



El documento como dato, conocimiento e información¹

Document as Data, Knowledge and Information

Dr. Mario Pérez-Montoro Gutiérrez
Estudis de Ciències de la Informació i de la Comunicació
Universitat Oberta de Catalunya, Internet
Interdisciplinary Institute

Resumen

Las Ciencias de la Documentación se caracterizan por fundamentarse conceptualmente sobre una discriminación básica: la distinción entre los conceptos de *dato*, *información*, *conocimiento* y *documento*. Esta distinción, aun siendo crítica, arrastra una más que profunda confusión conceptual que evita que se pueda realizar un aprovechamiento adecuado (no sólo de tipo conceptual, sino también de tipo pragmático) de la misma dentro de esas disciplinas. En este trabajo presentamos algunas aclaraciones relacionadas con esa distinción. En primer lugar, defendemos un análisis que permite establecer una clara diferenciación entre los conceptos de dato, información y conocimiento. Finalmente, en segundo lugar, mostraremos cuál es el papel articulador del concepto de documento frente a esa distinción anteriormente defendida.

Palabras clave: Documento, Dato, Información, Conocimiento, Ciencias de la Documentación, Gestión del Conocimiento.

Resum

Les Ciències de la Documentació estan fonamentades conceptualment damunt una discriminació bàsica: la distinció entre els conceptes de *dada*, *informació*, *coneixement* i *document*. Aquesta distinció, tot i ser crítica, arrossega una pregona confusió conceptual que impedeix que se'n pugui fer un aprofitament adient (no només de tipus conceptual, sinó també pragmàtic) dins aquestes disciplines. En aquest treball presentem algunes aclaracions relacionades amb aquesta distinció. En primer lloc, defensem una anàlisi que permet d'establir una clara distinció entre els conceptes de dada, informació i coneixement. En segon lloc mostrarem quin és el paper articulador del concepte de document davant aquesta distinció anteriorment postulada.

Paraules clau: Document, Dada, Informació, Coneixement, Ciències de la Documentació, Gestió del Coneixement.

Abstract

Information Science is based conceptually on the distinction between the following concepts: *data*, *information*, *knowledge* and *document*. However, this distinction exhibits a deep conceptual confusion that has hindered an adequate development of the discipline. This paper's goal is to show some clarification relative to this distinction. On the one hand, this article presents an analytical definition of the concepts of data, information and knowledge. On the other hand, it offers a description of the role of the concept of document in this conceptual context.

Key words: Document, Data, Information, Knowledge, Information Science, Knowledge Management.

¹ Los resultados que aparecen en este trabajo se integran dentro del proyecto de investigación KAIMI (*Knowledge Assets Identification and Methodology of Implementation in Organizational Knowledge Management*) que se está desarrollando gracias a la financiación económica del IN3 (*Internet Interdisciplinary Institute*) de la Universitat Oberta de Catalunya.

1. Introducción

En las últimas décadas estamos asistiendo a la consolidación de un importante avance tecnológico: la franca mejora en nuestra capacidad de gestionar datos. En un proceso que comenzó sobre la segunda mitad del siglo XX de la mano del desarrollo de una serie de nuevas tecnologías, se está logrando la gestión de ingentes volúmenes de datos a unas velocidades nunca antes imaginadas.

Pero no pensemos que nos encontramos frente a un recurso del que sólo se benefician las grandes empresas que invierten fuertes sumas en sus departamentos informáticos o los laboratorios universitarios de investigación donde se desarrolla la más puntera de las tecnologías informáticas. Sin ir más lejos, si abrimos cualquier ordenador de sobremesa que tengamos alrededor, seguramente entre su software básico podremos encontrar un programa (Access o File Maker suelen ser los más populares) —que técnicamente se conoce bajo el nombre de “Sistema de Gestión de Bases de Datos”— capaz de gestionar (buscar, recuperar, dar de alta y baja, etc.) de forma casi instantánea y sin apenas problemas una base de datos formada, por ejemplo, por la nada despreciable cantidad de 15.000 registros. Y no olvidemos, además, que la aparición de internet ha permitido que toda esa gestión pueda realizarse *online*, abriéndose la posibilidad del compartimiento de todos esos datos y su gestión independientemente de la ubicación física del usuario.

Como se puede imaginar, estos avances tecnológicos han facilitado e influido en el desarrollo de la mayoría de las disciplinas. En el terreno de la economía, por ejemplo, estos avances han permitido el cálculo y control más adecuado de los balances económicos y la gestión de clientes en el seno de las empresas. Y en el ámbito del derecho, por poner sólo un ejemplo más, el aprovechamiento de estos recursos ha permitido que los abogados tengan acceso instantáneo a la jurisprudencia implicada y necesaria para el correcto desarrollo cotidiano de su praxis jurídica.

Sin embargo, existe un ámbito del saber donde las consecuencias han sido más directas si cabe: en el de las Ciencias de la Documentación. La gestión clásica de documentos se ha visto claramente beneficiada a partir del progreso en el tratamiento tecnológico de los datos. Las bibliotecas y los centros de documentación han podido mejorar (y en muchos casos ampliar) de una manera sustancial la calidad de los servicios que ofrecen a sus usuarios.

Es importante señalar que estos avances han superado la dimensión de lo exclusivamente instrumental y han comportado también otros desarrollos dentro de las Ciencias de la Documentación. En concreto, han permitido, al menos en parte, el nacimiento y la consolidación de una serie de disciplinas aplicables en el seno de las organizaciones y dotadas de una clara dimensión documental.

Entre esas disciplinas podríamos destacar especialmente tres: la gestión del conocimiento, la Auditoría de la Información y la Inteligencia Competitiva. La gestión del conocimiento, por un lado, persigue, mediante el compartimiento y la gestión adecuada, que el conocimiento contribuya de una manera fluida a que los miembros de una organización cubran sus objetivos.² La Auditoría de la Información, en cambio, persigue la identificación y evaluación de la información crítica para el correcto funcionamiento de una empresa. La Inteligencia Competitiva, por último, se encarga de la captura, el análisis y la explotación de la información

² En otro trabajo (Pérez-Montoro Gutiérrez, 2003b) hemos intentado realizar una caracterización más fina de esta disciplina haciendo hincapié especialmente en su dimensión documental.

de tipo técnico-tecnológico que garantice la supervivencia y crecimiento de una empresa frente a la competencia.

Todas esas disciplinas, además de solaparse en algunos puntos, se caracterizan por fundamentarse conceptualmente sobre una discriminación básica: la distinción entre los conceptos de dato, información y conocimiento. Esta distinción, aun siendo crítica, arrastra una más que profunda confusión conceptual que evita que se pueda realizar un aprovechamiento adecuado (no sólo de tipo conceptual, sino también de tipo pragmático) de la misma dentro de esas disciplinas.

En este trabajo nos hemos puesto como objetivo tratar de realizar algunas aclaraciones relacionadas con esa distinción. En primer lugar, en el apartado 2, vamos a intentar presentar un análisis que permita establecer una clara diferenciación entre los conceptos de dato, información y conocimiento. Finalmente, en segundo lugar, en el apartado 3, mostraremos cuál es el papel del concepto de documento frente a esa distinción anteriormente defendida.

2. Dato, información y conocimiento

Como acabamos de señalar, la distinción entre los conceptos de dato, información y conocimiento debe ser considerada como el anclaje conceptual y pragmático de una serie de disciplinas. Pero, ¿cómo podemos entender cada uno de estos conceptos?³

Comencemos ofreciendo un análisis del concepto de *dato*.⁴ Intuitivamente, podemos identificar los datos como acaecimientos físicos (pequeñas parcelas (o *trozos*) de la realidad) susceptibles de transportar asociada cierta información. Poseen una naturaleza material y pueden ser considerados como el soporte físico de la información.⁵ Cada uno de los enunciados (impresos o en soporte electrónico, dependiendo del formato o versión que esté utilizando el lector) que aparecen en este documento pueden ser considerados como datos. Podemos recoger esta propuesta a partir de la siguiente definición:

(a) Dato = soporte físico de la información.

Es importante señalar algunas características de los datos de la mano de esta caracterización. Por un lado, como ya apuntábamos al principio del trabajo, que al ser acaecimientos físicos, los datos son sencillos de capturar, estructurar, cuantificar o transferir. Por otro, que un mismo dato puede informar o no a un agente dependiendo, como veremos a continuación, del stock previo

³ Es importante señalar que existen muchos trabajos, principalmente desde el ámbito de la Gestión del Conocimiento, que intentan clarificar esta distinción. Así, por ejemplo, si revisamos algunas de esas obras ((Nonaka y Takeuchi, 1995), (Davenport, 1997), (Davenport y Prusak, 1998), (Wilson, 1996) o (Webb, 1998), por ejemplo) podemos comprobar que dedican un primer esfuerzo a definir el conocimiento apoyándose en un análisis previo de los conceptos cercanos de *dato* e *información*. La propuesta que presentamos en este trabajo, sin embargo, no recoge estas aportaciones por considerarlas confusas y poco clarificadoras. En este sentido, nuestro análisis debe ser considerado como una alternativa conceptual a esas propuestas.

⁴ Algunas de las ideas básicas sobre las que se fundamenta este análisis del concepto de dato, junto al de información, se encuentran en parte defendidas en Pérez-Montoro, 2003a.

⁵ En un contexto más general, en el de la teoría de la información, podemos identificar la expresión "dato" como sinónima de "señal".

de conocimiento del agente. Por otro, que en el seno de una organización los datos acostumbra a ser conjuntos de caracteres alfanuméricos materializados sobre un documento (físico o electrónico). Y, por último, que en el mismo contexto, el de las organizaciones, la acumulación indiscriminada de datos no siempre lleva necesariamente a una mejora en la toma de decisiones.

Podemos justificar esta manera de definir el concepto de dato revisando cómo se entiende este mismo concepto en otros contextos. Así, por ejemplo, nuestra caracterización recoge sin tensión el sentido que se le da al concepto de *dato* en informática y telecomunicaciones: conjunto de caracteres asociados a un concepto. El conjunto de caracteres "35.879.987" respecto al concepto *número del documento nacional de identidad (DNI)*, podría ser un ejemplo.

En la misma línea, nuestra propuesta encaja perfectamente también con el uso que se hace de la palabra "dato" cuando se definen ciertas aplicaciones informáticas. Un sistema de gestión de bases de datos (SGBDD), sin ir más lejos, suele definirse como un recurso informático que permite la gestión de registros a partir de los datos o conjuntos de caracteres que aparecen en esos registros. En cierto sentido, se puede defender la idea de que la gestión de registros que esas herramientas habilitan es una gestión de tipo sintáctico (a partir de los conjuntos de caracteres que aparecen en los registros) y no de tipo semántico (a partir del contenido informativo asociado a esos conjuntos de caracteres). Frente a una ecuación de búsqueda, el SGBDD recupera aquellos registros donde aparezcan los datos que conforman esa ecuación.⁶ En los mismos términos, un sistema de Data Mining o de Text Mining permite, entre otras cosas, detectar correlaciones o patrones entre datos (o conjuntos de caracteres) que aparecen en los registros que conforman el sistema para que, posteriormente, de una manera intelectual, alguien pueda decidir si ese patrón se corresponde o no con alguna genuina correlación semántica.

Centrémonos ahora en el análisis del concepto de *información*. La información debe ser identificada como el contenido semántico de los datos. En este sentido, la información no posee una naturaleza física o material (como pasaba en el caso de los datos), sino que posee una naturaleza conceptual, pertenece al territorio de lo conceptual.

En el contexto de las organizaciones,⁷ lo que permite que un dato transporte cierta información es la existencia de un código (o clave de codificación) que le asocia cierto contenido informativo. La información no depende de los receptores, sino de la clave de codificación que pone en relación el dato y aquello sobre lo que éste informa. El dato o conjunto de caracteres "Real Madrid – FC Barcelona: X" que aparecen en la página del periódico del lunes informa de que el Barça y el Real Madrid han empatado gracias a la existencia de un código (una clave de codificación) que le asocia ese contenido semántico al dato físico. Estas ideas podrían resumirse presentando la siguiente definición:

(b) Información = contenido semántico del dato derivado de una clave de codificación.

⁶ Esto ocurre, evidentemente, en aquellos SGBDD en los que no se ha implementado un thesaurus como recurso para mejorar la recuperación de información.

⁷ Insistimos en remarcar el tipo de contexto para indicar indirectamente que la mayoría de los datos que suelen circular en el seno de una organización suelen ser de tipo convencional y, por tanto, que esa codificación es también de naturaleza convencional (no natural).

En esta misma línea, es importante distinguir entre dos fenómenos relacionados pero a la vez distintos: *transportar información* y *adquirir información*. *Transportar información* es una propiedad que poseen los datos gracias a la existencia de claves de codificación. *Adquirir información*, en cambio, es una propiedad que poseen los individuos (o agentes) que ejemplifican cuando son capaces de asimilar, a partir de su stock previo de conocimiento, la información que transporta un dato al interpretar ese dato a la luz de la clave de codificación que está en juego.

Pasemos, por último, al análisis del concepto de *conocimiento*. El conocimiento debe ser identificado con un tipo especial de estados mentales (o disposiciones neuronales) que posee un individuo y que presentan una serie de características propias. Por un lado, son estados mentales que adquiere el individuo a partir de un proceso de asimilación o metabolización de información.⁸ En este sentido el contenido semántico de esos estados mentales coincide con esa información asimilada. Y, por otro, actúan de guía en las acciones y la conducta de ese individuo. Podemos plasmar estas ideas en la siguiente definición:

(c) Conocimiento = estados mentales de un individuo contruidos a partir de la asimilación de información y que rigen las acciones del propio sujeto.

Sin embargo, las características del conocimiento no acaban aquí. Podemos abundar un poco más sobre este tipo especial de estados mentales. El conocimiento, a diferencia de los datos y la información, se encuentra estrechamente relacionado con las acciones y las decisiones del sujeto que lo realiza; se llega incluso a poder evaluar ese conocimiento utilizando como indicativo esas acciones y decisiones. El conocimiento, además, es el factor crítico que permite la asimilación de nueva información —y la creación de nuevo conocimiento, por tanto— por parte del sujeto que lo posee y suele verse reestructurado continuamente por las entradas de nueva información asimilada.

Para acabar la presentación de esta distinción, nos queda señalar también que la caracterización de estos tres conceptos que acabamos de introducir permite, en la mayoría de las circunstancias, la discriminación física entre los datos, la información y el conocimiento. En concreto, según nuestra propuesta los datos, información y conocimiento se ubicarían respectivamente en tres niveles diferentes. En primer lugar, los datos se hallarían en el territorio de lo físico. Los datos, como acaecimientos físicos (como trozos o parcelas de la realidad) estarían dotados de naturaleza material. En segundo lugar, la información no sería un conjunto especial de datos, sino que se situaría en el territorio de lo conceptual. La información sería el contenido semántico de los datos.⁹ Un mismo dato podría transportar diferentes informaciones y una misma información podría ser transportada simultáneamente por diferentes datos. Por último, el conocimiento, como estado cerebral o disposición neuronal, pertenecería al territorio de lo mental.

⁸ Esta característica ayuda a distinguir los estados mentales del sujeto que se corresponden con el conocimiento de aquellos estados mentales del sujeto que se corresponden con meras creencias que no alcanzan el nivel epistémico necesario para poder identificarlas como conocimiento.

⁹ Estrictamente hablando la información debería identificarse con una proposición, con un contenido proposicional dotado de un valor de verdad.

3. El concepto de documento frente a la distinción

Una vez que hemos esbozado la distinción entre los conceptos de dato, información y conocimiento, nos queda describir cuál es el papel que juega el concepto de documento dentro de todo este escenario.

Comencemos definiendo qué es un documento. En términos generales, podemos afirmar que los documentos siempre se han visto involucrados en la actividad intelectual del ser humano. Desde el principio de la historia del pensamiento, el hombre ha utilizado una serie de objetos o materiales donde poder plasmar y almacenar aquello que pensaba o sentía. Las pinturas rupestres, las tabletas de arcilla mesopotámicas, los muros de los edificios sagrados egipcios, los papiros, los pergaminos o, posteriormente, el papel, son claros ejemplos de tipo de objeto o materiales. En la actualidad, con el desarrollo de las tecnologías de la información y la comunicación, se está apostando cada día más por los formatos electrónicos para recoger nuestra producción intelectual.

Para referirnos a todo este tipo de objetos o materiales utilizamos normalmente el término "documento". O, dicho de otra manera, podemos identificar como documento todo aquel soporte donde se represente algún tipo de información. En este sentido, podemos incluir bajo el concepto de documento una hoja de papel escrito, un libro, una fotografía, una cinta de video, un DVD, un archivo creado con un procesador de textos, una base de datos o una página web. Expresándolo en forma de definición:

(d) Documento = todo soporte donde se represente información.

Como se desprende de la definición, los documentos se caracterizan por poseer dos dimensiones. Por un lado son algo físico y, por otro, contienen asociado un contenido informativo o información. Veamos, a partir de estas dos dimensiones, la relación con los otros tres conceptos.

La relación entre los conceptos de documento y dato parece bien sencilla. Si, como hemos ya señalado, el dato es el soporte físico de la información, el documento debe ser entendido como un conjunto de (un tipo especial de)¹⁰ datos.

Veamos, en cambio, qué ocurre con la relación que mantiene con el concepto de información. Si, como hemos defendido, la información debe entenderse como el contenido semántico del dato derivado de una clave de codificación, el documento aparece como ese objeto material donde se puede representar y materializar información. Esta representación y materialización permite explicar varias cosas. Por un lado, permite explicar cómo se puede transmitir la información: la información se representa (se asocia) en (a) un documento a partir de un código y su transmisión se produce a partir de la propia transmisión material del documento. Y, por otro lado, permite también dejar claro por qué la conservación y almacenamiento del documento significa también la conservación y almacenamiento de la información que éste contiene. Sólo cabe analizar ese documento bajo la misma clave de codificación (o código) que

¹⁰ Decimos "especial" ya que los documentos coincidirían exclusivamente con conjuntos de datos que poseen una naturaleza convencional.

se utilizó para asociarle ese contenido semántico concreto para poder recuperar esa información después del almacenamiento del documento.

Abordemos, por último, la articulación del concepto de documento frente al de conocimiento. Como ya hemos señalado, el conocimiento debe entenderse como aquellos estados mentales de un individuo construidos a partir de la asimilación de información y que rigen las acciones del propio sujeto. Frente a estos estados mentales, y a partir de su dimensión física y su capacidad de transportar información, el documento juega un papel muy importante: éste aparece como ese objeto material donde se puede representar y materializar esos estados mentales que residen exclusivamente en la cabeza de las personas. Y, al igual que pasaba en el caso de la información, esta representación y materialización permite explicar la transmisión y el almacenamiento de conocimiento (explícito)¹¹ a partir de la transmisión y almacenamiento de documentos.

En este sentido, por un lado, el conocimiento residente en la cabeza de un individuo se representa (se plasma) en un documento a partir de un código y su transmisión se produce a partir de la propia transmisión material del documento. Cuando un segundo individuo es capaz de obtener la información asociada a ese documento transmitido y formar un nuevo estado mental a partir de la misma, podemos afirmar que se ha producido la transmisión de ese conocimiento. Y, por otro lado, a partir del mismo mecanismo, la conservación y almacenamiento del documento que se obtiene como fruto de la representación de un conocimiento concreto permite también la conservación y almacenamiento de ese conocimiento. Basta analizar ese documento bajo la misma clave de codificación (o código) que se utilizó en la representación de esos estados mentales para poder recuperar la información asociada y crear nuevos estados mentales en otros individuos después del almacenamiento del documento. De esta manera ese conocimiento puede ser recuperado por cualquiera que lo necesite en el momento adecuado.

En esta misma línea, para concluir este breve esbozo, es importante señalar también una cosa más que puede aclarar todo este escenario conceptual. No debemos olvidar que, en ciertas ocasiones y en términos coloquiales, solemos clasificar como información o conocimiento un dato concreto. En el contexto de las organizaciones, solemos utilizar también las expresiones “conocimiento” e “información” para referirnos a las representaciones físicas de esos estados mentales o de esos contenidos informativos, para referirnos a los documentos (en cualquiera de sus soportes (papel, electrónico, óptico, magnético, etc)) que utilizamos para representar y difundir ese conocimiento o esa información. Así, por ejemplo, si un documento (un dato, un acaecimiento físico) transporta cierta información o se ha obtenido como fruto de la representación de un conocimiento que posee un sujeto, en un sentido lato, solemos decir también que ese documento es, respectivamente, *información* o *conocimiento*.

¹¹ Es importante señalar que nos estamos refiriendo exclusivamente al conocimiento de tipo explícito. El conocimiento de tipo tácito, en cambio, necesita un tratamiento ostensiblemente diferente. Para transmitir el conocimiento tácito es imprescindible el contacto humano entre los que lo poseen y los que quieren acceder a él. De esta manera, las estrategias para gestionar ese tipo de conocimiento en el seno de las organizaciones suelen ir encaminadas a fomentar el intercambio a través de comunidades de práctica, ferias de conocimiento o simplemente creando una cultura organizacional que facilite los contactos informales entre sus miembros.

4. Bibliografía

- Davenport, Thomas H. (1997). *Information Ecology*. Oxford: Oxford University Press.
- Davenport, T.; Prusak, L. (1998). *Working Knowledge*. Boston: Harvard Business School Press.
- Devlin, Keith (2001). *InfoSense. Turning information into Knowledge*. New York: W. H. Freeman and Company.
- Nonaka, Ikujiro y Takeuchi, Hirotaka (1995). *The Knowledge Creating Company*. Oxford: Oxford University Press.
- Pérez-Montoro Gutiérrez, Mario (2003a). *The Phenomenon of Information*. Lanham (Maryland): Scarecrow Press.
- Pérez-Montoro Gutiérrez, Mario (2003b). "La Gestión del Conocimiento y su dimensión documental". En Fuentes Pujol, Eulàlia (ed.) (2002). Anuario SOCADI de Documentación e Información. Barcelona: Societat Catalana de Documentació i Informació.
- Pérez-Montoro Gutiérrez, Mario (2001). "La información como fundamento cognitivo de una definición adecuada de conocimiento". En Extreño Placer, Ana (ed.) (2001) *La representación y organización del conocimiento: metodologías, modelos y aplicaciones*. Alcalá de Henares, págs. 79-87.
- Wilson, D. A. (1996). *Managing Knowledge*. Oxford: Butterworth-Heinemann & Institute of Management
- Webb, Sylvia H. (1998). *Knowledge Management: Linchpin of Change*. London: Aslib. ISBN 0-85142-414-7.