrada cuando nos cuestionamos la integridad de su contenido.

La autenticación de recursos digitales es más difícil de establecer. Lo digital puede reproducirse sin coste, existir en múltiples ubicaciones, tener diferentes nombres en documentos idénticos o nombres idénticos en documentos diferentes, e implicar otras complicaciones que no se manifiestan en los ítems físicos. Las soluciones tecnológicas para asegurar la autenticidad digital incluyen sellos de tiempo, marcas de agua, encriptamiento y firmas digitales. No obstante, aunque los académicos generalmente confían en los métodos técnicos, los tecnólogos son más escépticos porque son capaces de imaginar formas de sortearlos o quebrantarlos. Incluso cuando se establece un sofisticado sistema tecnológico para establecer la autenticidad, solo podemos confiar en ella desde el momento en que el sistema se hace cargo de la “cadena de custodia” del documento.



¿TE LO CREEES?

Ustar.net vende fotos firmadas por famosos y cada una viene con su certificado de autenticidad que incluye una réplica de la foto y una firma de un empleado de la empresa, para garantizar que el autógrafo está hecho a mano. Pero Ustar no facilitar un certificado que garantice que la firma del empleada esté autorizada para certificar lo anterior.

(Screenshot by R. Glushko. Source: ustar.net.)

2.5.4 PROCEDENCIA

Cuando se aborda la selección de los recursos, una de las cosas que se recomienda es analizar cualquier evidencia disponible o registro que informe sobre el uso de los recursos desde sus orígenes, para asegurarse de que han mantenido su calidad en el transcurso del tiempo. El concepto de procedencia (provenance) transforma el enfoque pasivo de preguntarse “¿Qué le ha pasado a este recurso?” en acciones específicas que se llevan a cabo para asegurarse de que nada negativo le ha sucedido a un recurso o que permitan detectarlo si hubieran sucedido.

La idea de que los documentos importantes deben crearse de forma que puedan ser autenticados y, entonces, preservados, dentro de una cadena de custodia sin fisuras, nos retrotrae hasta la antigua Roma. Los notarios actuaban de testigos en la creación de esa clase de documentos, los cuales luego eran protegidos para mantener su integridad como valor y evidencia. En los Sistemas de Organización del tipo museo y archivo, que preservan objetos o documentos raros o de importante relevancia cultural, estas preocupaciones quedan expresadas en



el conocido como “principio de procedencia”. Se trata de la historia de la propiedad de una colección o de los recursos que la componen, manteniendo registros de dónde han estado custodiados y de quienes han tenido acceso a ellos.

En la antigua China se conoce una técnica única en los Sistemas de Organización, que es la impresión de sellos muy intrincados en documentos, libros y pinturas, que de forma colectiva registran la procedencia de los propietarios que incluyen reseña y aprobación sobre objeto por parte el emperador o de sus altos funcionarios.

No obstante, no son solamente los historiadores del arte o responsables de la custodia de documentos críticos los que necesitan prestar atención a la procedencia. Cualquiera que, por ejemplo, piensa en comprar un coche de segunda mano, será prudente si comprueba el historial del vehículo (usando el Número de Identificación de Vehículos, el identificador persistente del coche) para asegurarse de que no ha tenido un siniestro grave, sufrido una inundación o es robado.



MANUSCRITO CHINO CON SELLOS DE PROCEDENCIA

Este bello manuscrito, preservado en el National Palace Museum de Taipei, fue creado por Zhao Ji (赵佶), emperador Huizong, el octavo emperador de la dinastía Song, hace unos mil años. Era famoso por su destreza en poesía, pintura y caligrafía. El manuscrito contiene dos poemas; el de la derecha describe las técnicas para la pintura de paisajes, mientras que el de la izquierda expresa el aprecio del emperador por las flores de los ciruelos, que anuncian el comienzo de la primavera. Los sellos rojos son los usados durante generaciones por muchos emperadores de la dinastía Ching. Los más antiguos son posteriores en al menos quinientos años después de que Huizong crease el poema. Estampar el sello rojo sobre un recurso es análogo, aunque mucho más elegante e informativo que hacer ahora un “like” en una página web.

(Foto: R. Glushko)



CAPÍTULO 3

DESCRIPCIÓN DE RECURSOS

Y METADATOS

Robert J. Glushko / Kimra McPherson / Ryan Greenberg /
Robyn Perry / Matthew Mayernik / Graham Freeman / Carl Lagoze

CONTENIDO

105	3.1. Introducción
109	3.2. Visión general de la descripción de recursos
109	3.2.1. Nombrar {y, o, vs.} Describir
111	3.2.2. “Descripción” como término inclusivo
	3.2.2.1. La descripción bibliográfica
	3.2.2.2. Los metadatos
	3.2.2.3. Etiquetado de los recursos web
	3.2.2.4 RDF (Marco para la descripción de recursos)
	3.2.2.5. Los objetos de información agregados
118	3.2.3. Marcos para la descripción de recursos
120	3.3. El proceso de descripción de recursos
123	3.3.1. Determinar el alcance y el foco
	3.3.1.1. Describir contenidos individuales o describir colecciones
	3.3.1.2. Nivel de abstracción en los recursos descriptivos
	3.3.1.3 El alcance, la escala y la descripción de recursos
128	3.3.2. Establecer los propósitos
	3.3.2.1. La descripción de recursos para apoyar la selección
	3.3.2.2. La descripción de recursos para apoyar la organización
	3.3.2.3. La descripción de recursos para apoyar las interacciones
	3.3.2.4. La descripción de recursos para el mantenimiento y administración
	3.3.2.5. La descripción de recursos para la ciencia y el “sensemaking”
137	3.3.3 Identificar propiedades



		3.3.3.1. Propiedades intrínsecas estáticas
		3.3.3.2. Las propiedades extrínsecas estáticas
		3.3.3.3 Las propiedades intrínsecas dinámicas
		3.3.3.4 Las propiedades extrínsecas dinámicas
142	3.3.4.	El diseño del vocabulario de descripción
	3.3.4.1	Los principios de una buena descripción
	3.3.4.2	¿Quién utiliza las descripciones?
	3.3.4.3	Los vocabularios controlados y las normas de contenido
	3.3.4.4	El control del vocabulario como reducción de la dimensionalidad
148	3.3.5	Diseño de la forma de la descripción
149	3.3.6	La creación de descripciones de recursos
	3.3.6.1	Las descripciones hechas por profesionales
	3.3.6.2	Las descripciones hechas por autores o creadores
	3.3.6.3	Las descripciones hechas por usuarios
	3.3.6.4	La descripción de recursos automatizados y computacionales
157	3.3.7	Evaluar las descripciones de recursos
	3.3.7.1	Evaluar la creación de descripciones de recursos
	3.3.7.2	Evaluar el uso de las descripciones de recursos
	157	3.4 La descripción de recursos no textuales
	157	3.4.1 La descripción de recursos artísticos y museísticos
	159	3.4.2 La descripción de imágenes
	160	3.4.3 La descripción de música
	163	3.4.4 La descripción de videos



3.1. INTRODUCCIÓN

Este capítulo incorpora al ámbito de la organización de información un proceso central que tiene que ver con **cómo describir el contenido y las características de los recursos de información en un sistema**. No podemos olvidar que la razón más importante para organizar las cosas es poder encontrarlas en un futuro. Esto puede ser sorprendentemente difícil de hacer, porque las personas reconocemos las cosas por nombres distintos o recordamos cualidades diferentes de cada una de ellas. Lo que parecen operaciones sencillas a primera vista, se vuelve complejo en la práctica real cuando se acumulan volúmenes muy amplios de contenidos.

Las ideas de este capítulo se apoyan en aspectos que requieren tratarse con mayor detalle, como son: identificar y seleccionar los recursos que se van a organizar, y sobre ellos, organizarlos y mantenerlos, y administrar el propio Sistema de Organización en el que se integran. Un concepto esencial que recorre este capítulo es la diferencia que se establece al plantearse cuál es el enfoque o foco que adoptamos sobre un recurso (resource focus). De esta forma se establece una primera distinción entre recurso primario (el contenido, el material original) y el recurso descriptivo (o recurso de descripción, metadatos o recurso secundario) que cumple una función de sustituto del recurso original dentro del sistema de información, a través de una representación de sus propiedades. Ambos recursos están vinculados entre sí, y será habitual que un recurso primario esté relacionado con múltiples y diferentes recursos descriptivos. A menudo, tratar un recurso como primario o secundario puede depender del contexto en el que se incluye, y por lo tanto depender del punto de vista o foco con el que lo observemos¹.

El famoso cuadro del pintor americano del siglo XIX James Whistler se exhibe en el Museo de Orsay en París y es descrito como la *Mona Lisa* de la época victoriana. ¿Por qué nombre se lo conoce? ¿Cómo debería ser descrito?

Las descripciones de recursos para obras de arte suelen contener el nombre del artista, el medio en el que se creó, el año de su realización y, por supuesto, su título. La mayoría son bastante obvias en cuanto a las propiedades que describen; el título, debido a su importancia y poder expresivo, es a menudo una excepción.

¹ Recomendamos al lector de este libro repasar estos conceptos en los capítulos 2, 3 y 4 de la edición completa original del libro “The discipline of organizing”. Cuando se ha considerado necesario, se ha incorporado de ellos alguna aclaración necesaria para una comprensión más clara de este capítulo (N.E.)



La mayoría de las veces, el título de un cuadro describe el tema. Si reconocemos la pintura anterior, lo más probable es que la conozcamos por su nombre coloquial, *La madre de Whistler*. Aunque es un retrato de Anna McNeill Whistler, madre del pintor James Abbott McNeill Whistler, el artista le dio un título completamente distinto, *Arrangement in Grey and Black, No. 1 (Arreglo en gris y negro, No. 1)*, porque creía que la propiedad más importante de un cuadro no era el tema que representaba, sino sus propiedades puramente estéticas y su efecto en el espectador. Así que Whistler dio nombres (títulos, podríamos también decir) a sus pinturas, que eran en su mayoría paisajes y retratos, a modo de composiciones musicales: *Nocturno en negro y oro*; *Sinfonía en blanco*; *Arreglo en rosa, rojo y púrpura*; y así sucesivamente.



“Arreglo en gris y negro N° 1” (1871). Cuadro al óleo de James Abbot McNeil Whistler. Su título alternativo es: “Retrato de la madre del artista” y “Madre de Whistler”. El cuadro pertenece a la colección del Museo d’Orsay, en París.

Fotografía: Wikimedia Commons, Google Art Project

Si el título elegido por Whistler sorprende, porque describiríamos el cuadro como el retrato de una anciana, esto ayuda a reforzar lo extremadamente diferentes que pueden ser los nombres para un mismo recurso. Las descripciones de los recursos y los metadatos proporcio-



Evolución del Mt. St. Helena

Antes de 1980, el Monte Santa Helena era un pico nevado digno de postales. Ahora, el cráter de más de una milla de ancho de su cima nos recuerda lo violentas que son las erupciones volcánicas (Créditos: Imagen en dominio público procedente del US Forest Service)

nan significados, pero ¿para quién? Lo más destacado de un recurso puede depender del contexto en el que se experimenta y por lo tanto puede cambiar con el tiempo. Las descripciones que tienen sentido para algunas personas pueden no tenerlo para otras. Ciertas descripciones o metadatos pueden parecer “erróneos” (en realidad, no adecuados), a las personas que buscan en sistemas y que esperan encontrar la información de cierta manera, cuando en realidad los responsables de catalogarla lo han enfocado de otra manera. “Erróneo”, en este caso, significa desajuste con las expectativas y marco mental del usuario.

El Monte Santa Helena, en el estado de Washington, era considerado una montaña hasta 1980. Entonces, el accidente volcánico más mortal y económicamente destructivo de la historia de los Estados Unidos voló la cima de la montaña, matando a 57 personas y dejando un cráter de una milla de ancho. Hoy en día casi todas las descripciones del Monte Santa Helena mencionan la erupción volcánica. Pero ¿cómo manejar en las búsquedas a los documentos anteriores a la explosión que describen el Santa Helena simplemente como una montaña?

Puede parecer que se podría evitar esta “búsqueda que se basa en una descripción incorrecta” si en todo momento los sistemas mantuvieran las descripciones de un recurso actualizadas, incorporando cualquier nueva información que les afecte, y que a la luz de la tremenda potencia de los motores de búsqueda web hoy en día ya se supone que acceden, manejan y producen información superabundante, detallada y actualizada al momento. **Pero los avances tecnológicos en búsqueda y recuperación de información no eliminan el esfuerzo cognitivo de recordar qué son las cosas, cómo se describen mejor y dónde pueden**



encontrarse. El diseño de descripciones para los recursos y los metadatos dependen de la razón que motive nuestra necesidad de encontrar información posteriormente, de casos de uso futuros. Este capítulo trata sobre el cómo y el porqué.

PÁRATE A PENSAR: LOS LUGARES TIENEN SUS MOMENTOS

Nuestra percepción del Monte Santa Helena cambió tras su erupción. Seguro que somos capaces de recordar lugares que ya no son como antes porque formaron parte de un evento importante. Como una boda. O los juegos olímpicos Un desastre natural. O las Torres Gemelas.

Es fácil encontrar imágenes del antes y el después del Monte Santa Helena con una búsqueda en la web. ¿Qué información podría estar asociada a estas imágenes? Las cámaras modernas asignan un identificador a la fotografía almacenada y también capturan la descripción técnica de la producción de la imagen: el tipo de cámara, la lente, la velocidad del obturador, la sensibilidad a la luz, la apertura y otros ajustes. Muchas cámaras modernas también registran información sobre las circunstancias geográficas y temporales de la imagen: la fecha, la hora y el lugar del mundo en el que se tomó la fotografía.

Quando la imagen se transfiere de la cámara y se publica para que todos la vean, puede ser útil registrar información biográfica sobre el fotógrafo para ayudar a los espectadores a empatizar con él y comprender mejor el contexto de la fotografía. También puede haber diferentes licencias e información de derechos de autor asociados a la fotografía: quién es su propietario y cómo puede utilizarse. **Además de la mera fotografía, existe un amplio espectro de metadatos que construyen su significado completo.**





Consideremos un contexto completamente diferente. Cuatro niños de 7 años están eligiendo piezas de Lego para completar su última construcción. El primer niño está buscando “cilindros”, otro “tapones de botella”, el tercero “limpiaparabrisas”, y el último está buscando “¿ladrillos redondos?”. Resulta que todos buscan lo mismo; cada niño ha ideado su propio conjunto de términos descriptivos para los pequeños bloques de construcción. Algunas de sus descripciones se basan sólo en el color (“rojo”), otras en el color y la forma (“túnel azul”), otras en el papel (“conector”), otras en los conocimientos culturales comunes (“espada láser”). Otras, como “caracol de la cárcel” y “babosa”, parecen inidentificables, a menos que estés dentro de la mente de un niño de 7 años concreto. No importa si los niños usan diferentes vocabularios de descripción cuando juegan solos, pero tendrán que estar de acuerdo si juegan juntos.

Las pinturas, las fotografías digitales y las piezas de Lego son muy diferentes, pero en conjunto estos escenarios plantean importantes preguntas sobre la descripción de los recursos que son las que intentamos responder en este capítulo:

- **¿Cuál es el propósito de los recursos descriptivos?**
- **¿Qué propiedades de los recursos deberían ser descritas?**
- **¿Cómo se crean los recursos descriptivos?**
- **¿Qué hace que un recurso descriptivo sea bueno?**

3.2. VISIÓN GENERAL DE LA DESCRIPCIÓN DE RECURSOS

Los recursos se describen para poder referirnos a ellos, distinguirlos, buscarlos, gestionar el acceso a ellos, preservarlos y hacer predicciones sobre lo que les podría ocurrir o sobre para lo que podrían servir. Cada uno de estos propósitos requiere diferentes recursos descriptivos. Utilizamos descripciones de recursos en cada acto de comunicación y en cada conversación; son los **habilitadores de los sistemas de organización**.

3.2.1. NOMBRAR {Y, O, VS.} DESCRIBIR

En todo Sistema de Organización se tomarán decisiones sobre qué cosas deben tratarse como recursos y sobre la manera en que los nombres, denominaciones o títulos que les asignemos y los identificadores usados distinguirán un recurso de otro. Los nombres pueden sugerir



las propiedades y los principios que un sistema de organización utiliza para organizar sus recursos. Podemos ver cómo las sociedades, históricamente, han “organizado” o “descrito” a su población observando que entre los apellidos más comunes en inglés se encuentran descripciones de ocupaciones (Smith (Herrero), Miller (Molinero), Taylor (Sastre)), descripciones de relaciones de parentesco (Johnson (Hijo de John), Wilson (Hijo de Will), Anderson (Hijo de Ander)) y descripciones de apariencia (Brown (Marrón), White (Blanco)).

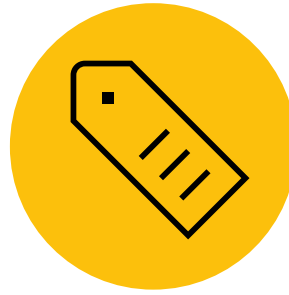
En muchas culturas, uno de los cónyuges toma un nombre que refleja su relación marital. En muchas partes del mundo de habla inglesa, las mujeres casadas se refieren a menudo a sí mismas usando el apellido de su marido.

De manera parecida, muchos otros tipos de recursos tienen nombres que son descripciones de propiedades, incluyendo edificios (El Pentágono, la Casa Blanca), ubicaciones geográficas (América del Norte, Mar Rojo) y ciudades (Grand Forks, Baton Rouge).

A cada recurso se le puede dar un nombre o un identificador. Los identificadores son descripciones de recursos especialmente eficientes porque, por definición, los identificadores son únicos en algún dominio o dentro de una colección de recursos. Pero los nombres e identificadores no suelen servir para describir el recurso, porque por lo general se asignan al recurso en lugar de registrar una propiedad del mismo: *rotulan, pero no describen*.

Sin embargo, la arbitrariedad de los nombres y los identificadores hace que no sirvan para distinguir entre recursos si una persona no los conoce. Por eso utilizamos lo que los lingüistas llaman expresiones referenciales o *descripciones definidas*, como “el pequeño perro negro” en lugar del más eficiente “Blackie”, cuando hablamos con alguien que no sabe que ese es el nombre propio del animal.

Del mismo modo, cuando utilizamos un catálogo de biblioteca o un motor de búsqueda para localizar un recurso conocido, lo buscamos utilizando el título de la obra (su nombre, en sentido amplio) o alguna información específica que conozcamos sobre él, para que sea más fácil de encontrar. Por el contrario, cuando buscamos recursos para satisfacer una necesidad de información, pero no tenemos en mente recursos específicos, los buscamos utilizando descripciones de su contenido u otras propiedades. En general, la recuperación de información puede caracterizarse como la comparación de la descripción de las necesidades de un usuario con las descripciones de los recursos que podrían satisfacerlas.



3.2.2. “DESCRIPCIÓN” COMO TÉRMINO INCLUSIVO

Hasta ahora hemos utilizado el concepto de “descripción” en su sentido ordinario para referirnos a la clasificación o la explicación de los rasgos visibles o importantes que caracterizan o representan algo. Sin embargo, el concepto se utiliza a veces con mayor precisión en el contexto de los Sistemas de Organización, en los que la descripción de recursos suele ser más formal, sistemática e institucional. En el contexto de la actividad bibliotecaria, en los procesos técnicos de descripción bibliográfica, un *descriptor* es uno de los términos que en un determinado lenguaje diseñado específicamente para esa función puede asignarse a un recurso para designar sus propiedades, características o significado, o sus relaciones con otros recursos. En los contextos del diseño de sistemas de información y del modelado conceptual, se usan términos como “palabras clave”, “términos de índice”, atributos, valores de atributos, elementos, “elementos de datos”, “valores de datos” o “vocabulario” para aspectos relativos a la descripción de recursos. En el campo de la inteligencia de negocio, el análisis predictivo u otros contextos de la ciencia de datos, estos se llaman “variables”, “características”, propiedades o “medidas”. En contextos en los que las descripciones son menos formales o más personales, los términos de descripción suelen llamarse “etiquetas” o “tags”. En lugar de intentar hacer distinciones específicas entre estos sinónimos o cuasi-sinónimos, utilizaremos “descripción” como un término inclusivo, excepto cuando el uso convencional favorezca abrumadoramente uno de los otros términos.

Muchos de estos términos provienen de un ámbito semántico muy delimitado en el que el propósito de la descripción es identificar y caracterizar sobre lo que trata un recurso (aboutness) o, dicho de otro modo, la esencia de su contenido. Sin embargo, a medida que se vuelve trivial asociar a los recursos información descriptiva que ha sido generada por procesos informáticos, aparecen muchos tipos adicionales de información, más allá de la que estrictamente se refiere a “de qué trata”, que pueden apoyar interacciones adicionales. Describiremos muchas de



estas nuevas orientaciones y los tipos de información necesarios para permitirlos en el apartado §3.3.2 *Determinación de los propósitos*. Aplicamos los conceptos de descripción de los recursos de manera expansiva para acomodarlos a todos ellos. Usamos el término *descripción de recursos* de manera general para implementar todos los significados.

Partimos de la anteriormente expresada distinción y coexistencia de recursos primarios y recursos que los describen, a los que denominamos como recursos de descripción, metadatos o recursos descriptivos. Elegimos este término (recurso descriptivo) como una alternativa más inclusiva y más fácil de entender que dos términos bien establecidos en la organización de sistemas de recursos de información: descripción bibliográfica y metadatos. También distinguiremos la descripción de recursos como concepto general de otros usos específicos como en la descripción estadística, el etiquetado de recursos web y el lenguaje RDF (Resource Description Framework) utilizado para hacer declaraciones sobre recursos web y recursos físicos que pueden ser identificados en la web.

3.2.2.1. LA DESCRIPCIÓN BIBLIOGRÁFICA

El campo de la biblioteconomía y documentación (Library and Information Science) ha debatido y sistematizado durante casi dos siglos los propósitos y la naturaleza de la descripción bibliográfica, que constituye uno de sus cimientos. *Las descripciones bibliográficas* caracterizan los recursos de información y las entidades que pueblan el universo bibliográfico, que incluyen las obras, las ediciones, los autores y los temas.

La descripción bibliográfica de un recurso de información suele realizarse como un registro estructurado en un formato estandarizado que describe un recurso específico que ha sido, por lo general, publicado.

3.2.2.2. LOS METADATOS

Los metadatos suelen definirse como “datos sobre datos”, definición que es casi tan ubicua como inútil. Una definición más completa de metadatos sería “una descripción estructurada de recursos de información de cualquier tipo”. Los metadatos son más útiles cuando están respaldados por un esquema de metadatos que define los elementos para una descripción estructurada.

El concepto de metadatos se originó en el diseño de sistemas de información y bases de datos en la década de 1970, por lo que es mucho más reciente que el de descripción bibliográfica. Los primeros esquemas de metadatos, llamados diccionarios de datos, documentaban

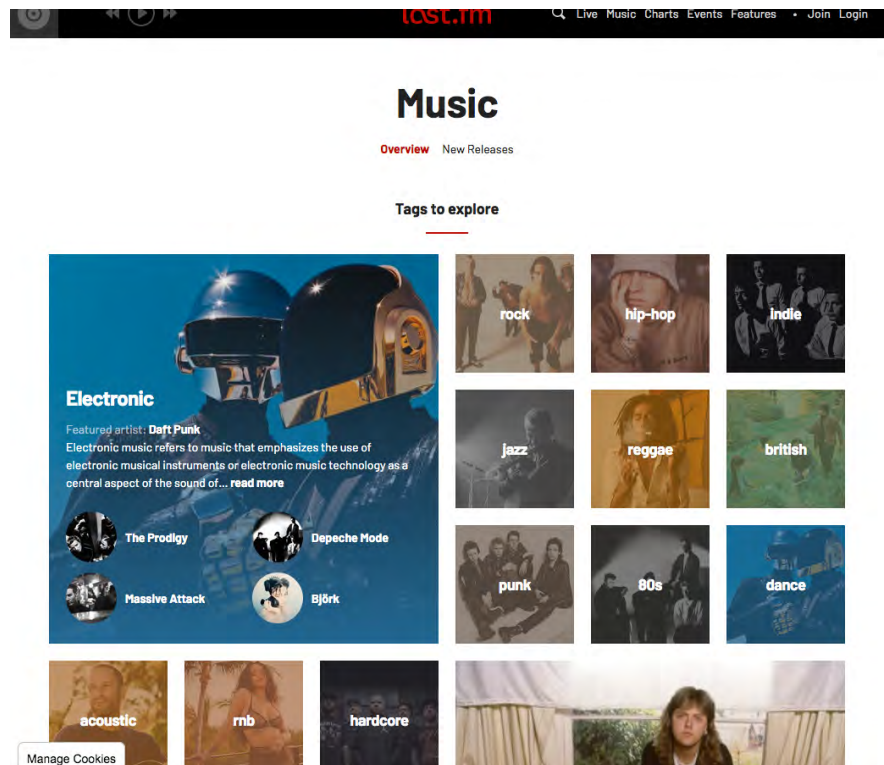


la disposición y el contenido de los campos de datos en los registros utilizados por las aplicaciones transaccionales en los grandes ordenadores centrales. Un tipo más sofisticado de metadatos surgió con la documentación de los modelos de datos en los sistemas de gestión de bases de datos, llamados esquemas de bases de datos, que describían la estructura de las tablas relacionales, los nombres de los atributos y los valores de contenido y tipos de datos válidos en cada campo.

Hoy en día, los esquemas XML y otros formatos de descripción de recursos de fácil uso en las páginas web y en aplicaciones informáticas han ampliado la idea de descripción de recursos mucho más allá de la descripción bibliográfica para incluir la descripción de componentes de software, conjuntos de datos empresariales y científicos, servicios web y objetos computacionales tanto en formato físico como digital. Las propias descripciones de recursos sirven para permitir el descubrimiento, la reutilización, el control de acceso y la invocación de otros recursos necesarios para que los usuarios o los agentes informáticos interactúen eficazmente con los recursos primarios descritos por los metadatos.

3.2.2.3. ETIQUETADO DE LOS RECURSOS WEB

El concepto de metadatos se ha ampliado para incluir las etiquetas, las clasificaciones, los marcadores en libros electrónicos u otros tipos de descripciones que una persona puede aplicar a fotografías personales, entradas de un blog, noticias, o cualquier otro recurso con presencia en la web. La práctica del etiquetado (tagging) surge como una forma de aplicar rótulos o etiquetas (labels) al contenido para describirlo e identificarlo. Los conjuntos de etiquetas son útiles para administrar una recopilación de sitios web o medios digitales que interesan a una persona, para compartirlos con otros y para permitir nuevos tipos de interacciones y servicios. Por ejemplo, los usuarios de **Last.fm** etiquetan la música con etiquetas que describen su naturaleza, época, estado de ánimo o género, y Last.fm utiliza esas etiquetas para generar estaciones de radio que reproducen música similar a esa etiqueta y otras relacionadas.



Last.fm

Last.fm analiza etiquetas y otros metadatos para crear páginas multimedia enriquecidas que sirvan para el “descubrimiento”, esto es para, aplicando su algoritmo, proponer artistas, canciones nuevas, descargas gratuitas y videos musicales que satisfagan los gustos de los usuarios. Esto permite que los usuarios exploren las novedades musicales de manera más intuitiva que la mera búsqueda por género y artista

No obstante, el etiquetado tiene un inconveniente. La tendencia de los usuarios a etiquetar intuitiva y espontáneamente revive el “problema de vocabulario” porque “árbol” para un fotógrafo es “roble” para otro. Del mismo modo, la elección poco sistemática de palabras conduce a una incoherencia morfológica; la misma fotografía puede ser etiquetada con “quemar” y “árboles”, y también con “arder” y “árbol” por otro usuario. Esta disparidad en los descriptores que la gente utiliza para categorizar recursos que son los mismos o similares puede convertir los sistemas que utilizan etiquetas en una “sopa de etiquetas” carente de estructura².

² Se utilizan aquí conceptos como el “problema del vocabulario” y la “incoherencia morfológica”, sobre los que el lector puede encontrar información más detallada en los apartados §2.4. y §2.4.3 de este mismo libro.



3.2.2.4 RDF (MARCO PARA LA DESCRIPCIÓN DE RECURSOS)

El Marco o Infraestructura para la Descripción de Recursos (RDF, Resource Description Framework) es un modelo estándar para hacer declaraciones procesables por ordenador sobre recursos en la web; es la base de la visión de la Web Semántica. Hemos estado usando la palabra “recurso” para referirnos a cualquier cosa que se esté organizando. En el contexto de RDF y la web, sin embargo, “recurso” significa algo más específico: un recurso es cualquier cosa a la que se le ha dado un Identificador Uniforme de Recursos (URI). Las URIs pueden adoptar diversas formas, pero probablemente se esté más familiarizado con los URI utilizados para identificar páginas web, como <http://springfield-elementary.edu/>. (Probablemente esté acostumbrado también a llamarlas simplemente URL - o más coloquialmente “dirección web” o “enlace” - en lugar de URI). La idea clave detrás de RDF es que podemos usar URIs para identificar no sólo cosas “en” la web, como páginas web, sino también cosas “fuera” de la web como personas o países. Por ejemplo, podemos usar la URI <http://springfield-elementary.edu/> para referirnos a la propia escuela primaria de Springfield y no sólo a la página web de la escuela.

RDF modela todas las descripciones como conjuntos de “tripletas”, donde cada tripleta consiste en el recurso que se describe (identificado por un URI), una propiedad y un valor. Las propiedades también son recursos, lo que significa que están identificadas por URIs.

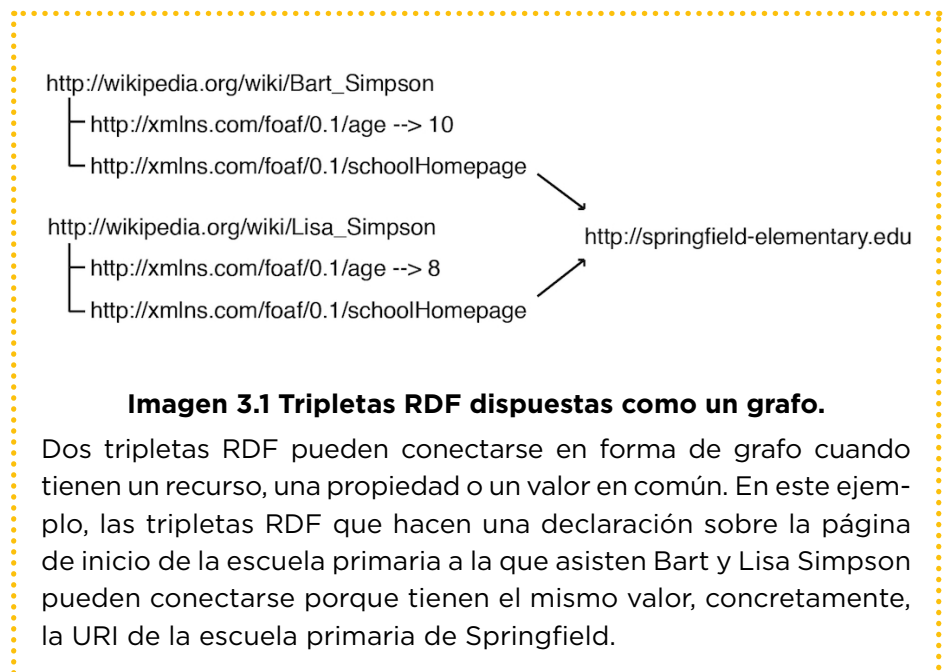


Imagen 3.1 Tripletas RDF dispuestas como un grafo.

Dos tripletas RDF pueden conectarse en forma de grafo cuando tienen un recurso, una propiedad o un valor en común. En este ejemplo, las tripletas RDF que hacen una declaración sobre la página de inicio de la escuela primaria a la que asisten Bart y Lisa Simpson pueden conectarse porque tienen el mismo valor, concretamente, la URI de la escuela primaria de Springfield.



El uso de URIs como identificadores de recursos y propiedades permite que las descripciones modeladas conforme a RDF se interconecten en una red de “datos enlazados” (o también, “datos vinculados”), de la misma manera que la red internet permitió que la información se interconectara en una red masiva de “documentos vinculados”. Los defensores del RDF afirman que esto beneficiará enormemente el descubrimiento y la inferencia de conocimientos. Pero los beneficios de la forma de descripción altamente prescriptiva que supone RDF deben sopesarse con sus costes; la conversión de las descripciones ya existentes a RDF puede ser un trabajo extenuante y técnicamente exigente.

3.2.2.5. LOS OBJETOS DE INFORMACIÓN AGREGADOS

En la era pre-digital, los objetos de información tenían unos límites tangibles explícitos. Los libros consistían en páginas encuadradas dentro de una cubierta, un disco de vinilo solo poseía un conjunto de canciones (incluso se podía ver el patrón de ranuras que separaba las canciones), una película se entregaba en una tira de celuloide enrollado en un carrete y una colección se demarcaba (normalmente) como una serie de estanterías o depósitos identificados en una biblioteca.

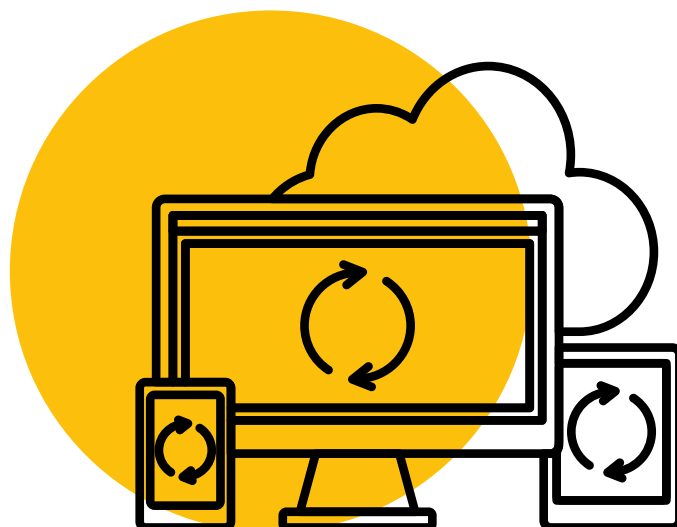
Los límites de los objetos de información en el reino digital no son ni tangibles ni obvios. Consideremos la simple noción de una página web. Nuestra noción cognitiva de lo que se muestra en la ventana de nuestro navegador (por ejemplo, algún texto formateado con una imagen asociada) es en realidad, en términos de arquitectura web, tres objetos de información (también conocidos como recursos); el código HTML que codifica el texto, el CSS que define las reglas de formato y el JPEG que codifica la imagen. Los tres tienen URLs y pueden ser recuperados y enlazados independientemente. La situación es aún más ambigua en lo que respecta a la noción común de un sitio web, cuyos límites no están definidos técnicamente y son cognitivamente difíciles de expresar.

Las agregaciones pueden ser mecanismos de conveniencia para simplificar la difusión o la organización, pero también pueden ser transformadoras; casi todo el valor de un recurso puede derivar de su inclusión en una agregación. En una página web, el archivo CSS es virtualmente inútil por sí solo, ya que su función es dar estilo al archivo HTML. En iTunes, las funciones de reproducción y organización están optimizadas para la música pop, donde las canciones pueden funcionar por sí solas, independientemente de las demás del álbum al que pertenezca.



Los aficionados a la música clásica suelen tener estar menos cómodos con este enfoque, porque las “pistas” individuales de una grabación, divididas para reducir el tamaño del archivo y facilitar la navegación por las obras largas, no son separables; las piezas están pensadas para ser escuchadas en su totalidad, y puede ser difícil asegurar que se agreguen juntas y que se asignen los metadatos adecuados a sus agregaciones. En otras palabras: no se pueden escuchar las sinfonías en modo aleatorio.

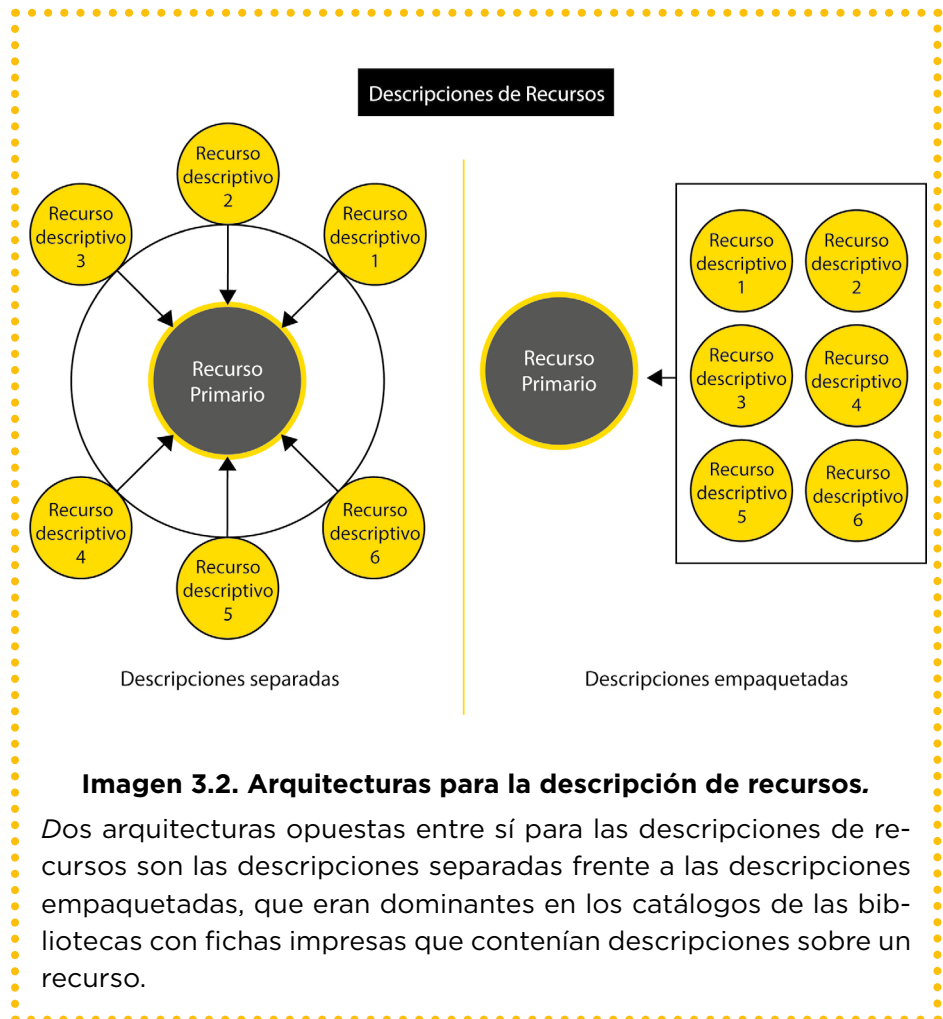
El problema aquí es cómo expresar arquitectónica y técnicamente la idea de una agregación: **un conjunto de objetos de información que, cuando se consideran juntos, componen otro objeto de información concreto.** Las agregaciones se encuentran en todo nuestro espacio de información digital: la página web y el sitio mencionados anteriormente; una publicación académica que consiste en texto, figuras y datos; un conjunto de datos que es la composición de múltiples archivos de datos. En particular, la idea de agregación es al mismo tiempo recurrente y no exclusiva. Un objeto que es en sí mismo una agregación puede ser agregado a otro objeto. Los objetos de información incluidos en una agregación pueden también incluirse en otras agregaciones, lo que permite la reutilización y la “refactorización” (producción de un resultado análogo al existente previamente, aplicando una combinación diferente y más optimizada de procedimientos y recursos) de los objetos de información existentes. La solución a este problema es un aspecto crítico para el campo de la organización de información digital porque, sin límites bien definidos, no podemos identificar, referenciar o describir los objetos de información.





3.2.3. MARCOS PARA LA DESCRIPCIÓN DE RECURSOS

El amplio espectro de recursos a los que se pueden aplicar descripciones y las diferentes comunidades que las describen significa que se han propuesto muchos marcos y clasificaciones para ayudar a dar sentido a la descripción de recursos.



El punto de vista histórico dominante trata las descripciones de recursos como un paquete de declaraciones; este punto de vista es el que está incorporado en la clásica ficha impresa del catálogo de biblioteca y su análogo informatizado en el formato MARC21 (un formato de intercambio para registros de catálogo de biblioteca), que contiene muchos campos sobre las características bibliográficas de un objeto en campos tales como autor, título, año de publicación, editor y paginación. Una arquitectura alternativa para la descripción de recursos se centra en cada descripción o declaración individual sobre un solo re-



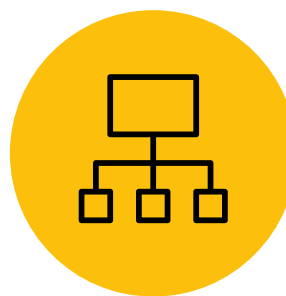
curso, como se hace en RDF y los enfoques de datos enlazados. Estas dos alternativas se contrastan en la Imagen 3.2, Arquitecturas para la descripción de recursos.

En cualquier caso, estas formas de pensar sobre la descripción de recursos enfatizan, o tal vez incluso exageran, dos decisiones de implementación:

- La primera consiste en combinar múltiples descripciones de recursos en un paquete estructurado o mantenerlas como declaraciones descriptivas separadas.
- La segunda es la elección de la sintaxis en la que se codifican las descripciones.

Ambas decisiones de implementación tienen importantes repercusiones, pero son secundarias para las cuestiones relativas a los fines de la descripción de recursos, la forma en que se seleccionan sus propiedades como base para la descripción, la mejor forma de crearlas y otras consideraciones lógicas o de diseño. De acuerdo con una idea fundamental de la disciplina de la organización (introducida en el §1.6 *El concepto de “principio de organización”*), **es imperativo distinguir los principios de diseño de las decisiones de implementación**. Entendamos el conjunto de decisiones de implementación sobre notación de caracteres, sintaxis y estructura como la forma de descripción de recursos y aplazaremos su abordaje todo lo que podamos.

La descripción de recursos no es un fin en sí misma. La variedad de objetivos que se persiguen al describir recursos son, en realidad, medios para habilitar y usar un Sistema de Organización para alguna colección de recursos. Como resultado, nuestro marco para la descripción de recursos se ajusta a las actividades existentes en los Sistemas de Organización³.

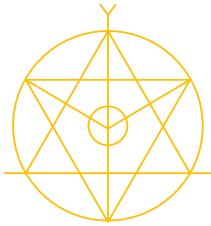


³ Recomendamos al lector repasar estos conceptos sobre *forma e interacciones* en los capítulo 9 y 3 respectivamente, de la edición completa online del libro “TDO core concepts edition”.



3.3. EL PROCESO DE DESCRIPCIÓN

DE RECURSOS



Preferimos, por tanto, el concepto general de descripción de recursos a los más especializados de descripción bibliográfica y metadatos porque facilita la visión de las cuestiones que trascienden los dominios en los que predominan esos términos. Además, nos permite proponer un proceso más estándar que podemos aplicar ampliamente al uso de descripciones de recursos en cualquiera de los Sistemas de Organización que se utilicen en un contexto concreto. **Un vocabulario compartido posibilitará compartir lecciones y buenas prácticas.**

El proceso de descripción de recursos se compone de siete pasos interdependientes e iterativos. Comenzamos con un resumen genérico del proceso para establecer el escenario previo a una discusión paso a paso.

- 1 // Identificar los recursos que vamos a describir es el primer paso. El *dominio* y el *alcance* de los recursos circunscriben las propiedades descriptibles y los potenciales propósitos para los que servirán las descripciones. El *foco* de un recurso determinará cuáles son los recursos primarios de información y cuáles serán tratados como sus correspondientes recursos descriptivos. Dos decisiones importantes en esta etapa son la granularidad de la descripción - ¿Describimos recursos individuales o colecciones de ellos? - y el nivel de abstracción- ¿estamos describiendo instancias de recursos, partes de ellos o tipos de recursos?⁴
- 2 // En general, el propósito de la descripción de recursos es apoyar las actividades comunes a todos los Sistemas de Organización: *seleccionar, organizar, interactuar y mantener los recursos*⁵. El dominio específico y el contexto de los recursos, y el contexto en el que se crean y utilizan las descripciones, impondrán requisitos y limitaciones más específicas a los fines que se pretenden conseguir con ellas.
- 3 // Una vez que se han determinado los propósitos de la descripción en términos de actividades e interacciones, se pueden identificar qué *propiedades* específicas de los recursos se necesi-

4 Estos aspectos se trataron con más detalle en el apartado §2.3 “Identidad de los recursos”.

5 Las actividades e interacciones son el tema central del capítulo 3 de la edición completa original del libro “TDO core concepts edition”.



tan para posibilitarlas. El objetivo de la descripción no es ser exhaustivo; siempre hay más propiedades posibles de las que se pueden describir razonablemente⁶. En cambio, el reto consiste en utilizar las propiedades más sólidas y fiables para apoyar las interacciones deseadas.

- 4 // Este paso incluye varias *decisiones lógicas y semánticas* sobre cómo se describirán las propiedades de un recurso. ¿Qué términos o nombres de elementos deben ser usados para identificar las propiedades de los recursos que hemos elegido para describirlo? ¿Existen reglas o restricciones sobre los tipos de datos o valores que las descripciones de las propiedades pueden asumir? Cuando se trata de descripciones numéricas, sus tipos de datos y niveles de medición limitan los tipos de procesamiento a los que pueden someterse. Los tipos de dato nominales, ordinales, de intervalo y de coeficientes están cada uno de ellos limitados a una serie de transformaciones propias que están basadas en lo que representan. Un buen vocabulario de descripción tendrá que ser fácil de asignar cuando se creen descripciones de recursos y fácil de entender cuando se utilicen en la búsqueda.

- 5 // Las decisiones lógicas y semánticas sobre el vocabulario de descripción se definen mediante *decisiones sobre la notación, la sintaxis y la estructura de las descripciones*. En conjunto, estas decisiones determinan lo que llamamos la forma o codificación de las descripciones de recursos. La implementación de las descripciones implica decisiones sobre cómo y dónde se almacenan y la tecnología utilizada para crearlas, editarlas, almacenarlas y recuperarlas.

- 6 // Las descripciones de recursos son *creadas por individuos, por grupos informales o formales de personas, o por medios automatizados o computacionales*. Algunos tipos de descripciones sólo pueden ser creadas por personas, otros tipos de descripciones sólo pueden ser producidas mediante técnicas automatizadas o algorítmicas, y algunas pueden ser generadas combinando ambas maneras.

⁶ Se da la paradoja de que la creación de metadatos descriptivos podría hipotéticamente extenderse tanto que resultara en una copia a escala 1:1 del propio contenido, como un mapa que tuviera la misma escala que el territorio representado, e incluso extenderlo, de forma que la “edición anotada” o “marcada” contenga más datos y ocupe más espacio que el documento original. De hecho la descripción puede, desde cierto punto de vista, no solo sustituir al contenido sino multiplicarlo (N.E. en español)



7 // Los recursos descriptivos deben ser *evaluados* con respecto a los propósitos para los que estaban previstos. Los resultados de esta evaluación ayudarán a determinar si hay que rehacer alguno de pasos anteriores.

Las siguientes siete subsecciones discuten cada uno de estos pasos en detalle. Una guía de referencia rápida es la Imagen 2.3, *El proceso de descripción de los recursos*.

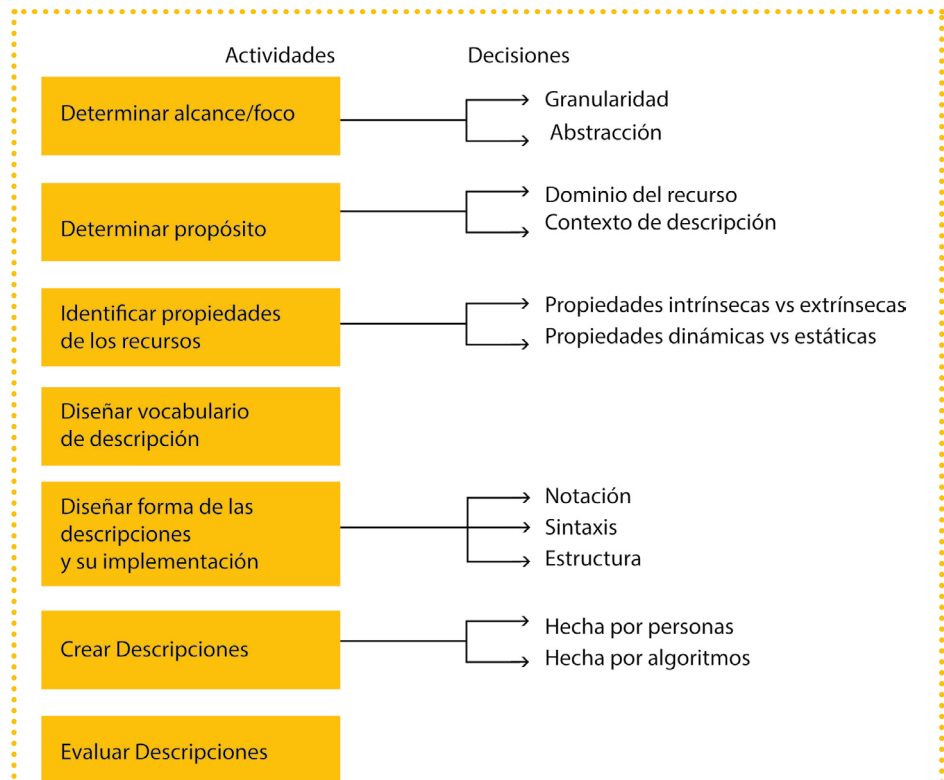


Imagen 2.3. El proceso de descripción de recursos

El proceso de descripción de recursos consta de siete pasos: Determinar el alcance y el enfoque, determinar el propósito, identificar las propiedades de los recursos, diseñar el vocabulario de descripción, diseñar la forma de descripción y su implementación, crear las descripciones y evaluarlas.

El grado de explicitación y sistematicidad que debe tener cada paso depende del dominio y alcance del recurso, y especialmente de cuáles serán los usuarios previstos del sistema de organización. Si observamos con atención, podemos ver que la mayoría de estos pasos tienen lugar incluso en contextos muy informales, como los niños que juegan con piezas Lego con los que empezamos este capítulo. El objetivo de construir cosas con las piezas lleva a los niños a identificar qué propie-



dades son más útiles para analizarlas. Desarrollan descripciones de las piezas que capturan los valores específicos de propiedades que ellos consideran relevantes. Finalmente, ponen a prueba – evalúan - sus descripciones usándolas cuando juegan juntos; se hace inmediatamente obvio que una descripción no cumple su propósito cuando un niño le entrega a otro una pieza que no es la que él pensaba que había pedido.

Por el contrario, un escenario como el de la toma de fotografías digitales implica un proceso mucho más explícito y sistemático de descripción de recursos. Las propiedades de los recursos, el vocabulario de descripción y la forma de descripción que utiliza automáticamente una cámara digital fueron seleccionados por una asociación de empresas y publicados como una especificación técnica implementada por los fabricantes de cámaras y teléfonos móviles de todo el mundo.

Las descripciones de recursos utilizadas por las bibliotecas, los archivos y los museos suelen crearse de manera aún más explícita y sistemática. Al igual que las descripciones de la fotografía digital, las propiedades, el vocabulario y la forma de las descripciones utilizadas por sus sistemas de organización se rigen por normas. Sin embargo, no existe un equivalente a la cámara digital que pueda crear estas descripciones automáticamente. En su lugar, son creadas meticulosamente por profesionales.

Muchos recursos y sus descripciones asociadas en los sistemas de organización empresariales y científicos se crean mediante procesos automatizados o computacionales, por lo que el proceso de descripción de los recursos individuales no es en absoluto similar al de las bibliotecas y otras instituciones de la memoria. Sin embargo, el proceso de diseño de los modelos o esquemas de datos para la clase de recursos que se generarán es igualmente sistemático y suele ser realizado por analistas de datos y analistas de sistemas altamente cualificados.

3.3.1. DETERMINAR EL ALCANCE Y EL FOCO

¿Qué recursos queremos describir? Determinar lo que se considerará como un recurso individualizado no siempre es fácil, especialmente en el caso de los recursos compuestos por partes, así como en el caso de los recursos de información en los que su propiedad más importante sea su contenido, el cual no es directamente perceptible. Identificar lo que se quiere describir con la mayor precisión posible es el primer paso para crear una descripción útil.



Al hablar del “enfoque” sobre los recursos, estamos introduciendo, como vimos anteriormente, la diferenciación entre los recursos primarios y los recursos descriptivos; lo que denominamos “el foco” en los recursos. La determinación del enfoque sobre un recurso va de la mano de la decisión sobre cuáles son los recursos que pretendemos describir; estas decisiones, a menudo arbitrarias, marcan por tanto una enorme diferencia en la naturaleza y el alcance de la descripción de los recursos. **Los metadatos de una persona pueden ser los datos de otra.**

- ❑ Para un bibliotecario, el precio de un libro puede ser simplemente otro atributo más que forma parte del registro del libro.
- ❑ Para un contable en una librería, el precio de ese libro, tanto el coste de comprar el libro como el precio al que se vende a los clientes, es una información crítica para mantenerse en el negocio.
- ❑ En el contexto de los registros médicos, el proveedor de seguros de un paciente no es de mucha importancia para el médico, pero para la persona responsable de la facturación es fundamental. Para la enfermera, los signos vitales actuales del paciente pueden ser de suma importancia, mientras que para el médico puede ser más importante comprender cómo esos han evolucionado esos datos en conjunto y servir para indicar un pronóstico a largo plazo de la salud del paciente.
- ❑ Un científico que estudia anatomía comparada preserva los especímenes animales y registra descripciones físicas detalladas sobre ellos, pero un científico que estudia la ecología o la migración descarta los especímenes y se centra en la descripción del contexto en el que se encontraba el espécimen.

3.3.1.1. DESCRIBIR CONTENIDOS INDIVIDUALES O DESCRIBIR COLECCIONES

Lo más simple es pensar que la descripción de un recurso está asociada a otro recurso individual (lo que en términos comunes entenderíamos como “su ficha”). Pero ya hemos señalado que supone un desafío determinar qué se puede considerar un recurso individual cuando los recursos son en sí mismos objetos o sistemas compuestos de otras partes o recursos. Por ejemplo, a veces describimos un equipo de fútbol como un único recurso y otras veces nos centramos en cada jugador



individual. Sin embargo, después de decidir sobre la granularidad del recurso, la pregunta sigue siendo si cada recurso puede necesitar también una descripción separada.

Las bibliotecas y los museos se especializan en la conservación de descripciones de recursos sobre los componentes individuales (instancias) de sus colecciones. Las descripciones de recursos también se aplican a clases o colecciones de recursos, porque una colección también es un recurso (Como se trata en el apartado §1.4 *El concepto de "colección"*). Los archivos y las colecciones especiales de mapas suelen tener asignadas descripciones de recursos por agregación, pero cada documento o mapa que forma parte de la colección no tiene necesariamente su propia descripción bibliográfica. Del mismo modo, los conjuntos de datos comerciales y científicos se describen invariablemente a nivel de granularidad de la colección porque a menudo se analizan en su totalidad.

3.3.1.2. NIVEL DE ABSTRACCIÓN EN LOS RECURSOS DESCRIPTIVOS

La forma de hacer una descripción también puede estar asociada a un tipo o dominio completo de recursos. Una colección de descripciones de recursos es mucho más útil cuando cada recurso se ha descrito utilizando elementos de descripción o términos comunes que se aplican a todos ellos. Un *esquema* (o modelo, o norma de metadatos) especifica el conjunto de descripciones que se aplican siempre sobre un tipo de recurso (un disco, un mapa, una ley...). A veces este esquema, modelo o norma se infiere de una colección de recursos existentes o se impone sobre ella para garantizar definiciones más coherentes, pero con mayor frecuencia se utiliza como especificación cuando los recursos se crean o generan por primera vez.⁷

Una base de datos relacional, por ejemplo, se conceptualiza fácilmente como una colección de registros organizados en una o más tablas, en las que cada registro en su propia fila tiene una serie de campos o atributos que contiene algún tipo de contenido prescrito. Cada registro o fila de la tabla de la base de datos es una descripción de un recurso (un empleado, un producto, cualquier cosa) y los valores de los atributos individuales, organizados por columnas y filas de la tabla, son partes distintas de la descripción para alguna instancia particular del recurso, como el empleado 24 o el producto 8012C.

⁷ Sobre estos aspectos puede ampliarse leyendo los apartados §3.5.2.4 "Preservar los tipos de recursos"; §4.2.1 "El dominio de recursos" y el bloque titulado "¿Qué sucede con la "creación" de recursos?" en la introducción del apartado §3.1 de la edición online completa del libro "TDO core concepts edition".



Los recursos de información que comúnmente llamamos documentos son, por su naturaleza, menos homogéneos en cuanto a su contenido y estructura que los que se pueden gestionar en bases de datos. Los esquemas de los documentos, comúnmente representados en SGML o XML, suelen permitir una mezcla de elementos descriptivos de datos y texto: amalgaman contenido y estructura.

3.3.1.3 EL ALCANCE, LA ESCALA Y LA DESCRIPCIÓN DE RECURSOS

Si sólo tuviéramos una cosa que describir, podríamos usar una sola palabra para describirla: “eso”. No necesitaríamos distinguirla de cualquier otra cosa. Un segundo recurso implica al menos un término más en el lenguaje de descripción: “no es eso”. Sin embargo, a medida que una colección crece, las descripciones deben hacerse más complejas para distinguir los recursos entre sí.

Cada elemento o término en un lenguaje de descripción crea una dimensión, o eje, a lo largo del cual se pueden distinguir los recursos, o define un conjunto de preguntas sobre los recursos. Las distinciones y preguntas que surgen con frecuencia deben ser fáciles de abordar, como, por ejemplo:

- **¿Cuál es el nombre del recurso?**
- **¿Quién lo creó?**
- **¿Qué tipo de contenido o material contiene?**

Por lo tanto, a medida que crece una colección, el lenguaje para describir los recursos debe ser más riguroso, y las descripciones creadas cuando la colección era pequeña tendrán que ser revisadas porque ya no serán adecuadas para los fines previstos.

Dado que la tarea de descripción de recursos de las bibliotecas se ha normalizado a nivel nacional e internacional, la labor de catalogación se distribuye entre muchos agentes descriptores (centros que catalogan recursos) cuyos resultados se comparten. El principio de la normalización ha sido la base de la descripción bibliográfica centralizada y coordinada durante un siglo.

La descripción centralizada de los recursos por profesionales cualificados funciona para las bibliotecas, pero incluso en los primeros días de la web, muchas voces, tanto desde las ciencias de la biblioteconomía y documentación, como de visionarios sobre de evolución de la web



reconocieron que este enfoque no se podría extender para describir los recursos de la web. En 1995, se propuso el conjunto de metadatos *Dublin Core (DC)* con sólo 15 elementos como un vocabulario de descripción mucho más sencillo que podrían utilizar personas sin formación técnica en catalogación. Desde entonces, la iniciativa Dublin Core ha sido muy influyente para inspirar a otras numerosas comunidades a crear vocabularios de descripción minimalistas, a menudo simplificando vocabularios que habían sido ideados por profesionales, permitiendo así su uso fuera de estos ámbitos profesionalizados, reduciendo barreras de entrada a la catalogación y descripción.



Por supuesto, un vocabulario de descripción más sencillo hace menos distinciones que uno complejo; la sustitución de “autor”, “artista”, “compositor” y muchas otras descripciones de la persona o el recurso no humano responsable del contenido intelectual de un recurso por sólo “creador” (como hace Dublin Core) da lugar a una pérdida sustancial de precisión cuando se crea la descripción y puede causar malentendidos cuando se reutilizan las descripciones.

Los efectos negativos del aumento del alcance y la escala en la descripción de los recursos pueden evitarse, a veces, si se contempla el alcance y la escala finales del Sistema de Organización cuando se está creando. No sería inteligente que una empresa con clientes en seis estados de los Estados Unidos creara un campo de dirección en su base de datos de clientes que sólo manejara esos seis estados; un diseño más extenso abarcaría cualquier estado o provincia e incluiría un código de país. Sin embargo, en general, al igual que hay problemas para adaptar un vocabulario sencillo conforme se amplía el alcance y la escala, el diseño y la aplicación de descripciones de recursos que funcionan para una colección grande y en continuo crecimiento seguramente supongan demasiado trabajo en el caso de aplicarse sobre una colección de pequeño tamaño.



3.3.2. ESTABLECER LOS PROPÓSITOS

La descripción de recursos sirve para muchos fines, y la mezcla de propósitos y los tipos de descripciones resultantes en cualquier sistema de organización particular depende del alcance y la escala de los recursos que se organizan. Podemos identificar y clasificar los propósitos más comunes utilizando las cuatro actividades que se producen en todo sistema de organización: seleccionar, organizar, interactuar y mantener los recursos del sistema. La descripción de los recursos también tiene un propósito más abierto en el campo del *sensemaking* y el conocimiento científico (aspectos que trataremos poco más adelante en el apartado §3.3.2.5); observamos y describimos el mundo para dar sentido a nuestras experiencias y para predecir observaciones futuras.

3.3.2.1. LA DESCRIPCIÓN DE RECURSOS PARA APOYAR LA SELECCIÓN

Al definir la selección como el proceso mediante el cual los recursos se identifican, evalúan y luego se añaden a una colección en un sistema de organización, se enfatizan las descripciones de recursos creadas por alguien que no sea la persona que los utiliza. Podemos distinguir varias formas diferentes en las que la descripción de recursos apoya la selección:

DESCUBRIMIENTO

- ¿Qué recursos disponibles podrían añadirse a una colección? Los nuevos recursos suelen aparecer en directorios, registros o catálogos. Algunos tipos de recursos se seleccionan y se adquieren automáticamente mediante suscripciones o contratos.

CAPACIDAD Y COMPATIBILIDAD

- ¿Cumplirá el recurso con los requisitos funcionales o de interoperabilidad? Los recursos con un componente tecnológico intensivo (una app, un software) suelen tener numerosos tipos de descripciones especializadas que especifican su funcionalidad, fiabilidad y otras “- ilidades” que determinan si son compatibles con otros recursos en un Sistema de Organización. Algunos servicios tienen cualidades de niveles de servicio, términos y condiciones, o interfaces documentadas en sus descripciones que afectan a su compatibilidad e interoperabilidad. Algunos recursos tienen licencias o restricciones de uso que pueden impedir que los recursos se utilicen eficazmente para los fines previstos. Las decisiones sobre la “selección de personas” están cada vez más basadas en datos y los



equipos deportivos, los empleadores empresariales y los sitios de citas dependen ahora de cálculos estadísticos de predicción para encontrar a la persona más adecuada.

AUTENTICIDAD

- ¿Es el recurso lo que dice ser? Las descripciones de recursos que pueden apoyar la autenticación incluyen las tecnológicas, como sellos de tiempo, marcas de agua, encriptación, *checksums* y firmas digitales. La historia de la propiedad o custodia de un recurso, llamada su procedencia (provenance), se establece a menudo a través de la asociación con otros registros como compra-ventas o declaraciones de impuestos.

VALORACIÓN

- ¿Cuál es el valor de este recurso? ¿Cuál es su coste? ¿A qué ritmo se deprecia? ¿Tiene una vida útil? ¿Tiene alguna calificación, clasificación o medidas de calidad asociadas? Además, ¿cuál es la calidad de esas clasificaciones y medidas?

También consideramos otra perspectiva complementaria, la de la persona que crea o produce la descripción del recurso y su propósito principal, que suele ser fomentar la selección del recurso por parte de otra persona (situación que es central en el contexto del comercio electrónico). La comercialización de productos consiste en idear nombres y descripciones para hacer que un recurso sea distintivo y atractivo en comparación con otras alternativas. Durante muchos años las ciruelas pasas se promovieron como un suplemento dietético que las personas (especialmente las de edad avanzada) necesitaban para “mantener la regularidad”. Pero después de que la Junta de Ciruelas Pajas de California (el mayor proveedor del mundo) las rebautizara como “ciruelas secas” y comenzara a comercializarlas como un snack (y al mismo tiempo cambió su nombre por el de Junta de Ciruelas Secas de California) las ventas aumentaron considerablemente.

Muchos países exigen que los productos importados lleven una etiqueta con su país u origen. Los consumidores suelen utilizar esta propiedad en las descripciones de los recursos como indicador de alta calidad, como puede ocurrir con los relojes suizos, la moda francesa o italiana o el beicon canadiense. Otra posibilidad es que los consumidores quieran comprar productos nacionales o de origen local por patriotismo económico o para cumplir con ciertas exigencias normativas que regulen un mercado. No es sorprendente que cuando los consumidores ven el origen de manera positiva, esta información aparezca destacada y fácil de leer. En cambio, cuando los consumidores consideran el origen de



manera menos positiva, tal vez como una advertencia de productos de baja calidad, es probable que el proveedor haga que la información sobre el origen sea lo menos llamativa posible dentro del margen permitido por la ley, o incluso que pueda dar a entender de forma engañosa que el producto sea nacional.

Esta tergiversación también es omnipresente en las citas online, aunque la cantidad de tergiversación debe equilibrarse con los objetivos de la relación y las posibilidades de que se descubra el engaño.

3.3.2.2. LA DESCRIPCIÓN DE RECURSOS PARA APOYAR LA ORGANIZACIÓN

Hemos definido la organización como las actividades para concretar, a partir de unos principios, la forma de describir y disponer recursos para crear capacidades en las que se basarán ciertas interacciones. Esta definición trata la creación de descripciones de recursos y su uso para organizar los recursos para las interacciones como actividades separadas y secuenciales. Esta separación es más fácil de percibir cuando quienes asignan palabras clave y clasificaciones a los documentos, o cuando los sensores producen datos, y estas descripciones de recursos se utilizan más tarde para permitir la recuperación de documentos o el análisis de datos. Un empleado de un gran almacén podría clasificar camisas de vestir usando etiquetas que describen sus marcas, sus tallas y otras propiedades. Las normas que rigen la recopilación, la integración y el análisis de información personal son también descripciones de recursos que influyen en la organización de los recursos de información.

PÁRATE A PENSAR: ANUNCIOS PARA VENDER CASAS

Los anuncios de inmobiliarias son conocidos por el uso de descripciones con una gran dosis de “creatividad optimista”; una casa que se vende como “bien comunicada” estará probablemente al lado de una autopista concurrida, y una casa “tranquila” estará en una zona remota de la ciudad o en las afueras. ¿Cómo describiría su propia casa para convertir las cualidades negativas en positivas?

Sin embargo, aunque la descripción y la organización de recursos son lógicamente separables, a menudo están entrelazadas. Cuando organizamos nuestra propia ropa, no utilizamos descripciones de recursos explícitas y en su lugar nos basamos en las implícitas que usan propiedades fácilmente percibidas como el color, la forma y el material del que está hecha. Cuando son los algoritmos, en lugar de personas, las



que analizan textos para identificar y extraer sus características descriptivas para hacer funcionar aplicaciones como la recuperación de información, la clasificación de spam y el análisis de sentimientos, las descripciones de recursos y la organización de recursos coevolucionan de forma continua, a medida que el algoritmo se adapta y aprende con cada nuevo recurso que describe. Esta estrecha conexión entre la descripción de recursos y la organización de recursos también se explota en los sistemas de organización que utilizan los datos de uso de los registros de sesión (logs), la navegación o las actividades de descarga, como recursos de interacción, vinculándolos a los pagos por el uso de los recursos o analizándolos para influir en la selección y organización de los recursos en futuras interacciones personalizadas. (Recuérdese la explicación del apartado §1.9 *El concepto de “recurso de interacción”*).

3.3.2.3. LA DESCRIPCIÓN DE RECURSOS PARA APOYAR LAS INTERACCIONES

La mayoría de los debates sobre los propósitos de las descripciones de recursos y los metadatos enfatizan las interacciones que se basan en las descripciones de recursos que se han asignado de forma intencionada y explícita. Los Requisitos Funcionales para Registros Bibliográficos (FRBR) de IFLA, definidos en el campo profesional de la biblioteconomía y la documentación, especifican cuatro interacciones: la búsqueda, la identificación, la selección y la obtención de recursos, más una quinta añadida con posterioridad, Navegar o Explorar (Que forma parte del Modelo Conceptual de Referencia para Bibliotecas LRM que extiende y reformula el inicial de FRBR). Estas funciones cuadran sin problemas en casi cualquier otro Sistemas de Organización, no sólo en los de las bibliotecas.

ENCONTRAR

- ¿Qué recursos están disponibles que “corresponden a los criterios de búsqueda expresados por el usuario” y que, por lo tanto, pueden satisfacer una necesidad de información? Los usuarios modernos aceptan que la indización automatizada hace posible la búsqueda no sólo en todo el recurso de descripción, sino a menudo en todo el contenido del recurso primario.

IDENTIFICAR

- Otra finalidad de la descripción de recursos es permitir que un usuario confirme la identidad de un recurso específico o que distinga entre varios que tienen algunas descripciones que se solapan parcialmente. El color puede utilizarse en la descripción de un recurso cuando es necesario identificar rápidamente los recursos físicos.



SELECCIONAR



Seleccionar en este contexto significa la actividad del usuario de utilizar descripciones de recursos para apoyar una elección entre recursos de una colección, y no la actividad institucional previa de seleccionar recursos para formar la colección. Los motores de búsqueda suelen utilizar un breve “fragmento de texto” (snippet) con los términos de la consulta resaltados como descripciones de recursos para apoyar al usuario en su selección.

OBTENER



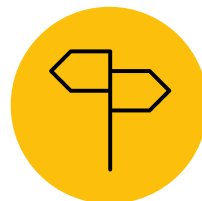
Obtener recursos físicos suele requerir un esfuerzo considerable después de haber sido seleccionados. Coger un autobús o un avión implica coordinar la ubicación y tiempo actuales con la hora y el lugar en que el recurso está disponible. En el caso de los recursos de información en forma física, la obtención de un recurso seleccionado suele significar un paseo por las estanterías de la biblioteca. En el caso de los recursos de información digital, un motor de búsqueda devuelve una lista de los identificadores de los recursos a los que se puede acceder con un solo clic, por lo que no se requiere mucho esfuerzo para pasar de la selección entre los resultados de la consulta a la obtención del recurso primario correspondiente. En la tienda online supone iniciar un proceso de compra, y dar el definitivo paso de gastar el propio dinero.

Elaine Svenonius propuso que una quinta interacción llamada Navegación se añadiera a la lista del modelo FRBR, lo que tuvo lugar en 2016, pero bajo el nombre de “Explorar”:

NAVEGAR O EXPLORAR



Si los usuarios no pueden especificar sus necesidades de información de la manera que requiere la funcionalidad de búsqueda, deberían poder utilizar descripciones relacionales y estructurales entre los recursos para navegar de cualquier recurso a otros que puedan ser mejores. Svenonius enfatiza la generalización, la agregación y las relaciones derivadas. Sin embargo, en principio, cualquier relación o propiedad podría servir como “puente” de navegación entre los recursos.





3.3.2.4. LA DESCRIPCIÓN DE RECURSOS PARA EL MANTENIMIENTO Y ADMINISTRACIÓN

Muchos tipos de descripciones de recursos que hemos comentado anteriormente para apoyar la selección, también son útiles a lo largo del ciclo de vida de un sistema, para dar soporte al mantenimiento tanto de los recursos específicos, como a la colección a la que pertenecen. En particular, la información técnica sobre los formatos y las especificaciones técnicas de los recursos (software, ordenadores u otros) necesarias para utilizarlos, así como la información necesaria para garantizar la integridad de los mismos, y que se denomina a menudo “metadatos de preservación” en un contexto de mantenimiento.

Las descripciones de recursos que se asocian más exclusivamente a las actividades de mantenimiento incluyen información sobre la versión y la eficacia, o información sobre la vida útil. Los programas de mantenimiento de equipos suelen estar relacionados con el número de kilómetros recorridos (indicados por el cuentakilómetros de un coche), el número de horas operadas (almacenadas por muchos motores), el número de páginas impresas, u otra información de fácil registro sobre el uso o las interacciones de los recursos. Con recursos inteligentes que ahora son capaces de capturar, analizar y comunicar más datos sobre su rendimiento en tiempo real, ahora es posible una predicción y programación más sofisticada de los trabajos de mantenimiento. También es más fácil identificar los recursos que no se están utilizando tanto como se esperaba, lo que podría implicar que ya no se necesitan y que, por lo tanto, pueden archivarse o desecharse con seguridad.

3.3.2.5. LA DESCRIPCIÓN DE RECURSOS PARA LA CIENCIA Y EL “SENSEMAKING”

Hasta ahora, a lo largo de todo este apartado §3.3.2, hemos discutido cómo se utilizan las descripciones de recursos para realizar tareas bien definidas dentro de un sistema de organización existente. Sin embargo, existe un propósito más amplio y menos definido para la descripción de recursos que es más antiguo y más fundamental: el uso de las descripciones de recursos como materia prima para dar sentido al mundo (*sensemaking*) o comprenderlo.

Durante miles de años, incluso antes de la invención del lenguaje escrito, la gente ha recopilado sistemáticamente cosas, información sobre esas cosas y observaciones de todo tipo para entender cómo funciona su mundo. Los humanos del paleolítico realizaron pinturas rupestres que representan los resultados de las cacerías y las migraciones de



animales; los antiguos egipcios registraron las inundaciones anuales del río Nilo en tallas de piedra; y los astrónomos babilonios, egipcios, chinos y mesoamericanos organizaron observaciones lunares, solares y planetarias en forma de calendarios que comenzaron hace unos cinco mil años.

Estos diversos esfuerzos por imponer un significado a la experiencia mediante el registro, el análisis, la organización y la reorganización de las observaciones pueden describirse colectivamente como “*sensemaking*”. (Ver el recuadro: El *sensemaking* y la organización).

Algunos aspectos del *sensemaking* están conectados con la evolución, que ha dado a nuestros cerebros poderosos mecanismos que automáticamente simplifican y organizan los datos perceptivos que obtenemos del mundo. No obstante, este proceso automático de *sensemaking* está dominado y amplificado por la necesidad de dar sentido a lo que nos rodea.

El ***sensemaking* intencional tiene lugar cuando se siguen conscientemente métodos estadísticos, experimentales y científicos sistemáticos para extraer y organizar el conocimiento de colecciones de muestras, observaciones o mediciones**. Es fundamental reconocer aquí que el contenido de estas colecciones representa las decisiones tomadas sobre lo que se debe recoger, porque la mayoría de las cosas y de los fenómenos tienen una gran cantidad de descripciones o propiedades que podrían registrarse, y los expertos o investigadores han de optar por unas u otras, hasta encontrar la más óptima según el esfuerzo que implica obtenerla y los resultados que permiten obtener.

Una vez que se han reunido las cosas o los datos, los métodos cuantitativos resumen los valores de las propiedades en una colección o conjunto de datos y las relaciones entre ellos. Dotar de sentido a una sola colección o conjunto de datos, determinando las propiedades que diferencian y clasifican sus componentes es el comienzo para alcanzar el objetivo más importante: comprender el conjunto o población más amplio del que la colección inicial es sólo una muestra. No hay mejor ejemplo de esto que la tabla periódica de los elementos desarrollada por Mendeléiev en 1869, que organizó los elementos conocidos sobre la base de sus propiedades químicas comunes y luego predijo con éxito algunas propiedades de los aún no descubiertos.

Los modelos computacionales desarrollados a partir de un conjunto de datos iniciales pueden predecir futuras observaciones. Los modelos de clasificación asignan una nueva instancia a una categoría (por ejemplo, un mensaje de spam o no spam, Madison o Hamilton como autor, si



una escena fotografiada es al aire libre o en interior); los modelos de regresión predicen un valor específico de alguna medida (dada una descripción de una nueva película, ¿cuánto dinero ganará?); los modelos de regresión ordinales predicen valores para medidas no métricas (¿cuál será su valoración de la película?). Los métodos experimentales de comprobación de hipótesis ayudan a desarrollar y perfeccionar modelos de cualquier tipo, variando sistemáticamente las condiciones en que se realizan las observaciones para descubrir cómo cambian los resultados en diferentes situaciones.

EL SENSEMAKING Y LA ORGANIZACIÓN

Las personas organizan para dar sentido a la información equívoca que reciben y reestablecer dicho sentido en el mundo con el fin de hacerlo más ordenado.

(Weick 2005)*

Los procesos de *sensemaking* (dar sentido) y organizar están entrelazados. Las culturas de la antigüedad realizaron observaciones a lo largo del tiempo y analizaron patrones sobre los ciclos de cultivo, los precios de los productos básicos, las condiciones meteorológicas y los avistamientos astronómicos. Pensemos en los primeros astrónomos, que orientaban los edificios de los templos para alinearlos con los fenómenos astronómicos y que decoraban las paredes de los templos con los signos del zodiaco.

- ✘ ¿Qué planetas y estrellas del firmamento deberían observar y cómo deberían registrar los detalles de esas observaciones?
- ✘ ¿Qué técnicas matemáticas y estadísticas deberían usar para analizar y describir estas observaciones?
- ✘ ¿Qué subconjunto de observaciones es más útil para predecir la llegada de las inundaciones del río Nilo, causadas por lluvias imposibles de ver a miles de kilómetros de distancia?

Cada decisión sobre qué observar y cómo describirlo refleja un conjunto de supuestos y posibles hipótesis que no suele hallarse formulado de forma explícita. Las decisiones que se toman con el fin de esclarecer algo parten de esta base. No obstante, estas se abandonan cuando no logran aportarnos conocimientos nuevos, pues el proceso iterativo de elegir lo que se observa y se describe no puede garantizar que lleguemos a comprender algo correctamente.

*WEICK, Karl, SUTCLIFFE, Kathleen, and OBSTFELD, David. "Organizing and the process of sensemaking": Organization science. 2005. pp. 409-42.



Los trabajos de Tycho Brahe y Johannes Kepler ejemplifican el principio según el cual un conjunto de datos precisos o extensos no basta para producir un modelo correcto. Brahe fue un noble astrónomo danés del siglo XVI que pasó décadas recogiendo datos sobre la posición de cientos de estrellas y planetas. Sin embargo, debido a los prejuicios religiosos y científicos imperantes en la época, aceptó las suposiciones erróneas de que el Sol y los planetas giraban alrededor de la Tierra en órbitas circulares. Después de su muerte en 1601, Kepler pasó una década analizando los datos de Brahe y acabó por rechazar la idea de que las órbitas planetarias eran circulares y estaban centradas en la Tierra. En su lugar, defendió que las órbitas eran elípticas y giraban en torno al Sol. Los nuevos principios en los que Kepler se basó para organizar los datos de Brahe simplificaron el modelo del sistema solar y de esta manera Kepler logró descubrir las leyes del movimiento planetario, que constituyen los cimientos de la astronomía y la física modernas.

Cuando tratamos de dar sentido y construir un modelo se nos presenta el gran desafío de encontrar un equilibrio entre dos objetivos enfrentados: comprender una colección o un conjunto de datos en particular, y ser capaces de aplicar ese conocimiento a casos nuevos. Los modelos pueden diferir en el número de descripciones de recursos que utilizan como parámetros y cabe destacar que a veces es fácil y tentador sobrecargar un modelo. Esto ocurre cuando se añaden más parámetros que registran variaciones aleatorias en las observaciones. Al sobrecargar los modelos se produce una exactitud espuria en la reproducción de las observaciones originales, pero también se consigue que sea más difícil generalizar los modelos.

El paso de mayor nivel en el *sensemaking* consiste en crear modelos científicos o teorías que propongan mecanismos con los que interpretar y establecer las causas de las observaciones. Al igual que el proceso instintivo de dar sentido a las cosas tiende a elaborar explicaciones sencillas, los científicos suelen preferir las teorías más simples frente a las complejas. Se trata de una estrategia heurística de nuestro sistema cognitivo conocida como la navaja de Ockham o el principio de parsimonia. Aunque las teorías complejas a veces pueden ser más precisas, las más sencillas generan predicciones más comprobables, lo que hace más fácil que se verifique o refine una teoría. El famoso principio de Ockham, formulado hace ocho siglos, consiste en preferir modelos que impliquen el mínimo de supuestos, por lo general medido este mínimo según la cantidad de variables o parámetros necesarios para hacer una predicción.



3.3.3 IDENTIFICAR PROPIEDADES

Una vez establecidos los propósitos de una descripción, debemos identificar las propiedades específicas de los recursos que pueden satisfacer esos propósitos. Hay cuatro razones por las que esta tarea es más difícil de lo que parece en un principio:

- ▶ En primer lugar, un recurso cualquiera puede necesitar muchas descripciones, todas ellas relacionadas con propiedades diferentes, dependiendo de las interacciones que se vayan a apoyar y del contexto en el que tengan lugar. La selección de jugadores para un equipo de béisbol se centra en sus propiedades físicas: la altura, la fuerza, la capacidad de salto y la coordinación. La selección de un equipo de debate se centrará más en sus propiedades verbales e intelectuales.
- ▶ En segundo lugar, diferentes tipos de recursos necesitan incorporar diferentes propiedades en sus descripciones. En el caso de los recursos de un museo, se podrían incluir los materiales y las dimensiones de las obras de arte; para los archivos y servicios gestionados por un administrador de red, se incluyen los permisos de control de acceso; para libros electrónicos o DVD, se incluiría el código de gestión de derechos digitales (DRM del inglés: digital rights management) que expresa lo que se permite y se prohíbe hacer con el recurso.
- ▶ En tercer lugar, como ya hemos mencionado de pasada en el apartado §3.3.1.3, las propiedades que participan en la descripción de recursos dependen de quién la realiza. No podemos esperar que alguien sin una formación adecuada en la disciplina de la organización realice distinciones e interpretaciones meticulosas de las propiedades. Retomaremos este tema en los apartados §3.3.6 y §3.4.1.
- ▶ En cuarto lugar, desde el punto de vista conceptual, puede parecernos que dos propiedades son idénticas, pero es posible que estas difieran totalmente al implementarlas. La descripción de muchos recursos no es más que un sustituto o un resumen de la fuente primaria, del recurso original. Cuando el aspecto a captar es la esencia del recurso, como ocurre con las fotografías y las pinturas, entre otros muchos recursos, una descripción resumida y apropiada podría ser una fotografía de menor resolución del recurso original (conocida como miniatura o thumbnail). Este sustituto es fácil de crear y de relacionar con el recurso primario. También



cabe destacar que sintetizar un texto en un resumen breve o un *abstract* es una habilidad en sí misma. Los recursos basados en el tiempo plantean mayores dificultades a la hora de resumir. ¿Cuál debería ser el resumen de una película: una sinopsis textual de la trama, un clip significativo de la misma, un resumen en vídeo o una mezcla de todas estas cosas? ¿Y de una canción?

El desajuste al implementar propiedades relativas a personas puede resultar muy amplio, ya que las personas no son tan fáciles de medir como la mayoría de los recursos. Las empresas necesitan cuantificar el interés de una persona en sus productos para predecir qué precio estaría dispuesta a pagar, pero el “interés” no puede medirse directamente. En cambio, las predicciones se basan en medidas indirectas del “interés”, como el tiempo que el cliente ha estado mirando la página web del producto y si también ha entrado en la página de un competidor.

Si queremos comprender y contrastar las propiedades que se usan en las descripciones de recursos y los principios de organización, debemos tener en cuenta dos dimensiones: la esencia de las propiedades (las propiedades pueden asociarse de forma intrínseca o extrínseca con la fuente) y su persistencia (pueden ser estáticas o dinámicas). En conjunto, estas dos dimensiones dan lugar a cuatro categorías de propiedades, como se ilustra en la Imagen 3.4: *La esencia de las propiedades x La persistencia: Las cuatro categorías de propiedades*. Dichas categorías proporcionan un esquema útil para pensar en las propiedades de los recursos, incluso si a veces la exacta clasificación de propiedades puede ser discutible.

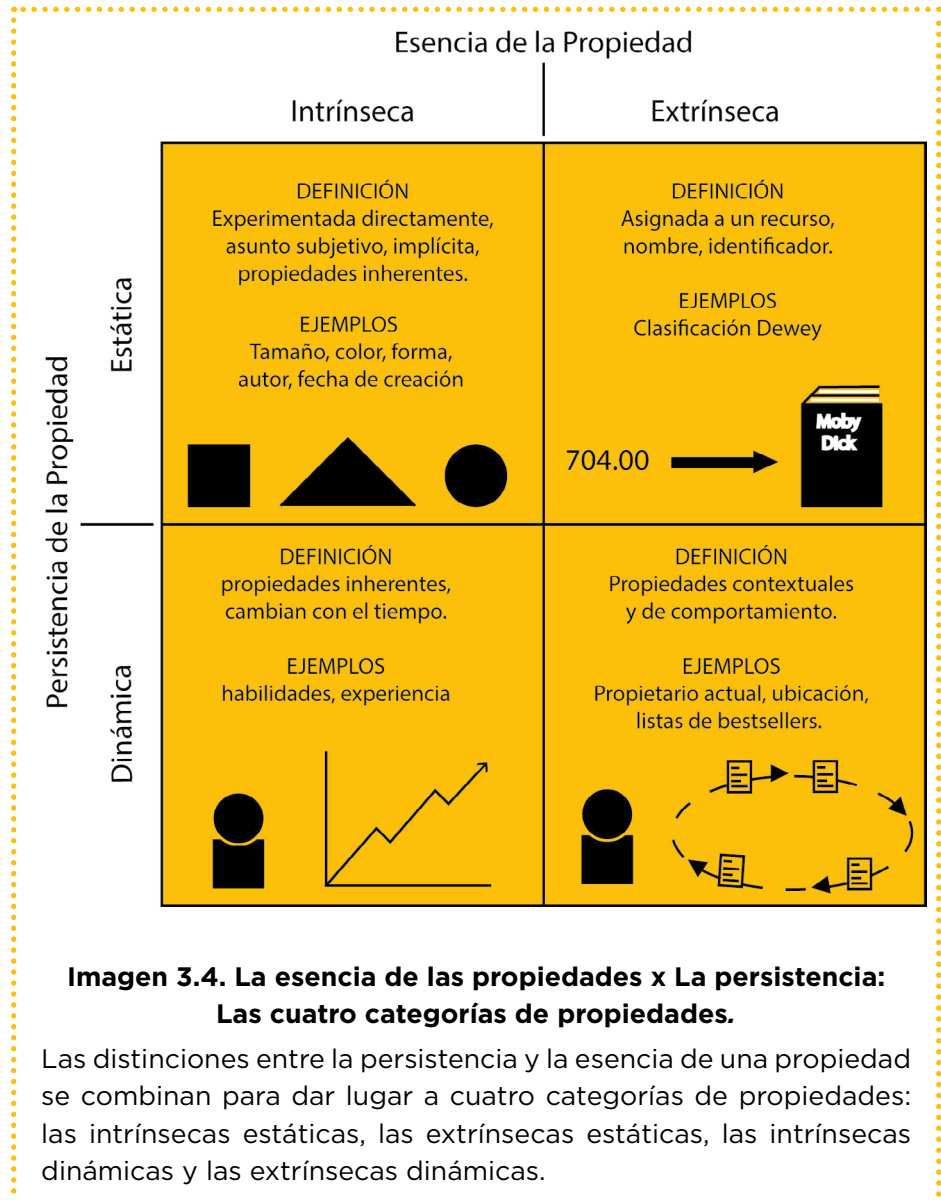
3.3.3.1. PROPIEDADES INTRÍNSECAS ESTÁTICAS

Las propiedades intrínsecas o implícitas son inherentes a los recursos y a menudo pueden percibirse o experimentarse de forma directa. Además, sus valores no cambian con el tiempo. La especie de un animal, el material de una silla de madera y el diámetro de una rueda son propiedades estáticas, ya que sus valores son invariables. En consecuencia, las propiedades estáticas como el color o la forma son utilizadas regularmente para describir y organizar los recursos físicos.

Las propiedades físicas e intrínsecas suelen ser sólo una parte de la descripción de recursos. En muchos casos, las propiedades materiales o físicas describen sólo la capa superficial de un recurso y revelan más bien poco sobre su verdadera naturaleza o su propósito original, lo que significa o cuándo y por qué fue creado. El autor de una canción y el contexto de su creación son ejemplos de propiedades intrínsecas y estáticas que no son perceptibles de forma directa.



Algunas descripciones intrínsecas pueden también extraerse o calcularse mediante procesos computacionales. Por ejemplo, un programa informático podría calcular la frecuencia de aparición y la distribución de las palabras en un documento específico. Del mismo modo, las firmas visuales (visual signatures) o las huellas digitales acústicas (audio fingerprints) son descripciones intrínsecas (Aspectos a los que se dedica, un poco más adelante, el apartado §3.4 *La descripción de recursos no textuales*).



Algunas relaciones entre recursos son intrínsecas y estáticas, como la relación entre un padre y su hijo o entre dos hermanos. Las relaciones parciales y de composición presentes en recursos compuestos por partes también son propiedades intrínsecas estáticas y, además, se em-



plean mucho en las descripciones de recursos. **Sin embargo, es mejor evitar tratar las relaciones entre recursos como propiedades y, en su lugar, expresarlas como relaciones.**

3.3.3.2. LAS PROPIEDADES EXTRÍNSECAS ESTÁTICAS



Las propiedades intrínsecas estáticas de un dálmata

Las manchas de un perro dálmata son propiedades intrínsecas estáticas que aparecen poco después del nacimiento y son tan distintivas que es imposible describir la raza sin mencionar las manchas.

Las propiedades extrínsecas o explícitas se asignan a un recurso en lugar de ser inherentes a él. El nombre o el identificador de un recurso es a menudo arbitrario, pero una vez asignado no suele cambiar. La disposición de los recursos según el orden alfabético o numérico de sus identificadores descriptivos es un principio de organización común. Los números de clasificación y los encabezamientos de materia o temas asignados a recursos bibliográficos son propiedades estáticas extrínsecas, al igual que los números de serie estampados o adjuntos a productos manufacturados.

En el caso de los recursos de información digitales, puede que las propiedades de sus versiones impresas o “renderizadas” (una presentación digital concreta) no sean intrínsecas. Algunos formatos de texto separan por completo el contenido de la presentación y, como resultado, las hojas de estilo pueden cambiar de forma radical el aspecto de un documento impreso o de una página web sin alterar el recurso



primario. Por ejemplo, si se aplicara un estilo diferente a este párrafo para resaltarlo en negrita o en letra de 24 puntos, su contenido seguiría siendo el mismo.

3.3.3.3 LAS PROPIEDADES INTRÍNSECAS DINÁMICAS

Las propiedades intrínsecas dinámicas cambian con el tiempo. Un ejemplo de estas son las características que cambian con el desarrollo de un individuo como la altura y el peso, el dominio de una destreza o la capacidad intelectual. Como estas propiedades no son estáticas, suelen emplearse solo para organizar recursos que permanecerán en la colección durante un tiempo limitado. Los programas deportivos o las ligas que separan a los participantes por edad o años de experiencia se sirven de propiedades intrínsecas dinámicas para describir y organizar sus recursos.

3.3.3.4 LAS PROPIEDADES EXTRÍNSECAS DINÁMICAS

Las propiedades extrínsecas dinámicas son arbitrarias en muchos aspectos y pueden cambiar porque se basan en el uso, el comportamiento o el contexto. El propietario o la ubicación actual de un recurso, su frecuencia de acceso, la frecuencia de conexión con otros recursos, su popularidad actual o relevancia cultural, o su ventaja competitiva sobre recursos alternativos son algunas de las propiedades extrínsecas y dinámicas que más se utilizan en las descripciones de recursos. Un libro de actualidad descrito como un *best seller* de un año concreto puede encontrarse en la sección de rebajas unos años más tarde. El promedio de las notas de un estudiante es una propiedad extrínseca y dinámica.

Las propiedades extrínsecas dinámicas son útiles para los científicos de datos que realizan modelos de predicción o clasificación. La ubicación actual de un cliente, lo que ha comprado hace unos segundos y el lugar donde se realizó una compra pueden considerarse manifestaciones de preferencias y valores no observables. En la escena de un crimen, las huellas dactilares que se encuentran en el pomo de una puerta son una propiedad extrínseca dinámica asociada a la puerta. Estas serán analizadas por detectives forenses junto con otros recursos de interacción, todo ello con el objetivo de identificar a la persona para la que las huellas dactilares son propiedades intrínsecas estáticas. Muchas relaciones entre recursos son propiedades extrínsecas y dinámicas, como la relación que podríamos denominar “mejores amigos”.

Las *propiedades contextuales* son las relacionadas con la situación o el contexto en el que se describe un recurso. El investigador Anind K. Dey, del Georgia Tech, define “contexto” como “cualquier información que



caracterice una situación relacionada con las interacciones entre los usuarios, las aplicaciones y el entorno que la rodee”. Esta definición abierta implica un gran número de propiedades contextuales que podrían utilizarse en una descripción; algunas definiciones más específicas de “contexto” podrían ser “ubicación + actividad” o “quién, cuándo, dónde, por qué”. Como el contexto varía, los descriptores basados en el contexto son apropiados en el momento en el que se asignan, pero su persistencia y su eficacia son limitadas; la descripción de un documento como “recibo de una compra reciente” no será útil durante mucho tiempo⁸.

Los recursos suelen describirse con “propiedades culturales” derivadas del lenguaje o la cultura convencionales, a menudo por analogía, ya que pueden ser muy evocadores y memorables.

A veces una descripción cultural pierde su notoriedad y, con ello, su poder para evocar cualquier otra cosa que no sea perplejidad cuando el usuario desconoce su significado, porque habita en “otro tiempo”, “otro mundo” u “otra cultura”. Para los amantes del Lego del ejemplo usado al inicio de este capítulo, siempre al corriente de las últimas películas de Star Wars, la “espada láser” era sólo la descripción obvia de un largo tubo de neón con mango. Sin embargo, alguien que no esté familiarizado con Star Wars podría no comprender “espada láser”, por lo que es posible que describa la pieza de otra manera.

3.3.4. EL DISEÑO DEL VOCABULARIO DE DESCRIPCIÓN

Una vez que hemos establecido las propiedades que utilizaremos en las descripciones de recursos, es necesario diseñar el vocabulario de descripción: un conjunto de palabras o valores que representen las propiedades. Los principios de la denominación, establecimiento de títulos o asignación de nombres, de los que ya hemos hablado anteriormente, son en parte aplicables para el diseño de un vocabulario de descripción⁹.

Sin embargo, como un vocabulario de descripción conlleva mucho más que la mera denominación de un recurso, debemos plantear algunos principios o directrices adicionales para esta etapa. Además, surgen aquí

8 Puede leerse sobre esta situación en el apartado §2.5 “Los recursos a lo largo del tiempo” de este libro.

9 El lector puede encontrar información más detallada en el apartado §2.4 “Nombrar recursos”, en este mismo libro.



nuevas preguntas relacionadas con el diseño, si consideramos todas las descripciones de recursos como un conjunto de descripciones separadas creadas por varios individuos a lo largo del tiempo.

LAS CARACTERÍSTICAS LATENTES Y LAS RECOMENDACIONES DE NETFLIX

Los últimos avances en computación y en ciencia de datos están permitiendo descubrir o crear propiedades de recursos denominadas “latentes”, ya que se infieren en lugar de observarse.

Algunas empresas usan muchas de estas características para clasificar a los clientes o hacerles recomendaciones basadas en su comportamiento reciente, por lo que estas características son también extrínsecas y dinámicas. No es tan sencillo deducir el sexo y la edad de un individuo en base a sus preferencias cinematográficas, pues no diferencian lo suficiente a los espectadores como para hacer recomendaciones personalizadas, ni siquiera si se estableciesen categorías amplias como “estudiantes varones-solteros de entre 18 y 25 años”.

Netflix descubrió que era necesario combinar las propiedades demográficas, el historial de visitas y el comportamiento de navegación con calificaciones detalladas de docenas de películas para lograrlo. Pensemos en los usuarios de Netflix a los que les gustan las películas de acción, pero que no dejan de pasar el ratón por las películas románticas. Se necesita un equipo computacional muy potente para interpretar esta característica latente y recomendar a estos usuarios un subgénero (de casi 100.000 espectadores) llamado “Películas románticas de acción y aventura”.

3.3.4.1 LOS PRINCIPIOS DE UNA BUENA DESCRIPCIÓN

En su obra *The Intellectual Foundation of Information Organization*, Svenonius propone un conjunto de principios o “directrices de diseño” de un lenguaje de descripción. Sus principios, enmarcados en el contexto específico de la descripción bibliográfica, se pueden generalizar para toda la amplia gama de tipos de recursos que nos interesan en este libro.

USO ADECUADO

Al elegir los términos de descripción, tenga en cuenta al usuario; es recomendable que dichos términos sean de uso común entre el público al que se dirigen.



REPRESENTACIÓN

Use descripciones que reflejen la forma en que los recursos se describen a sí mismos; suponga que las autodescripciones son correctas.

SUFICIENCIA Y NECESIDAD

Las descripciones deben proporcionar exclusivamente la información necesaria para cumplir sus propósitos. Esto podría implicar la exclusión de algunos aspectos de las autodescripciones que no son relevantes.

NORMALIZACIÓN

Normalice las descripciones para que sean prácticas, pero utilice también variantes y nombres alternativos (alias) para términos de uso común.

INTEGRACIÓN

Escoja preferentemente las mismas propiedades y términos para todos los recursos.

Cualquier conjunto de principios generales de diseño se enfrenta a dos retos.

- ▼ El primero es que, al aplicar cualquier principio, se requieren muchas elecciones adicionales y específicas dependientes del contexto para las que el principio general ofrece muy poca orientación. Por ejemplo, ¿cómo se aplica el principio de normalización si ya existen múltiples normas en un determinado dominio de recursos? ¿Cuál de las normas debería adoptarse y por qué?
- ▼ El segundo problema es que los principios generales pueden a veces dar lugar a recomendaciones contradictorias. Volvamos al principio del uso adecuado, que consiste en elegir términos de descripción de uso común. Pues bien, este fracasa si la comunidad de usuarios incluye tanto personas corrientes como científicos, colectivos que utilizan diferentes términos para los mismos recursos; ¿cuál es el “uso común” que debe prevalecer?

3.3.4.2 ¿QUIÉN UTILIZA LAS DESCRIPCIONES?

Centrémonos en el usuario de las descripciones. En cada paso del proceso de descripción se debe tener en cuenta al usuario, por lo que se trata de un elemento fundamental. Todos los principios de diseño de la sección anterior comparten la idea de que el diseño del vocabulario



de descripción debe centrarse en el usuario de las descripciones. ¿Se están organizando recursos personales, para fines personales y sobre todo privados? En ese caso, las propiedades y los términos de descripción pueden ser muy personales o idiosincrásicos y, no obstante, seguir los principios de diseño.

Del mismo modo, cuando los usuarios de los recursos comparten conocimientos relevantes o se encuentran en un contexto en el que pueden comunicarse y negociar, en caso de ser necesario, con el fin de identificar los recursos, sus descripciones pueden permitirse ser menos precisas y rigurosas de lo que deberían. Esto ayuda a explicar las curiosas descripciones en la historia de Lego con la que comenzamos este capítulo. Los niños que jugaban con las piezas estaban hablando teniendo todos delante las piezas Lego. Si no hubieran podido ver las piezas de las que hablaban o si hubieran tenido que describírselas a alguien que no hubiera jugado nunca antes con Lego, sus descripciones habrían sido necesariamente muy diferentes.

Sin embargo, es más habitual que en las descripciones de recursos no exista ese grado de contexto compartido. Además, no se diseñan solo para unos individuos concretos, sino para categorías de usuarios: empleados o clientes de empresas que utilizan catálogos de piezas y productos, científicos que analizan los conjuntos de datos de experimentos o simulaciones o usuarios de bases de datos que buscan libros o artículos científicos. En cada una de estas situaciones las descripciones de recursos tendrán que ser comprensibles para personas que las tienen que usar, pero no las han creado, por lo que el diseño del vocabulario de descripción debe ser más deliberado y sistemático. Así se garantiza que los términos no son ambiguos y que son convenientes para permitir una interpretación fiable y no dependiente del contexto. No es habitual que un único individuo posea la gran cantidad de conocimientos y la experiencia que se necesitan para idear un vocabulario de descripción que pueda satisfacer a diversos usuarios con propósitos diferentes. De hecho, suelen ser muchas personas las que trabajan juntas cuando se necesita desarrollar un vocabulario de descripción. A los resultados se les conoce como “vocabularios institucionales” para diferenciarlos de los individuales o culturales¹⁰.

¹⁰ Este tema se trata más a fondo en el capítulo 7, “Categorización: Describir clases y tipos de recursos” en la edición completa online del libro “TDO core concepts edition”.



PÁRATE A PENSAR: LA DESCRIPCIÓN Y LA EXPERIENCIA

Todo el mundo sabe algo sobre árboles, pero algunos saben más que otros y su experiencia particular y su perspectiva influyen en la forma en que los describen. ¿Qué tipo de propiedades y descripciones utilizarían los estudiantes universitarios, los investigadores en botánica, los diseñadores de paisajes, los trabajadores de mantenimiento de parques o los pueblos indígenas que viven en la selva tropical?

Algunas descripciones de recursos están diseñadas para ser utilizadas por máquinas (por aplicaciones informáticas, podríamos decir). En principio, esto reduce la importancia de los principios de diseño que tienen en cuenta las preferencias del usuario o los usos comunes. Sin embargo, aunque los recursos se describan y organicen mediante algoritmos, cuando se necesita explicar las clasificaciones y las predicciones que estos producen, se prefieren las descripciones de recursos comprensibles y fáciles de comunicar frente a otras que podrían ser las más adecuadas desde el punto de vista estadístico. Además, los principios de normalización e integración cobran mayor importancia en la comunicación entre máquinas, para permitir un procesamiento eficiente, la reutilización de datos y software, así como una mayor interoperabilidad entre los Sistemas de Organización.

METAESTUPIDEZ

En un ensayo muy citado con un título bastante provocador: *Metaestupidez: Destapando siete subterfugios de la metautopía* (2001). Cory Doctorow mantiene que muchos de los metadatos creados por los humanos son de baja calidad porque “la gente inteligente, es perezosa y estúpida. En definitiva, una misión imposible. Conózcase a sí mismo, los esquemas no son neutrales, las métricas influyen en los resultados, (y) hay más de una manera de describir las cosas”.

3.3.4.3 LOS VOCABULARIOS CONTROLADOS Y LAS NORMAS DE CONTENIDO

Un vocabulario controlado es un conjunto fijo o cerrado de términos de descripción con definiciones precisas sobre un dominio determinado y que se utiliza en lugar de los términos de “uso común”. Por ejemplo, en vez de usar términos de uso popular para referirse a enfermedades o síntomas, los investigadores del campo de la medicina y los hospitales universitarios utilizan el vocabulario controlado de la National Library of Medicine, denominados “National Library of Medicine’s Medical Subject Headings” o MeSH.



Podemos distinguir una progresión en el control del vocabulario: primero se crea un glosario de términos permitidos; luego se elabora un tesoro de términos dispuestos en una jerarquía y anotados para indicar los términos que se prefieren, unos más amplios o específicos que otros; más complejo será elaborar una ontología que exprese las relaciones conceptuales entre los términos en un lenguaje formal y lógico para que puedan ser procesadas por ordenadores.

Las reglas de contenido o reglas de catalogación (*content standards*) se asemejan a los vocabularios controlados porque también limitan los valores que pueden utilizarse en las descripciones. En lugar de especificar un conjunto fijo de valores, las reglas de contenido suelen restringir las descripciones al exigir que sean de un tipo concreto de datos (entero, booleano, etc.). Los valores posibles están limitados por expresiones lógicas (por ejemplo, un valor debe estar entre 0 y 99) o por expresiones regulares (por ejemplo, debe ser una cadena de longitud 5 que debe comenzar con un número). Las reglas de contenido como éstas se utilizan para garantizar descripciones válidas cuando se introducen en formularios web u otras aplicaciones.

3.3.4.4 EL CONTROL DEL VOCABULARIO COMO REDUCCIÓN DE LA DIMENSIONALIDAD

En la mayoría de los casos, un vocabulario controlado es un subconjunto del vocabulario natural o no controlado, pero también puede ser un conjunto nuevo de términos inventados con un propósito funcional. Esto puede sonar extraño si no tenemos en cuenta que el objetivo del vocabulario controlado es reducir el número de términos descriptivos susceptibles de asignarse a un recurso. Dicho de esta manera, lo que nos preocupa es la “reducción de la dimensionalidad”: buscamos transformar un espacio de gran dimensión en uno más reducido. La disminución del número de componentes en una descripción multidimensional puede lograrse mediante muchas técnicas estadísticas diferentes como la “extracción de características”, el “análisis de componentes principales”, la “descomposición ortogonal”, el “análisis de semántica latente”, el “escalamiento multidimensional” y el “análisis factorial”.

Estas técnicas pueden sonar imponentes y complejas, pero todas comparten el mismo concepto subyacente: las características o las propiedades que describen un recurso a menudo están correlacionadas. Por ejemplo, es más probable encontrar los términos “conductor” en un documento que contenga la palabra “coche” que en otro que no la tenga. Existen correlaciones similares entre las características visuales



que se utilizan para describir imágenes y las características acústicas que describen la música. Las técnicas de reducción de la dimensionalidad analizan las correlaciones entre las descripciones de recursos para transformar un gran conjunto de descripciones en un conjunto mucho más pequeño de otras no correlacionadas. De alguna manera, con esto se aplica el principio de la suficiencia y la necesidad que mencionamos anteriormente en §3.3.4.1 *Los principios de una buena descripción*, porque elimina las dimensiones o las propiedades de descripción que no son estrictamente necesarias para distinguir los recursos.

A continuación, ofrecemos un ejemplo simplificado que ilustra esta idea. Supongamos que tenemos una colección de recursos, en la que cada recurso descrito como “grande” también se describe como “rojo” y que cada recurso “pequeño” es también “verde”. Esta perfecta correlación entre color y tamaño significa que cualquiera de estas propiedades es suficiente para distinguir las cosas “grandes rojas” de las “pequeñas verdes” y no necesitamos ningún algoritmo para averiguarlo. No obstante, si tenemos miles de propiedades y las correlaciones son solo parciales, necesitamos de sofisticados enfoques estadísticos para elegir el conjunto óptimo de propiedades y términos de descripción. En algunas técnicas las dimensiones que quedan se llaman “latentes” o “sintéticas” porque son óptimas desde el punto de vista estadístico, pero no se corresponden directamente con las propiedades de los recursos.

3.3.5 DISEÑO DE LA FORMA DE LA DESCRIPCIÓN

Al llegar a esta etapa de la descripción de recursos, ya hemos tenido que tomar muchas decisiones importantes: qué recursos describir, los propósitos para los que estamos describiendo dichos recursos, sus propiedades y los términos que usaremos en las descripciones. En la medida de lo posible, hemos descrito los pasos a nivel conceptual. No obstante, hemos estado posponiendo el debate sobre su implementación y eludiendo consideraciones sobre la notación, la sintaxis y el despliegue de las descripciones de los recursos, como unidades separadas o por paquetes. Separar el diseño de las consideraciones de implementación supone idealizar el proceso de descripción de recursos, lo cual hace más fácil aprender y pensar en la descripción de recursos y en los Sistemas de Organización.

A veces tendremos que enfrentarnos a tecnologías heredadas que llevan en funcionamiento durante años dentro de una organización, o a ciertos acuerdos comerciales establecidos o previstos, a las regu-



laciones legales, o al cumplimiento de ciertas normas, o a requisitos de rendimiento u otros factores que influyen en la forma en que las descripciones de recursos deben o deberían ser implementadas, almacenadas y gestionadas. Esta perspectiva más pragmática la denominamos “hoja de ruta de un Sistema de Organización”, y merece ser tratada con más detalle, pero no en este momento ni en este documento.



3.3.6 LA CREACIÓN DE DESCRIPCIONES DE RECURSOS

Las descripciones de recursos pueden ser creadas por profesionales, por los autores o editores de los recursos, los usuarios que acceden y usan los contenidos o también por medios computacionales o automatizados. Cada uno de estos casos presenta unas ventajas y unas limitaciones particulares y conlleva perder algo en un sentido para ganar en otro. Una solución habitual es tratar de combinar en enfoques híbridos los aspectos deseables de cada uno de ellos. Por ejemplo, el vocabulario de un nuevo dominio de recursos puede tomarse del etiquetado que hacen los usuarios finales y luego ser perfeccionado por profesionales; clasificadores poco expertos pueden crear descripciones ayudados por herramientas de software que sugieren términos. También se puede usar un software que cree descripciones y mejorarlo después con descripciones hechas por humanos: una forma de “aprendizaje supervisado”.

A menudo, las descripciones de recursos existentes pueden o deben ser transformadas o mejoradas para satisfacer las necesidades cambi-



antes de un Sistema de Organización y a veces estos procesos de re-conversión pueden automatizarse. Esto puede enfocarse desde el punto de vista de las interacciones entre recursos, pero no lo trataremos en este documento. El siguiente punto se centra en la creación de nuevas descripciones de recursos desde cero, donde aún no existen.

3.3.6.1 LAS DESCRIPCIONES HECHAS POR PROFESIONALES

Antes de que Internet permitiera que casi cualquier persona creara, publicara y describiera sus propios recursos, y que describiera los creados y publicados por otros, la descripción de recursos solían realizarla profesionales en contextos institucionales. Los indizadores y los catalogadores expertos describían recursos bibliográficos y museísticos después de haber sido formados para aprender los conceptos, los vocabularios controlados para la descripción y las normas profesionales relevantes. En el campo de los sistemas de información, los analistas profesionales de datos y procesos, y los redactores técnicos, entre otros, han creado descripciones igual de rigurosas y recibido una formación especializada análoga. A estos tipos de descripciones de recursos las hemos denominado “institucionales” para resaltar que han sido creadas de acuerdo con unas normas y que no son resultado de decisiones informales ad hoc, sobre todo por personas no formadas o inexpertas.

3.3.6.2 LAS DESCRIPCIONES HECHAS POR AUTORES O CREADORES

Se supone que el autor o el creador de un recurso comprende las razones y los fines para los que se puede utilizar. Y, supuestamente, la mayoría de los autores desean que se les lea, por lo que describirán sus recursos de forma que los usuarios los encuentren útiles y atractivos. Sin embargo, es poco probable que esas descripciones utilicen los vocabularios controlados y las normas que utilizarían los catalogadores profesionales.

3.3.6.3 LAS DESCRIPCIONES HECHAS POR USUARIOS

Hoy en día Internet alberga un número asombroso de recursos, de los cuales la mayoría son fuentes de información primaria publicadas



como contenido web. No obstante, muchísimos otros recursos circulan como medios físicos existentes en el “mundo real”. La mayoría de los recursos están siendo descritos por usuarios y no por profesionales o autores. La tendencia de conjunto para estos usuarios es que creen descripciones para su propio interés cuando asignan etiquetas o clasificaciones a los recursos de la web y rara vez se sirven de descriptores estándar o controlados. La variabilidad resultante puede suponer un problema si la tarea de crear una descripción en realidad requiere unos conocimientos específicos. Pongamos un ejemplo: casi nadie pondrá en duda la duración de un determinado archivo de música, pero, si hemos de determinar a qué género musical pertenece, lo más probable es que haya discrepancias. Afortunadamente, la mayoría de los usuarios de la web reconocen de forma implícita que el potencial de las aplicaciones “Web 2.0” o de “contenido generado por el usuario” sería aún mayor si se evitaran las descripciones egocéntricas. Además, las estadísticas elaboradas a partir de grandes cantidades de datos prueban que, en las aplicaciones más masivas, se tiende a llegar a un acuerdo sobre las descripciones. Esto se debe a que las descripciones idiosincrásicas son superadas, en lo que se refiere a la frecuencia de distribución, por las más convencionales.

Con esto no estamos diciendo que las descripciones profesionales sean siempre de alta calidad y utilidad, ni que las del resto de la sociedad sean peores. Más bien, buscamos que se comprendan las limitaciones y las cualidades de cada una de ellas. El etiquetado reduce las barreras de entrada para la descripción, lo que consigue que la organización de la información sea más accesible y que las descripciones reflejen diversos puntos de vista. Sin embargo, cuando se asocian muchas etiquetas con un recurso, aunque es más probable que se pueda recuperar, disminuye la precisión (Recall vs. Precisión; Exhaustividad vs. Precisión).





3.3.6.4 LA DESCRIPCIÓN DE RECURSOS AUTOMATIZADOS Y COMPUTACIONALES

El archivo EXIF de una imagen tomada con una cámara digital registra las propiedades asociadas a la cámara y sus ajustes, así como algunas propiedades del contexto en el que se tomó la fotografía. Hacer a mano una descripción tan detallada sería casi imposible. Sin embargo, el inconveniente de la descripción automatizada es que no capta el significado de la foto; es decir, captura el tiempo y el lugar, pero no nos dice si la fotografía fue tomada durante una luna de miel. Denominamos a este fenómeno “brecha semántica”: a la diferencia entre la descripción automatizada y la humana¹¹.

Cualquier recurso lo suficientemente inteligente como para recoger datos sobre su estado o su entorno estará creando descripciones de recursos de forma automática¹². Los recursos que incorporan capacidades computacionales digitales pueden procesar los datos que llegan directos desde sensores para identificar eventos importantes y crear descripciones más fáciles de interpretar.

Algunos abordajes computacionales crean descripciones de recursos que tienen un propósito similar a las descripciones creadas por catalogadores humanos. La minería de textos y los sistemas de resumen aplicados a los comentarios de clientes sobre productos pueden sintetizar miles de comentarios en una lista que contenga las características más relevantes. Cuando buscamos libros en Amazon obtenemos información sobre su contenido y su singularidad, pues Amazon puede comparar el texto completo de montañas de libros y detecta las frases menos corrientes que diferencian a unos de otros.

Las descripciones computacionales pueden usar cualquier variable observable o latente (ver el recuadro lateral: *Las características latentes y las recomendaciones de Netflix*), excepto algunas ilegales como la raza, la religión, el origen nacional y el estado civil, para evitar la discriminación. En la práctica, sin embargo, esta prohibición se elude fácilmente porque estas propiedades se suelen deducir a través de otras. Por ejemplo, la dirección de residencia y el apellido suelen aportar una indicación bastante fiable de la raza de un individuo.

Por supuesto, todos los sistemas de recuperación de información comparan la descripción de las necesidades de un usuario con las

¹¹ Fenómeno que se aborda en el apartado §2.4.2.5 de este mismo libro.

¹² Aspectos derivados de las interacciones entre recursos, tratados en el apartado §2.2.3.2 de este libro.



descripciones de los recursos que podrían satisfacerlas. Los sistemas de búsqueda y recuperación de información se diferencian en las propiedades de los recursos que destacan, el número de apariciones y la distribución de una palabra para los documentos de bibliotecas digitales, los enlaces y el comportamiento de navegación para las páginas web, la acústica para la música, etc. Estas propiedades de descripción determinan el funcionamiento de los algoritmos de comparación y la forma en que se determina la relevancia o la similitud de las descripciones.

3.3.7 EVALUAR LAS DESCRIPCIONES DE RECURSOS

El proceso de evaluación está implícito en muchas de las actividades de los Sistemas de Organización¹³. Y debe ser explícito o planificado cuando se mantiene una colección de recursos a lo largo del tiempo. En este apartado, nos centraremos únicamente en la evaluación de las descripciones de recursos.

Evaluar significa determinar la “calidad” con respecto a unos criterios o unas dimensiones. Se han propuesto muchos conjuntos diferentes de criterios; para los repositorios de recursos digitales los más utilizados son la exactitud, la integridad y la consistencia. Otros muy comunes son la puntualidad, la interoperabilidad y la facilidad de uso. No obstante, es fácil que estos criterios entren en conflicto entre sí: si nos esforzamos mucho por cumplir con los criterios de exactitud e integridad, podríamos estar descuidando la puntualidad; mejorar la coherencia podría impedir modificaciones y personalizaciones que mejorarían la usabilidad.

En este capítulo hemos planteado un proceso que consta de múltiples etapas y cuya calidad resultante dependerá de la calidad creada o perdida en cada paso. Un ámbito demasiado específico o abstracto, propósitos demasiado ambiciosos o poco definidos, un vocabulario de descripción difícil de usar o un plazo insuficiente para crear buenas descripciones pueden causar problemas de calidad, pero ninguna de estas decisiones es visible al final del proceso, es decir, cuando los usuarios interactúan con las descripciones de los recursos.



¹³ Estas actividades se detallan en el capítulo 3 “Las actividades de los sistemas de organización”, disponible en la edición online íntegra del libro “TDO core concepts edition”.



PÁRATE A PENSAR: DEFINIR LA CALIDAD

¿Qué características o criterios utilizaría para determinar la calidad de un coche? ¿De la comida? ¿De la ropa? ¿De un lugar para vivir? ¿Cuáles de estos criterios son específicos de un dominio y cuáles se aplican de manera más general a muchos tipos de recursos?

3.3.7.1 EVALUAR LA CREACIÓN DE DESCRIPCIONES DE RECURSOS

Cuando los profesionales crean descripciones de recursos de manera centralizada, (esta es la práctica que más se aplica a los recursos de las bibliotecas) se centran en la calidad de las mismas para garantizar que se han utilizado vocabularios controlados y normas apropiadas. Sin embargo, la necesidad de describir recursos se extiende más allá del dominio bibliográfico tradicional y en esos contextos surgen otras consideraciones de calidad.

Las descripciones de recursos son esenciales en las empresas privadas para dirigir el negocio e interactuar de forma eficiente con proveedores, socios y clientes. En cambio, el sector público hace mucho más hincapié en los costes y la estrategia de descripción de recursos. ¿Cuál es el valor de la descripción de recursos? ¿Quién asumirá los costes de su producción? ¿Qué normas sectoriales se seguirán? Algunas de estas decisiones no se toman a la ligera, sino que se imponen como condición para hacer negocios con un socio comercial dominante, que a veces puede ser una entidad gubernamental.

Por ejemplo, una empresa como Wal-Mart, con un enorme poder de mercado, puede dictar términos y normas a sus proveedores porque los beneficios a largo plazo del contrato hacen que la adaptación inicial valga la pena. Asimismo, los gobiernos suelen exigir a sus proveedores que los servicios electrónicos proporcionados se basen en estándares abiertos, para no depender de tecnologías patentadas.

Tanto en el sector público como en el privado se utilizan cada vez más técnicas informáticas para crear descripciones de recursos. Esto se debe a que hay demasiados recursos como para elaborar descripciones profesionales una a una manualmente. Por esto, se están desarrollando mucho los campos de la minería de textos, la clasificación de páginas web, el enriquecimiento semántico y otras áreas de investigación y desarrollo similares. De esta manera se está reduciendo el coste de



producir descripciones útiles de recursos. De hecho, algunos museos han optado por crear de forma automática descripciones de recursos orientadas al usuario, así como nuevas interfaces para la búsqueda y la navegación, algo que han conseguido por medio de sistemas internos de gestión de colecciones, con los que han transformado las descripciones profesionales. Google también se embarcó en un proyecto ambicioso: digitalizar millones de libros, aunque ha sido muy criticado por la calidad de sus descripciones de recursos, elaboradas por algoritmos. A pesar de todo, es probable que los informáticos encuentren buenos usos usando su corpus de libros como banco de pruebas de investigación para mejorar las técnicas de descripción.



3.3.7.2 EVALUAR EL USO DE LAS DESCRIPCIONES DE RECURSOS

Independientemente de (o además de) los criterios de calidad aplicados a la creación y la selección de descripciones de recursos, las descripciones de recursos en algún momento terminan en manos de sus usuarios previstos. En ese instante el criterio de calidad más importante es la capacidad de las descripciones para satisfacer los fines previstos. En muchos sentidos, la respuesta es decepcionante: no lo hacen.

El diseño de un vocabulario de descripción se circunscribe a lo que se puede decir sobre un recurso, pero es importante reconocer que también determina de forma implícita lo que no se puede decir, con consecuencias negativas para los usuarios. El esquema de descripción de recursos puesto en práctica en el sistema de gestión de pacientes de un médico define ciertos tipos de información registrable sobre la visita de un paciente: la fecha de la visita, cualquier prueba o diagnósti-



co que se haya hecho, una remisión a un especialista, etc. El esquema, así como todo el trabajo que lleva asociado, impone limitaciones que afectan al tipo de información que los profesionales médicos pueden registrar y a la cantidad de espacio que pueden utilizar para esas descripciones. Además, ese esquema podría eliminar un espacio fundamental que no está estructurado y que pueden proporcionar los registros en papel: aquel en el que los médicos comunican sus fundamentos para un diagnóstico o explican una decisión sin incluir dicha información en ninguna casilla.

3.3.7.3 LA IMPORTANCIA DE LA EVALUACIÓN ITERATIVA

Entre los objetivos de calidad surgen conflictos inevitables que suponen que tengan que hacerse ciertas concesiones en cuanto a los criterios de calidad. Además, el aumento de escala en un sistema de organización y las mejoras constantes de las técnicas computacionales de descripción de recursos implican que la naturaleza de los equilibrios cambiará con el tiempo. En consecuencia, no bastará con evaluar las descripciones de los recursos en un único momento dado.

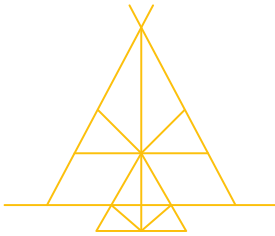
Esto hace que algunos tipos de descripciones de recursos como los registros de uso (logs), el historial de navegación y los datos transaccionales cobren gran importancia, dado que permiten centrarse en mejorar la calidad donde sea necesario. Además, en el caso de los Sistemas de Organización con muchos tipos de recursos y comunidades de usuarios, esta información puede permitir adaptar la naturaleza y el alcance de la descripción de recursos para encontrar el equilibrio adecuado entre “rico y exhaustivo” (“rich and comprehensive”) y “simple y eficiente” (“simple and efficient”). Cada combinación de un tipo de recurso y una comunidad de usuarios podría tener una solución diferente.

La noción de que la calidad es una propiedad resultante de todas las fases de un proceso completo está en el centro del “movimiento por la calidad” y en el control estadístico de procesos para las actividades industriales, pero también puede aplicarse a la descripción de recursos. La idea principal es que la calidad no puede comprobarse mediante la inspección de los productos finales. Al contrario, se logra por medio del control de procesos, la medición y la eliminación de la variabilidad de cada paso necesario para crear los productos. La retroalimentación explícita de los usuarios o la implícita obtenida al registrar sus necesidades de interacción con los recursos son esenciales a medida que iteramos a través del proceso de diseño y volvemos a examinar las decisiones tomadas.



3.4 LA DESCRIPCIÓN DE RECURSOS

NO TEXTUALES



Muchos de los principios y los métodos de descripción más conocidos se desarrollaron para el tratamiento de recursos textuales en formatos físicos. Esos principios han tenido que evolucionar para manejar los diferentes tipos de recursos que es necesario describir y organizar: desde pinturas y estatuas hasta archivos MP3, JPEG y MPEG.

Algunas descripciones de recursos no textuales se basan en textos y, en la mayoría de los casos, están elaboradas por personas. Otras descripciones en formatos no textuales se extraen del contenido de un recurso no textual por medio de algoritmos. Estas últimas, basadas en el contenido, captan propiedades técnicas intrínsecas y en algunos dominios son capaces de describir el tema del que trata el recurso con gran exactitud y todo gracias a los grandes avances del aprendizaje automático.

3.4.1 LA DESCRIPCIÓN DE RECURSOS ARTÍSTICOS Y MUSEÍSTICOS

Los problemas asociados con la descripción de los recursos multimedia no son del todo nuevos. Los conservadores de los museos han estado lidiando con ellos desde que empezaron a recopilar, almacenar y describir artefactos hace cientos de años. Muchos de estos objetos pueden manifestar una misma técnica de trabajo (pensemos en varios fragmentos de cerámica que pertenecieron al mismo jarrón). Los materiales y las formas no transmiten la semántica por sí mismos: sin una investigación y una descripción adicional, no sabemos nada sobre el jarrón, pues este no viene con ningún tipo de rótulo o etiqueta que lo relacione con un asentamiento Maya del siglo IX, por ejemplo. Dado que los museos pueden adquirir grandes lotes de artefactos de una sola vez, deben decidir qué recursos pueden permitirse describir y con qué detalle pueden hacerlo.

El historiador de arte alemán Erwin Panofsky fue el primero en codificar un método para estos problemas de descripción: en una de sus obras más importantes, *Studies in Iconology*, definió tres niveles de descripción que pueden aplicarse a una obra artística o a un artefacto de un museo. En la imagen 3.6: *Contraste entre las descripciones de una obra de arte*, se emplean estos tres niveles para describir una estatua de mármol. La ilustración también muestra las sorprendentes diferen-



cias entre los metadatos EXIF de una foto digital, una descripción automática estandarizada y producida por el propio dispositivo, y los de las descripciones de la estatua realizadas por personas.



EXIF Summary

Make	NIKON CORPORATION
Model	NIKON D90
Aperture	9
Exposure Time	1/320 (0.003125 sec)
Lens	ID AF-S DX VR Zoom-Nikkor 18-105mm f/3.5-5.6G ED
Focal Length	21.0 mm
Flash	Auto, Did not fire
File Size	4.7 MB
File Type	JPEG
Image Height	4288
Image Width	2848
Date & Time	2012:12:03 10:31:14

3 NIVELES

Primario

Estatua de mármol de una mujer desnuda de pie en una concha.

Secundario

Estatua hecha en 2005 por Lucio Carusi de Carrara, Italia. Se titula "Venus". Hecha de mármol de la zona.

Interpretativo

Esta es una transformación en 3D de El nacimiento de Venus, obra hecha en 1486 por el pintor italiano Sandro Botticelli y que actualmente se encuentra en la Galería Uffizi, en Florencia. La Venus de Carusi es más delgada que la de Botticelli debido a que los estándares de la belleza femenina han cambiado.

Las descripciones de obras de arte pueden diferir mucho, sobre todo si unas han sido capturadas por un dispositivo como una cámara digital y otras creadas por personas. Además, estas últimas varían según la experiencia del creador y el grado de interpretación subjetiva aplicado en la descripción.

(Fotografía de R. Glushko. La estatua, titulada "Venus", hecha por Lucio Carusi, de Carrara, Italia. Actualmente forma parte de una colección privada).



3.4.2 LA DESCRIPCIÓN DE IMÁGENES

Las cámaras digitales, incluidas las de los teléfonos móviles, toman millones de fotos cada día. A diferencia de las imágenes de los museos y galerías, la mayoría de estas reciben pocas descripciones más allá de las generadas desde el mismo dispositivo. Sin embargo, muchas llegan a ser descritas someramente en Facebook, Instagram, Flickr, Picasa, DeviantArt u otras plataformas donde se comparten imágenes, o incluso en aplicaciones de fotografía profesional como Light Room. Todos estos sitios permiten que los usuarios asignen etiquetas a las imágenes o las organicen por categorías.

Se han utilizado muchos métodos informatizados diferentes para describir o clasificar imágenes; por ejemplo, usar la “firma visual” de una imagen, extraída a partir de características de bajo nivel como el color, la forma, la textura y la luminosidad. Estas características se utilizan luego para detectar regiones y objetos significativos. Se calcula la similitud de las imágenes para crear categorías de imágenes que contengan los mismos tipos de colores, objetos o conjuntos, lo que facilita la búsqueda de imágenes duplicadas o modificadas.

Para que los ordenadores identifiquen objetos o personas en imágenes, es necesario enseñarlos con imágenes en las que ya están identificados. En 2005, Luis van Ahn ideó una forma ingeniosa de recoger numerosas imágenes etiquetadas con un juego basado en la web llamado ESP. Este emparejaba al azar a personas para que sugiriesen etiquetas con el fin de describir la imagen. Las opciones obvias fueron eliminadas desde el principio, de modo que para una fotografía de un pájaro en un cielo azul no estarían disponibles las etiquetas “pájaro” y “cielo”. Así los usuarios se veían obligados a sugerir otras palabras como “volando” y “menos nubes”. Van Ahn también inventó la técnica reCAPTCHA que consiste en mostrar imágenes de textos antiguos que están siendo digitalizados, lo que mejora la precisión de la digitalización al tiempo que verifica que el usuario de un sitio web es una persona y no un programa robot.

Sin embargo, las únicas características disponibles para que un ordenador reconozca elementos en una imagen son las descripciones cortas de texto o las propiedades de bajo nivel en las imágenes. De lo contrario, las pequeñas variaciones en la posición, la orientación o la iluminación de los objetos dificultarían la tarea de diferenciar elementos que se parecen, como un lobo blanco y la raza de perro blanco llamada Samoyedo. Este problema puede abordarse con técnicas computacionales como las redes neuronales profundas, que implementan la idea de que las características de bajo nivel en las imágenes pueden



combinarse en muchas capas anidadas: los bordes se combinan para formar motivos o patrones; los patrones, para crear partes de objetos familiares; y dichas partes, para constituir objetos completos. Esta composición jerárquica permite que las representaciones de más alto nivel no se vean afectadas por las variaciones de nivel inferior que son un quebradero de cabeza en otros enfoques de automatización.

En 2012, cuando se aplicaron las técnicas de aprendizaje profundo a un conjunto de datos de alrededor de un millón de imágenes que contenían mil categorías de objetos diferentes, la tasa de error se redujo a la mitad. Este espectacular avance, junto con el hecho de que las técnicas de aprendizaje profundo que extraen capas de características a partir de los datos de entrada son completamente generalizables, provocó que el Deep learning se aplicase rápidamente a muchos otros dominios con datos de alta dimensionalidad. Facebook utiliza el aprendizaje profundo para identificar personas en las fotos; Google, para el reconocimiento de voz y la traducción del lenguaje, y lo siguiente que pretende es aplicarlo al subtítulo rápido de imágenes y vídeo. Los dispositivos informáticos “wearables” pueden usar estas técnicas para ofrecer capas de información útil sobre el entorno con el que un determinado individuo está interactuando. De esta forma se crea una realidad aumentada en tiempo real.



3.4.3 LA DESCRIPCIÓN DE MÚSICA

Algunas etapas de la descripción de una canción no son tan diferentes de las de un texto: podemos incluir el nombre del cantante y/o el compositor de la canción, la longitud de la pieza o el nombre del álbum en el que aparece. Sin embargo, ¿qué ocurre si queremos describir el contenido real de la canción? Podríamos escribir la letra, pero para describir la música se requiere un enfoque diferente.

Describir la música presenta desafíos muy diferentes a los que implica la descripción de textos o imágenes. Los poemas y las pinturas son tangibles y se pueden mirar, contemplar. En cambio, por su naturaleza



COMO DESCRIBE Y ORGANIZA SU MÚSICA UN DJ

Los aficionados a la música podrían describir su música usando los nombres de las canciones o intérpretes, o por géneros como "Pop", "Rock" o "Clásica". Sin embargo, un DJ profesional, resalta-
rá otras propiedades de la música, especialmente las "pulsaciones por minuto". Esta foto anotada refleja una parte de la colección de música del conocido DJ "Kid Kameleon" (<http://kidkameleon.com/>). (Foto y anotaciones de Matt Earp. Reproducida con autorización.)

auditiva, la música es un fenómeno fugaz que sólo se puede experimentar en el momento en el que se interpreta. Incluso las partituras y las grabaciones musicales, aunque son tan tangibles como las pinturas y los poemas, no dejan de ser el continente donde se encierra el potencial para producir una experiencia musical y no la música en sí misma. La mayoría de la música popular de hoy en día se presenta en forma de canciones, en las que los textos se ajustan a una melodía y se apoyan en armonías instrumentales. Si queremos categorizarla o describirla según su contenido lírico, podemos confiar en los métodos para describir textos. No obstante, si queremos describir la música en sí misma, debemos adoptar un enfoque diferente.

Muchas personas y empresas del sector multimedia han experimentado con diferentes procesos para describir canciones. Dentro del ámbito de las nuevas tecnologías, destacan las aplicaciones de software como *Shazam* y *Midomi*, que pueden crear una huella digital de audio



a partir de un fragmento de música. Esta huella supone una representación digital de una pieza musical. Luego un ordenador la puede procesar y compararla con otras descripciones digitales de música de una biblioteca.

A primera vista, parece que los servicios actuales de reproducción de música son la cúspide de la clasificación y la descripción de música. *Pandora*, por ejemplo, contrata a musicólogos formados específicamente para escuchar la música y luego clasificar los géneros y los materiales musicales de acuerdo con un vocabulario controlado. El algoritmo resultante, el “Genoma de la Música”, puede aprender a definir los gustos musicales de un oyente por medio de esta clasificación musical y servirse de esa información para sugerirle música similar.

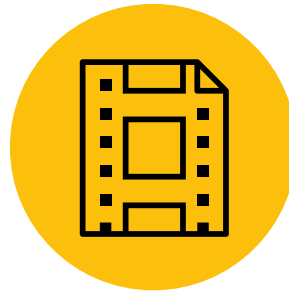
Aunque es cierto que el Genoma de la Música rebosa de complejidad, palidece en comparación con el antiguo y sofisticado trabajo de “pluma y papel” desarrollado por músicos que trabajaron durante siglos para describir la música. Se puede decir que la etnomusicología (definida en términos generales como el estudio de las prácticas musicales globales en sus contextos sociales) ha avanzado más hacia las descripciones exhaustivas de los recursos musicales que cualquier otro campo de estudio musicológico. Desde finales del siglo XIX, los etnomusicólogos han creado métodos complejos de notación y taxonomías estilísticas para captar y categorizar la música de las culturas occidentales y no occidentales.

Si miramos más de cerca, veremos que los músicos son extremadamente innovadores en lo que se refiere a las formas de categorizar, describir y analizar no sólo los géneros musicales a gran escala, sino las propias notas. En la foto que muestra la colección de discos del DJ profesional “Kid Kameleon”, vemos que están dispuestos no sólo según el género, sino también por compases por minuto (BPM). Para Kid Kameleon, estos discos representan los recursos para su proceso creativo musical y ordenarlos por BPM le permite obtener de forma exacta el material musical adecuado para que la música fluya durante una actuación. Por tanto, su sistema de clasificación es una taxonomía que va desde las amplias categorías por género hasta distinciones más meticulosas que ordenan por notas y ritmos. Esta foto no es una mera imagen de una colección de discos: es una representación visual del proceso creativo de un artista.

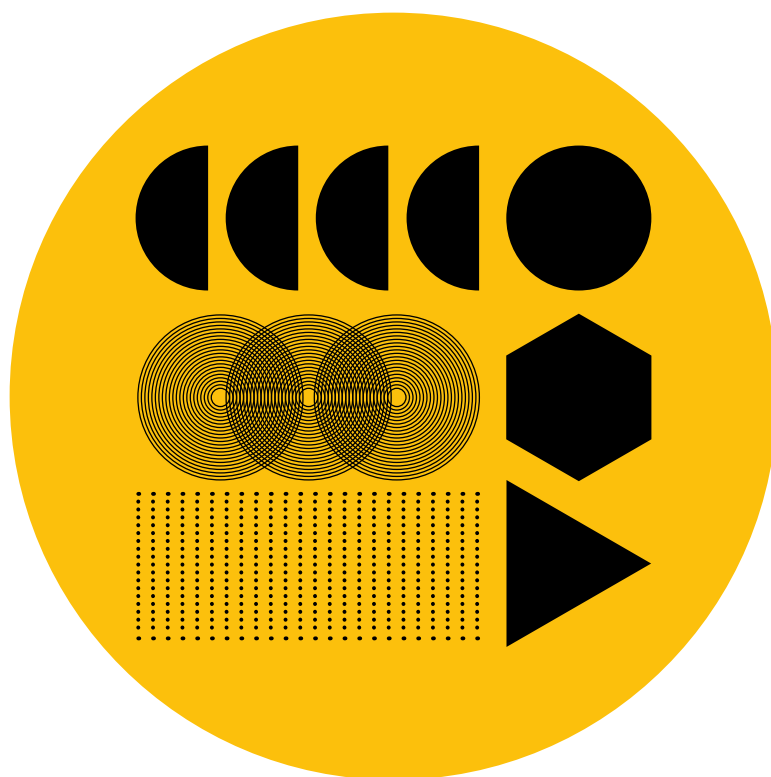


3.4.4 LA DESCRIPCIÓN DE VIDEOS

El vídeo es otro dominio en el que se continúa trabajando para crear descripciones de recursos que posibiliten una búsqueda más eficaz. Las técnicas de análisis disponibles pueden segmentar un vídeo y describir los fragmentos por color, dirección de movimiento, tamaño de los objetos, etc. La identificación de eventos anómalos y rostros de personas en un vídeo pueden aplicarse en la seguridad y la vigilancia. La tarea de detectar detalles específicos de un vídeo requiere actualmente una gran intervención humana, aunque es posible que los algoritmos de comparación de “firmas visuales” de imágenes tomen el relevo en el futuro, pues permitirían también introducir anuncios de forma automática en las plataformas de vídeo y en las emisiones de televisión.



DOS EPÍLOGOS Y UN DESEO





GIGO » LOS DATOS NO SON NÚMEROS, SON ORGANIZACIÓN

Cuando ponemos sobre el tapete el término “datos”, pronto los matemáticos, estadísticos y demás científicos pasan a tomar el control con sus ecuaciones, ratios y modelos. Sin embargo, hablar de datos en relación con la información, el conocimiento y los contenidos es hablar de su potencialidad para ser tratados de forma estructurada, para ser ordenados y organizados más allá de su discurso. Saber que el Ministerio de Transportes, Movilidad y Agenda Urbana se denominada MOPU en los años ochenta son datos; dividir un audiovisual en escenas son datos; relacionar un remake con la primera versión en la que se inspira son datos; conocer las exposiciones y restauraciones que ha tenido un cuadro de un museo son datos; entender los códigos y dependencias de las partidas presupuestarias son datos... datos que permiten organizar y dar sentido a los datos. Metadatos y esquemas de organización del conocimiento, diríamos desde el ámbito de la gestión de información. Ordenar y representar el mundo.

El movimiento por los datos abiertos y la transparencia ha puesto de manifiesto que no sólo se necesita publicar, distribuir o compartir datos y documentos, sino que es necesario hacerlo aplicando modelos y estándares que sostengan su sentido y permitan que sean empleados en circuitos de transformación del conocimiento. Y la cruda realidad es que los datos que la sociedad necesita no siempre pueden extraerse de los sistemas de información actualmente utilizados sin un alto coste en refinamiento y depuración, cumpliéndose la clásica maldición GIGO (*Garbage in, garbage out / Si metes basura, sacas basura*). No siempre es fácil adivinar dónde está lo que producirá basura en sistemas de información complejos. Los datos abiertos y significativos que necesitamos no son espontáneos, necesitan producirse. Es por ello que la organización y gestión de información es una disciplina necesaria para la sociedad de los datos, para que los procesos y sistemas incorporen la “transparencia por diseño”, de forma que los datos finales que se quieren poner al servicio de la sociedad para su reutilización y transformación en valor estén robustamente incorporados en las aplicaciones de gestión diarias, y su puesta en común sea natural y sencilla de producir. Al igual que en otras disciplinas se construye un marco para el “diseño centrado en el usuario”, en el campo de la gestión de información digital es necesario desplegar un amplio marco conceptual y práctico para el “diseño centrado en los datos” que haga viable – sostenible, operativo – el mandato de más y mejores datos como combustible de la sociedad del conocimiento.

Antonia Ferrer Sapena

*Cátedra de Transparencia y Gestión de Datos
de la Universitat Politècnica de València*



DIY » ORGANIZAR INFORMACIÓN

NO TIENE FIN

Esta publicación de FESABID nos recuerda que el reto de organizar información es al mismo tiempo un desafío intelectual y técnico, en el que están participando en paralelo y en competencia agentes de campos con intereses y puntos de vista diversos: la industria editorial, el análisis de datos, los servicios de información públicos, las empresas de diseño y producción digital, los medios de comunicación, las comunidades online... **¿Es el campo de la información tan transversal que se nos ha vuelto invisible?** ¿Es la información una ingeniería, una artesanía, una ciencia?

FESABID ha abierto el camino de traducir parcialmente esta obra sobre “La disciplina de organizar información”, para propiciar la reflexión, porque la consideramos que siempre es necesaria y también porque queremos que nos permita tender puentes de entendimiento entre profesionales en ejercicio y en formación en el macrosector de la información. Desde la Federación, llevamos un tiempo reivindicando la necesidad de garantizar el acceso significativo a la información y el papel importantísimo que para ello los servicios de información con profesionales expertos al frente pueden jugar. Trabajos como este nos ayudan a resituarnos, repensarnos y motivarnos para ser partícipes y, ¿por qué no?, protagonistas en ese campo de la información tan amplio y tan cambiante. En esta línea, también hemos arrancado el Grupo de Trabajo, denominado Nuevos Roles, que pretende explorar, estudiar y dar visibilidad a las muy diversas áreas de actuación y relaciones con otros sectores, y a la aportación de los profesionales de la información en todos estos campos. Como base sobre la que se puede fundamentar un discurso revalorizador y reivindicativo de nuestra profesión, que atraiga talento y refuerce la posición y conocimiento del sector en la sociedad; que genere conversación sobre la formación (universitaria y no universitaria) que estas nuevas realidades requieren.

Quizás no sea habitual en nuestra profesión, pero cada vez es más evidente que habrá que abordar estos asuntos a la manera “*Do it yourself*” y meterse en algo de bricolaje conceptual, juntando piezas, instrumentos y puntos de vista. Se pueden reconocer unos principios compartidos para identificar recursos de información, describir y comprender sus propiedades informativas, para organizarlos en sistemas que facilitan servicios e interacciones, con el fin último de incrementar el acceso intelectual a los contenidos digitales y su reutilización y transformación en valor para la sociedad, sea como negocio en la economía de la información, sea como recurso común en la sociedad digital.



Además, considerando que, esta es sólo una pequeña parte de una obra interesante y necesaria, esperamos que, mediante las diferentes vías que ofrecen las licencias de reutilización, podamos enredar a otros actores públicos y privados para ofrecer una versión completa de esta inspiradora y potente publicación de conceptos básicos que explora la frontera y revitaliza la emocionante empresa colectiva de entender mejor la forma en que podemos dar forma a la inmensa cantidad de información disponible en nuestro universo saturado, excesivo y sorprendente.

Alicia Sellés

Presidenta de Fesabid



ORGANIZAR ES CREAR CAPACIDADES

AL IMPONER DE FORMA INTENCIONADA

UN ORDEN Y UNA ESTRUCTURA

**ORGANIZACIÓN
Y DESCRIPCIÓN DE RECURSOS
DE INFORMACIÓN
DIGITAL**

CLASSIFICATION:
ASSIGNING
RESOURCES TO
CATEGORIES

CATEGORIZATION:
DESCRIBING RESOURCE
CLASSES AND TYPES

THE FORMS
OF RESOURCE
DESCRIPTIONS

DESCRIBING
RELATIONSHIPS
AND
STRUCTURES

INTERACTIONS
WITH
RESOURCES

THE
ORGANIZING
SYSTEM
ROADMAP

CASE STUDIES

**DESCRIPCIÓN
DE RECURSOS
Y METADATOS**

**FUNDAMENTOS DE
LOS SISTEMAS DE
ORGANIZACIÓN**

**LOS RECURSOS EN
LOS SISTEMAS DE
ORGANIZACIÓN**

DESIGN
DECISIONS IN
ORGANIZING
SYSTEMS

ACTIVITIES IN
ORGANIZING
SYSTEMS

Edición "lite"
en español
y en acceso abierto
del libro
**THE DISCIPLINE
OF ORGANIZING**
de Robert.
J. Glushko