

JUAN VOUTSSÁS M.

El concepto de documento
de archivo digital:
una propuesta de

Ontología

The concept of digital
record: an intended
ontology



La presente obra está bajo una licencia de:
<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>



Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)

Este es un resumen legible por humanos (y no un sustituto) de la [licencia](#). [Advertencia](#).

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

Adaptar — remezclar, transformar y construir a partir del material

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



NoComercial — Usted no puede hacer uso del material con [propósitos comerciales](#).



CompartirIgual — Si remezcla, transforma o crea a partir del material, debe distribuir su contribución bajo la la [misma licencia](#) del original.

**El concepto de documento de archivo digital:
una propuesta de ontología / The concept
of digital record: an intended ontology**

COLECCIÓN
SISTEMATIZACIÓN DE LA INFORMACIÓN DOCUMENTAL
Instituto de Investigaciones Bibliotecológicas y de la Información

**El concepto de documento de archivo digital:
una propuesta de ontología / The concept
of digital record: an intended ontology**

Juan Voutssás Márquez



**Universidad Nacional Autónoma de México
2022**

CD947
V68

Voutssás Márquez, Juan.

El concepto de documento de archivo digital : una propuesta de ontología = The concept of digital record : an intended ontology / Juan Voutssás Márquez.- México : UNAM. Instituto de Investigaciones Bibliotecológicas y de la Información, 2022.

88 p. - (Sistematización de la información documental)

ISBN: 978-607-30-6069-1

Teoría archivística. 2. Archivos (Documentos). 3. Ontologías (Recuperación de la información). I. Título II. ser.

Diseño de cubierta: Mario Ocampo Chávez

Primera edición: 03 de mayo de 2022

D.R. © UNIVERSIDAD NACIONAL

AUTÓNOMA DE MÉXICO

Instituto de Investigaciones Bibliotecológicas
y de la Información

Circuito Interior s/n, Torre II de Humanidades,
pisos 11, 12 y 13, Ciudad Universitaria, C. P.
04510, Alcaldía Coyoacán, Ciudad de México

ISBN: 978-607-30-6069-1

Esta edición y sus características son propiedad
de la Universidad Nacional Autónoma de
México. Prohibida la reproducción total o
parcial por cualquier medio sin la autorización
escrita del titular de los derechos patrimoniales.

Publicación dictaminada

Impreso y hecho en México

Contenido

Sección 1: Ontologías	1
Sección 2: La ontología propuesta	19
Referencias bibliográficas:	41

THE CONCEPT OF DIGITAL RECORD: AN INTENDED ONTOLOGY

Section 1: Ontologies	51
Section 2: The intended ontology	67
Bibliographic references	85

*Quiero expresar mi más profunda gratitud
a Kenneth Thibodeau, Luciana Duranti, Alicia
Barnard, Claudia Lacombe Rocha, Giovanni
Michetti y Rosely Rondinelli, de los equipos
de InterPARES, quienes aportaron generosamente
sus conocimientos y experiencia contribuyendo en
gran medida a la elaboración de las tablas, aunque
no siempre estuvieron de acuerdo con todas sus
interpretaciones o conclusiones.*

*Las cosas existen cuando tienen nombre; la
sustancia de las cosas proviene principalmente
del nombre, y aun desapareciendo, su esencia
permanece en el nombre.*

UMBERTO ECO
Apostillas a *El nombre de la rosa*

Sección 1

Ontologías

Tenemos un conocimiento común adquirido en torno al concepto de “documento de archivo” (*record*) en la Archivística, y es obviamente un requisito esencial e indispensable para cualquier persona responsable de su gestión y/o conservación. Durante mucho tiempo, tanto la Archivística como la Diplomática han estudiado, desmenuzado y analizado este concepto. Las definiciones resultantes sobre el mismo se han presentado principalmente a través de diccionarios o glosarios, o como “definiciones puntuales” en algunas normas, como la ISO-15489-1, o en proyectos relacionados con el tema, como MoReq. Durante muchos años, estas definiciones se refirieron a los documentos de archivo en “soportes tradicionales”, pero a la luz de los documentos de archivo digitales, en cierto momento fue necesario reconsiderar este conocimiento, ya que el anterior había quedado obsoleto con la llegada de este nuevo tipo de documentos. El aumento global de la información creada digitalmente –especialmente los documentos de archivo digitales– y su creciente importancia, requería el desarrollo de nuevas herramientas para su conocimiento, gestión y conservación. Por ello, era necesario trabajar con conceptos contemporáneos desarrollados en contextos de esta versión digital

y no solo con aquellos basados únicamente en las definiciones de los soportes tradicionales; no porque fueran erróneos: simplemente se habían vuelto obsoletos.

Pareciera que el concepto de “documento de archivo” en su acepción archivística está más que definido; pero el concepto es más complejo de lo que se ve a primera vista. La teoría archivística moderna establece una clara diferencia entre “documento” y “documento de archivo”, así como entre los distintos tipos de estos últimos, pero todavía hay entornos en los que estos conceptos se entremezclan y las diferencias no se entienden del todo bien. Además, el término es ahora más complejo desde que los “documentos de archivo” se han vuelto cada vez más digitales. Esto no ha sido solo un cambio de soporte, sino que ha supuesto un profundo cambio de la naturaleza inherente de aquellos. Durante algunos años, las definiciones trataban los documentos de archivo digitales de la misma manera que lo hacían con los documentos de archivo tradicionales, como si fueran lo mismo, solo que puestos en un soporte digital. Hoy sabemos que, debido a su naturaleza, los documentos de archivo digitales son muy diferentes de los tradicionales, especialmente en lo que respecta a su gestión y conservación. Por ello, muchas definiciones de “documento de archivo” han quedado obsoletas, incluso muchas de las que intentan definir “documento de archivo digital”.

Como ejemplo de ello, podemos utilizar algunas definiciones comunes, interesantes y compartidas desde hace tiempo acerca de “documento de archivo” (*record*) y “documento de archivo digital” (*digital record*):

- El “Glosario de Terminología Archivística” de la Asociación de Archivistas Estadounidenses presenta siete definiciones para “documento de archivo”. Las dos más cercanas a la Archivística son: 2) “Datos o información que ha sido fijada en algún soporte que tiene contenido, contexto y estructura; y que se utiliza como una extensión de la memoria humana o para demostrar la responsabilidad”. 3) “Datos o información en una forma fija que se crea o recibe en el

curso de una actividad individual o institucional y que es apartada (preservada) como evidencia de esa actividad para futuras referencias”. En cuanto al “documento de archivo digital” (*digital record*), el glosario se refiere al “documento de archivo electrónico” (*electronic record*): “Datos o información que han sido capturados y fijados para su almacenamiento y manipulación en un sistema automatizado y que requieren del uso de un sistema para hacerlos inteligibles por una persona” (Pearce-Moses 2005).

- La base de datos terminológica del proyecto InterPARES define “documento de archivo” como “Documento elaborado o recibido durante el curso de una actividad práctica –ya sea como instrumento o subproducto de dicha actividad, y que es separado (apartado, guardado) para acción posterior o referencia”. En cuanto al “documento de archivo digital”. El proyecto define “Un documento digital que es tratado y manejado como un documento de archivo” (InterPARES Terminology Database 2012). Esta definición se perfeccionó posteriormente como:

[...] documento producido (elaborado o recibido y apartado –es decir, conservado, guardado– para acción o referencia) por una persona física o jurídica en el curso de una actividad práctica como instrumento y subproducto de dicha actividad. Cuando el documento de archivo se aparta y se utiliza en forma digital, entonces es un documento de archivo digital, independientemente de la forma original en la que se haya elaborado o recibido (Duranti y Thibodeau 2006, 3).

- En la norma internacional ISO-15489, se define “documento de archivo” como “la información producida, recibida y retenida como prueba e información por una organización o persona en cumplimiento de sus obligaciones legales o en la transacción de sus negocios [...]” (ISO-15489 2001). La norma no incluye una definición específica para documento de archivo digital o electrónico.

El concepto de documento...

- La especificación MoReq2 define ‘documento de archivo’ como “documento(s) preparado(s) o recibido(s) por una persona u organización en el curso de su actividad, y conservado(s) por dicha persona u organización”. Para ‘documento de archivo electrónico’ define: “[...] un documento de archivo en formato electrónico como resultado de haber sido producido por un software de aplicación o como resultado de la digitalización, por ejemplo, por escaneo” (European Commission 2008).
- La base de datos terminológica internacional multilingüe del Consejo Internacional de Archivos (ICA) se refiere al “documento de archivo” prácticamente igual que la ISO-15489: “Un documento producido o recibido y mantenido por una agencia, organización o individuo en cumplimiento de obligaciones legales o en la transacción de negocios” (ICA 2014). Para “documento de archivo digital”, se refiere exactamente a las dos definiciones del “Glosario de Terminología Archivística de la Sociedad” de Archivistas Estadounidenses y a la del Proyecto InterPARES antes mencionadas.

Observando las definiciones presentadas –actuales como son– e incluso añadiendo algunas más, es fácil concluir que todas ellas se asemejan entre sí y, en efecto, dan una buena idea del término, destacando algunas de ellas varios elementos que parecen especialmente relevantes para sus autores. Pero más allá de las simples definiciones, pueden plantearse algunas cuestiones adicionales: ¿Cuáles son los componentes de un documento de archivo digital? ¿Tiene subcomponentes? ¿En cuántos niveles? ¿Cuáles son las relaciones y/o condiciones de estos subcomponentes? ¿Cómo se percibe el todo? ¿Cómo se desmenuza? Intentar responder a estas preguntas a partir de glosarios sería posible, pero para ello habría que leer, analizar y ponderar muchos documentos adicionales; habría que hacer muchas gráficas, tablas y bocetos para lograr un resultado lógico y coherente. Sería necesario leer una y otra vez listas de términos para establecer conjuntos, relaciones, reglas,

equivalencias, restricciones, términos ascendentes y descendentes, etcétera, entre todos estos conceptos.

Habría que leer todavía más para contextualizarlos, lo que llevaría más tiempo. Aún con ello, no podríamos responder directamente a partir de los glosarios a preguntas más profundas sin un bagaje adicional sobre el concepto. Esto se debe a que ellos sólo sirven para definir términos sueltos, pero no dentro de un contexto o estructura posterior que pueda definir no solo los términos sino sus interrelaciones, sus niveles y las condiciones que deben cumplirse para combinar estos términos; esto es, para entenderlos a fondo. Las definiciones de diferentes fuentes son similares, pero no idénticas, y algunas de ellas incluyen elementos que no se encuentran en otras. Las organizaciones y los proyectos que han emprendido la tarea de conceptualizar lo han hecho de diversas maneras, tratando de definir al máximo los elementos que les permitan establecer de la manera más clara y sencilla el concepto de “documento de archivo digital”. En suma, para conocer los conceptos resultantes se suele consultar diccionarios o glosarios; a veces libros o artículos, pero rara vez ontologías, y la razón principal para hacerlo es que son escasas en el ámbito de los documentos de archivo, y lo son todavía más en el de los documentos de archivo digitales. Cabe aquí la pregunta ¿por qué ontologías para definir el “documento de archivo digital”?

Con las nuevas tendencias mundiales, cada vez más se requiere que la información gubernamental y sus archivos estén disponibles para el público, y se busca acceder a ellos en mayor medida a través de la Web, por lo que esta información debe estar preparada para ello. El aumento global de la información creada digitalmente –especialmente los documentos de archivo digitales– y su creciente importancia han requerido el desarrollo de nuevas herramientas para su conocimiento, gestión y preservación. La comunidad archivística internacional ha elaborado en los últimos años valiosas conceptualizaciones y aportaciones teóricas y prácticas para la gestión moderna de enormes cantidades de este tipo de datos e información. La ciencia archivística, combinada con otras disciplinas emergentes, ha hecho en los últimos años

contribuciones fundamentales a la innovación del conocimiento para mejorar el acceso continuo y mejor a los documentos de archivo, incluyendo los digitales.

Una de las propuestas globales emergentes con soluciones técnicas innovadoras para la representación y gestión del conocimiento ha sido la “Web Semántica”, basada en una serie de recomendaciones realizadas por el World Wide Web Consortium (W3C), cuyo objetivo ha sido un mejor acceso a la información de todo tipo en la Web por parte de un gran número de usuarios e instituciones a través de un uso optimizado de la red mundial. Una de las premisas básicas de esta propuesta ha sido el uso de “representaciones semánticas” documentales. Se trata, en suma, de un conjunto de principios y herramientas para mejorar la descripción formal de los documentos, la cual a su vez incidirá favorablemente en la búsqueda y recuperación documental. El principio subyacente en ello consiste en que los documentos descritos formalmente bajo una serie de reglas y estructuras se hacen mucho más accesibles a través de sistemas interoperables en el ambiente de la red mundial. Aplicar esto en combinación con el tratamiento moderno de los documentos de archivo digitales, promete ser una mejor solución para la descripción y gestión de ese tipo de documentos, que mejorará su acceso tanto en los sistemas locales de gestión archivística como a nivel de la Web. Los documentos de archivo contruidos con estos principios se convertirían así en “documentos de archivo semánticos” que conformarían “archivos semánticos”.

Además, mejores representaciones formales de los atributos de los documentos de archivo digitales dentro de los nuevos sistemas de preservación de esos documentos contribuirá a mantenerlos más estables a lo largo de varias generaciones. Estas representaciones pueden ayudar definitivamente en el futuro a diseñar y aplicar mejores estrategias de conservación, migración, transferencia, etcétera, de los documentos de archivo digitales. Es sabido que la forma de los documentos de archivo debe ser fija y su contenido debe ser estable, pero la tecnología no lo será. También es sabido –por paradójico que esto suene– que para preservar los documentos de archivo digitales, estos no deben permanecer inamovibles

en su formato original, pues de lo contrario, este mismo hecho los hará inaccesibles con el tiempo. Está claro que, hoy en día y en el futuro, uno de los principales retos para la preservación de los documentos de archivo digitales es cómo hacer frente a la siempre cambiante tecnología: *hardware*, sistemas operativos, formatos, programas, etcétera. Todos estos elementos tienen un impacto directo y afectan el uso, la gestión, la autenticidad y la preservación de esos documentos de archivo. Mejores representaciones formales de esos documentos ayudarán sin duda a que transiten de manera más fluida y segura por todos esos cambios.

Adicionalmente, las comunidades de usuarios han cambiado en los últimos años; antes, los documentos de archivo eran para uso casi exclusivo de unos pocos funcionarios dentro de sus organizaciones, pero hoy en día se requiere cada vez más que un público usuario pueda acceder a ellos, con un número de documentos que está creciendo exponencialmente. Los principios de acceso a la información pública, transparencia, rendición de cuentas, gobierno abierto, democracia participativa, etcétera, se han vuelto cada día más universales y hacen que el público demande más y más información proveniente de archivos. Esos usuarios necesitan mejores herramientas para acceder, combinar, analizar y reutilizar esa información; pero para que ella adquiera realmente un mayor significado para los usuarios se requiere indispensablemente su vinculación con otras informaciones similares. Por lo mismo, es imprescindible combinar a corto plazo el conocimiento archivístico actual con una adecuada gestión de la información bajo estas nuevas representaciones formales como solución para un acceso mejorado a los mismos. Las aproximaciones semánticas prometen mejorar el acceso a la información así representada.

La Web Semántica ha introducido toda una estructura conceptual formada por ciertos componentes, algunos ya existentes y otros nuevos. Hay componentes tecnológicos y conceptuales, accesorios, etcétera. Es ya conocido el esquema estructural de la Web Semántica introducido por Berners-Lee (2001), el cual contiene capas tecnológicas que tienen más que ver con el *hardware*, el *software*, las redes, etcétera, así como nuevas capas conceptuales

como los estándares XML, RDF2 y OWL que pretenden convertir la web en esa infraestructura lógica global donde los datos y los documentos puedan ser compartidos y reutilizados entre diferentes tipos de sistemas y usuarios; es una estructura compleja. La capa OWL (Ontology Web Language o Lenguaje Ontológico para Web) es una de sus herramientas destinadas a desarrollar vocabularios específicos para asociar recursos documentales formalmente (W3C 2010). OWL proporcionó una estructura formal estándar de la Web destinada a representar explícitamente el significado de los términos y las relaciones entre esos términos en los vocabularios; se basa en un marco de la Lógica de la Descripción (DL) que constituye la base de varios lenguajes ontológicos conocidos (Baader *et al.* 2004, 3-28). OWL nunca llegó a consolidarse como metodología universal para ello, pero su principio de utilizar una ontología para realizar la tarea se ha vuelto algo bastante relevante y es la base de muchos otros esfuerzos en este aspecto. Los enfoques actuales de representación del conocimiento con ontologías se estudian cada vez más como una herramienta interesante, ya que, en teoría, el conocimiento debe representarse de una manera que sea formal, consensuada, reutilizable y eventualmente legible por computadora. Las ontologías proporcionan una forma de representar el conocimiento de ese modo. Pero, exactamente ¿qué son estas “ontologías”? ¿Cuál es su utilidad potencial en la Archivística moderna?

Originalmente, el término “ontología” se ha utilizado durante mucho tiempo en la Filosofía, en particular en la Metafísica, como un estudio que pretende definir qué cosas existen en la realidad y qué se puede decir de ellas. Tiene que ver con la ciencia de lo que “es”, el “ente”, y sus propiedades trascendentes; su esencia es el estudio del “ser” respecto a lo que “es” y “cómo es” para lograr una explicación sistemática de ello. En este sentido filosófico, una ontología es, por tanto, una estructura categorial particular que representa una determinada visión del mundo sobre la naturaleza y organización de la realidad; es decir, sobre lo que “existe”. En tiempos más recientes, el término y algunas de sus técnicas originales han sido adoptados por diversas disciplinas que pretenden,

más allá de definir el “ser”, establecer las categorías fundamentales de las cosas a partir del estudio de sus propiedades, atributos y estructuras; con ello, se ha convertido en un nuevo y variado concepto. Así, en suma, se puede distinguir entre la “Ontología” como conciencia filosófica del ser, y las “ontologías” como conceptualizaciones construidas a partir de ciertos conocimientos “[...] en un sentido amplio, las ontologías son conjuntos de reglas de inferencia para interpretar un sistema de símbolos [...] las inferencias basadas en ontologías pueden utilizarse para extraer información implícita de un conjunto de representaciones semánticas en una base de conocimientos” (Arakawa 2008, 2).

La ontología como conceptualización es un concepto complejo. Guarino y Giaretta (1995) realizaron un estudio donde encontraron siete enfoques diferentes del concepto con sus correspondientes interpretaciones; en resumen: 1) el mencionado enfoque tradicional: ontología como parte de la Filosofía; 2) ontología como un sistema conceptual informal; 3) ontología como una explicación formal semántica; 4) ontología como una especificación de una conceptualización; 5) ontología como una representación de un sistema conceptual a través de una teoría lógica, caracterizada ya sea por propiedades formales específicas o solo para propósitos específicos; 6) ontología como el vocabulario utilizado por una teoría lógica; 7) ontología como una especificación o un meta-nivel de una teoría lógica. Así pues, existen múltiples enfoques de las ontologías, los cuales comprenden diferentes interpretaciones; obviamente, “ontología” es un término polisémico, y cada una de sus interpretaciones tiene diversas manifestaciones y utilidad para diferentes usos y disciplinas: en lingüística y literatura, en inteligencia artificial, en la teoría del conocimiento, en ciencias biomédicas, en la geomática y en la informática –especialmente en el diseño de bases de datos–, en la recuperación de información y en la lingüística computacional. Muchas otras disciplinas las utilizan cada vez más y, por supuesto, también pueden ser útiles para la Archivística. Debido a esta gran variedad disciplinar, es natural observar diferentes enfoques y tratamientos del término, y por lo tanto es necesario estar conscientes de que tiene múltiples usos por parte

de diferentes comunidades con diferentes contextos que por su variedad podrían llevar a confusión durante su estudio y uso.

Para entender el concepto en su esencia básica, conviene partir de las definiciones más generales y comunes. En su forma más simple, *una ontología es una teoría que explica el conocimiento*.

- Una definición muy utilizada es la propuesta básica elaborada por Gruber (1995), “una ontología es una especificación explícita de una conceptualización”. Borst (1997, 20) la modificó ligeramente al afirmar que “la ontología es una especificación formal de una conceptualización compartida”. Según ellos, en estas definiciones ‘conceptualización’ significa un modelo o idea de algún objeto o hecho del mundo cuyos conceptos relevantes se identifican: objetos, conceptos y otras entidades que se supone que existen en algún área de interés, así como las relaciones que los contienen; es una visión abstracta y simplificada del mundo que debe representarse para algún propósito; ‘explícito’ significa que el tipo de conceptos utilizados y las restricciones para su uso se definen de forma clara y manifiesta; ‘formal’ se refiere a que la ontología es legible por una máquina o, al menos, factible de ser traducida a ella; ‘compartido’ significa que describe un conocimiento consensuado y aceptado por un determinado grupo .
- Swartout y colegas (1996) afirmaron que una ontología es “[...] un conjunto de términos estructurados que describen un tema o ‘dominio’. Su objetivo es proporcionar una estructura básica para una base de conocimientos”.
- Según Neches y colegas (1991), “[...] una ontología define los términos básicos y las relaciones que componen el vocabulario de una determinada área temática, así como las reglas de combinación de términos y relaciones para definir extensiones de dicho vocabulario”.
- Para Uschold y Grüninger (1996):

[...] una ontología es un conjunto de agrupaciones o ‘clases’ con diferentes tipos de relaciones entre ellas. La ‘clase’ principal es la que describe la taxonomía¹ de la ontología. Las relaciones tienen propiedades cuya máxima utilidad es la capacidad de hacer inferencias a partir de ellas sobre las clases[...] ontología es el término utilizado para referirse a la comprensión compartida y puede utilizarse como marco unificador para resolver problemas... una ontología implica o representa necesariamente algún tipo de visión del mundo con respecto a un dominio o tema concreto. Esta visión suele concebirse como un conjunto de conceptos (por ejemplo, entidades, atributos y procesos), sus definiciones y sus relaciones, lo que se conoce como conceptualización.

- Noy y McGuinness (2001) afirman que “[...] la ontología es una descripción formal y explícita de los conceptos en un dominio de acción que consiste en clases, donde cada una de ellas tiene propiedades –que describen varias características, roles y atributos del concepto– así como restricciones”.
- Arano (2005, 7) define ontología como

[...] una representación formal y explícita de la estructura conceptual de un campo de conocimiento. La ontología es un soporte semántico para las palabras que se describen como objetos lingüísticos en una base de datos léxica o terminológica. Las relaciones conceptuales representadas en una ontología tienen una amplia variedad que depende del campo de conocimiento a

1 “Una taxonomía es un vocabulario controlado organizado en una jerarquía. Cada término se refiere a una categoría, tipo o clase. Solo hay un tipo de vínculo, que significa ‘es una variedad de’ y corresponde a una relación de subclase. En sentido estricto, cada nodo de una taxonomía tiene exactamente un ‘padre’, pero el término ‘taxonomía’ suele referirse a jerarquías con múltiples ‘padres’. También se utiliza a veces para referirse a redes con más de un tipo de vínculo” (Uschold y Grüninger 1996).

estructurar. Una ontología se construye con el objetivo de compartir y reutilizar la información almacenada que, tras ser formalizada, puede ser interpretada tanto por personas como por programas informáticos.

- Para Chandrasekaran y colegas (1999), “[...] una ontología es una teoría de contenido sobre las clases, propiedades y relaciones de los objetos que son posibles en un determinado dominio de conocimiento y se explican mediante un lenguaje de representación”.
- Wand y Weber (2004) establecieron: “[...] aunque la mayoría de las ontologías son una taxonomía más que una teoría, siguen teniendo capacidad predictiva y explicativa. [También añadieron que] si un fenómeno se clasifica correctamente según una teoría, las personas pueden comprender y predecir mejor ese fenómeno y, en consecuencia, trabajar más eficazmente con él”.

Muchas de las ciencias de la información –especialmente la informática y la inteligencia artificial– han adoptado el uso de versiones modernas de “ontología”. En el ámbito de estas ciencias, “lo que existe es lo que puede ser representado”. Por lo tanto, bajo este enfoque disciplinario, “[...] la ontología es un artefacto de ingeniería, constituido por un vocabulario específico utilizado para describir una determinada realidad, más un conjunto de supuestos explícitos sobre el significado previsto de las palabras de ese vocabulario” (Guarino 1998).

Se puede resumir de forma breve y sencilla que dentro de las Tecnologías de Información y Comunicaciones (TIC), la idea de “descripción formal” o “especificación formal” significa esencialmente que lo que se describe de esa forma puede ser eventualmente programable en un sistema informático. De ahí que en estas disciplinas el término *ontología* sea básicamente una formulación programable de un vocabulario específico para describir el conocimiento común sobre una determinada realidad de un dominio dado. Debe tenerse en cuenta también que en estas disciplinas, el

término *artefacto* se utiliza para nombrar los conceptos y el vocabulario específico, y *dominio* para nombrar una determinada área o campo de conocimiento que se está definiendo. Más concretamente, el *artefacto* suele consistir en una estructura conceptual completa y exhaustiva dentro de un campo de conocimiento o *dominio*, y se construye para facilitar la comunicación y el intercambio de información entre usuarios, bases de datos y sistemas informáticos. Su construcción y uso pretende ir más allá de una organización o región; pretende llegar a toda una comunidad interesada en ese campo de conocimiento. En este trabajo no se pretende llegar a descripciones formales hasta el nivel de describir elementos de tal manera que puedan ser programables en un sistema informático; al menos, no de forma inmediata.

De las numerosas definiciones existentes, puede deducirse que los conceptos de ontología –en su forma moderna– no divergen sustancialmente entre sí, sino que existen diferentes “énfasis” o “enfoques” dependiendo tanto del planteamiento de cada disciplina, como del propósito principal de cada ontología. Sin embargo, esto tiende a producir muchas definiciones específicas que confunden a quienes las estudian. Green y Rosemann (2004, i) afirmaron que “[...] no solo el tipo de trabajo de investigación que se realiza bajo el término paraguas ‘ontología’ varía significativamente, sino que también la comprensión de una determinada ontología varía sensiblemente en términos de alcance, detalle y propósito”.

En este trabajo, se descartan los demás enfoques disciplinarios diversos, y por lo mismo sólo se considerará aquí el enfoque de la Archivística. Bajo este interés disciplinar particular y para el alcance y propósitos de este trabajo:

[...] las ontologías son conceptualizaciones construidas para describir formalmente –con un vocabulario de términos, atributos y relaciones– el conocimiento común y consensuado relacionado con un determinado tema, de manera que dicho conocimiento pueda ser comprendido, compartido y reutilizado principalmente por diferentes grupos de personas y eventualmente por aplicaciones informáticas (Green y Roseman 2004, i-ii).

Las ‘conceptualizaciones’ son representaciones formales de un cuerpo de conocimiento que comprende objetos, conceptos y otras entidades que se supone que existen en algún ámbito de interés, así como las relaciones existentes entre ellos; son visiones abstractas simplificadas del mundo que se pretende representar con algún propósito. Todo cuerpo de conocimiento o sistema basado en él se compromete explícita o implícitamente con alguna conceptualización en un determinado campo de interés o ‘dominio’ en el que existen entidades u ‘objetos’ que deben describirse de manera explícita, formal, coherente, consensuada y semántica. El vocabulario de representación proporciona un conjunto de términos para describir los objetos y sus relaciones en un campo de interés delimitado. ‘Formalmente’ significa que la ontología se describe en un lenguaje tal que puede ser procesado por un computador o, al menos, que es factible transferirlo a un programa.

Más allá de las definiciones y para intentar comprender mejor las ontologías, puede extraerse de todas las definiciones mencionadas una ‘caracterización’; es decir, una lista de todos los elementos característicos relevantes y/o comunes que las conforman. Por tanto, las ontologías:

- Pretenden explicitar los supuestos, puntos de vista, estructura y contenido acerca de un tema o campo específico.
- Describen algunos conocimientos acumulados y compartidos por una comunidad sobre un tema o campo.
- Definen los conceptos, las relaciones y otras características importantes que modelan un determinado tema o campo.
- Permiten compartir conocimientos sobre un tema común entre personas, entre computadores y, eventualmente, entre ambos.
- Pueden tener diferentes niveles de generalidad o especificidad. Dado que ninguna ontología es capaz de lograr una descripción completa de un cierto tema, un conjunto de ontologías permite construir de lo general a lo particular.
- Proporcionan un vocabulario para un determinado campo a través de un conjunto de términos.

- Permiten la creación de nuevos términos a partir de la reutilización de términos antiguos o de su combinación, y permiten acuñar nuevos conceptos para términos antiguos.
- Permiten la reutilización de las descripciones formales de sus términos en otras entidades.
- Deben especificar o describir de manera formal el significado o la semántica de sus términos, establecer sus restricciones y desambiguar su interpretación.
- Permiten el “mapeo”; es decir, especificar equivalencias, interrelaciones y combinaciones entre términos dentro de la misma ontología o entre otras ontologías.
- Permiten añadir reglas o condiciones requeridas por los conceptos (identidad, rigidez, unidad, dependencia, obligatoriedad, etcétera).
- No son únicas ni absolutas; puede haber múltiples representaciones de un mismo concepto, ya que éste puede representarse de muchas maneras y con distintos fines.
- Algunas ontologías fomentan la construcción y el uso de un conjunto de términos para representar hechos dentro de una instancia específica de un dominio, mientras que otras se construyen haciendo hincapié en el conjunto de hechos que se comparten.

Además de los diversos enfoques posibles derivados del uso o la disciplina, Souza y colegas (2011) establecieron diferentes niveles o profundidades de “precisión ontológica”; de la más simple hasta la más compleja se distinguen:

- Léxico o catálogo (vocabulario con definiciones en lenguaje natural).
- Taxonomía simple (vocabulario, diccionarios de datos).
- Taxonomía compleja (vocabulario, clasificación, diccionarios de datos, jerarquías).
- Tesauro (taxonomía con términos relacionados).

El concepto de documento...

- Modelo relacional/esquema (restricciones de tipo y relaciones arbitrarias entre entidades).
- Teoría axiomática completa.

Otros autores también hacen distinciones entre los niveles o la profundidad de las ontologías. Suelen denominar “ontologías ligeras” a las que solo cubren las etapas de vocabulario, clasificación o tesauro, mientras que denominan “ontologías complejas” a las que incluyen axiomas, restricciones, etcétera. Milton (1998) distingue entre las ontologías centradas en la teoría y las orientadas a la práctica. Para él, las primeras son las que se crean a partir de alguna teoría científica, humanística o social y hacen hincapié en ella, mientras que las ontologías orientadas pragmáticamente son las que provienen de la práctica consensuada de una disciplina y se construyen normalmente teniendo en cuenta que serán utilizadas por sistemas informáticos. Estas últimas son, por tanto, habituales en la práctica de las ciencias de la información y están dirigidas a ámbitos prácticos específicos, como en el caso de la gestión de documentos de archivo.

Dependiendo de la disciplina, el enfoque y el propósito, las ontologías pueden ser de muchos tipos. Al respecto existen varias propuestas de diferentes autores; por ejemplo, Steve *et al.* (1998) plantearon esta tipología:

- Ontologías genéricas: aquellas que representan teorías amplias o conceptos generales de un campo fundamental del conocimiento.
- Ontologías de dominio: aquellas que representan el conocimiento especializado y relevante sobre un determinado dominio o subdominio, como Medicina, Obstetricia, Bibliotecología o Archivística.
- Ontologías de representación: aquellas que especifican las entidades con representación general subyacente en formalismos de representación del conocimiento, sin definir lo que se representa; también llamadas meta-ontologías u ontologías de categorías de meta-nivel.

- Ontologías de tarea o de método: las creadas para una actividad, procedimiento, tarea o método específico.

Entre las numerosas variantes de las ontologías, se pueden distinguir algunos componentes que se utilizan en ellas con frecuencia; hay mucha variación entre los diferentes autores y propósitos; se puede enunciar una lista resultante:

- Clases: conjuntos, colecciones, tipos o categorías de objetos o conceptos, ya sean abstractos o concretos.
- Atributos: aspectos, propiedades, rasgos o características que pueden tener las clases y los objetos.
- Relaciones: tipos de interacción entre objetos o conceptos (*consiste en, es subclase de, está conectado a*, etcétera).
- Reglas: proposiciones de la forma “si-entonces” (antecedente-consecuente) que describen las inferencias lógicas que deben extraerse de un enunciado expresado de una cierta manera determinada.
- Restricciones: descripciones formales de lo que debe ser cierto para que un enunciado sea aceptado como válido.
- Funciones: Cálculos u operaciones realizadas y añadidas a un objeto o clase.
- Cardinalidad: El número de relaciones *mínimas, máximas o exactas* de un cierto elemento con otros entes.
- Axiomas: afirmaciones complejas y sus reglas presentadas en forma lógica que, en conjunto, comprenden la teoría global que la ontología describe en su dominio de aplicación.
- Instancias: cada miembro individual de una clase.

La importancia actual de las ontologías radica en que todo conocimiento debe expresarse en conceptualizaciones que inevitablemente requieren un vocabulario específico para un determinado campo. Las ontologías proporcionan una forma de facilitar estas conceptualizaciones y el vocabulario específico asociado, al tiempo que permiten explicitar los supuestos y la estructura del tema a definir. Facilitan el análisis del conocimiento, aportan orden y

estructura y aumentan la reutilización. Su importancia futura radica en que pueden ayudar eficazmente a conseguir a más corto plazo que la información se comparta entre individuos, entre sistemas y entre ambos de forma unificada y coherente, lo que a la postre redundará en una recuperación más precisa, pertinente y oportuna de dicha información. Al igual que los tesauros, las ontologías deben construirse, una por una, para un determinado campo o área de conocimiento; es decir, un 'dominio': Informática, Lingüística, Microbiología, Astronomía, Bibliotecología, Archivística, etcétera. Como cada uno de estos campos es muy amplio por sí mismo, se requerirán ontologías cada vez más detalladas para subcampos o áreas específicas del conocimiento. Se trata de un trabajo profundamente intelectual, no delegable a las máquinas –al menos no por ahora–. Por tanto, queda por construir todo un conjunto de ontologías, particularmente para la ciencia archivística, y es deseable que éstas sean construidas por archivistas profesionales. Gran parte del problema es que no sólo se necesita recopilar y sistematizar el conocimiento común ya acumulado sobre esta área, sino también integrar los nuevos conocimientos, a medida que surjan nuevas teorías de la investigación, así como los cambios y características emergentes impuestas por los adelantos tecnológicos; es decir, especialmente por el desarrollo e incremento de los documentos de archivo digitales.

Sección 2

La ontología propuesta

Bajo el actual campo de interés particular –la Ciencia Archivística– y para el alcance y los propósitos de este trabajo, la ontología desarrollada y propuesta es una conceptualización personal construida para describir formalmente –con un vocabulario de términos, atributos y relaciones– el conocimiento común y consensuado relacionado con el tema “documento de archivo digital”, de manera que dicho saber pueda ser entendido, compartido y reutilizado, de inicio por diferentes grupos de personas, y eventualmente por aplicaciones informáticas. De todo lo anteriormente estudiado sobre las ontologías, se puede afirmar que para este caso particular, la ontología propuesta:

- Es una *ontología de dominio*, pues representa el conocimiento especializado y relevante compartido sobre el dominio “documento de archivo digital”, sus componentes, términos y relaciones.
- Reúne muchas definiciones de los conceptos, relaciones y otras características importantes que conforman el concepto de “documento de archivo digital”, haciendo explícitas los supuestos, puntos de vista, estructura y contenido sobre este tema.

El concepto de documento...

- Es una *ontología ligera*, ya que pretende simplemente definir los términos básicos y las relaciones que componen el vocabulario relacionado con el “documento de archivo digital”, así como las reglas de combinación de términos y relaciones útiles para ampliar eventualmente dicho vocabulario.
- Tiene una *precisión ontológica* a nivel de taxonomía compleja, ya que sólo cubre las etapas de vocabulario, clasificación, diccionario de datos y jerarquías; reúne un vocabulario compartido para el campo de interés a través de un conjunto común de términos.
- Tiene dos dimensiones en su interior, ya que analiza el concepto de “documento de archivo digital” a partir de sus componentes intelectuales y de sus componentes digitales, y realiza el mapeo (equivalencias, interrelaciones y combinaciones) de los términos entre sí.
- No tiene la intención de llegar a descripciones formales hasta el nivel de describir elementos de manera que puedan ser inmediatamente programables en un sistema informático.
- No es única ni absoluta; puede haber múltiples representaciones de un mismo concepto, ya que éste puede representarse de muchas maneras y con distintos fines.
- Permite la creación de nuevos términos a partir de la reutilización de términos antiguos o de su combinación y permite la acuñación de nuevas acepciones para términos antiguos.
- Permite la reutilización de las descripciones formales de sus términos en otras entidades.

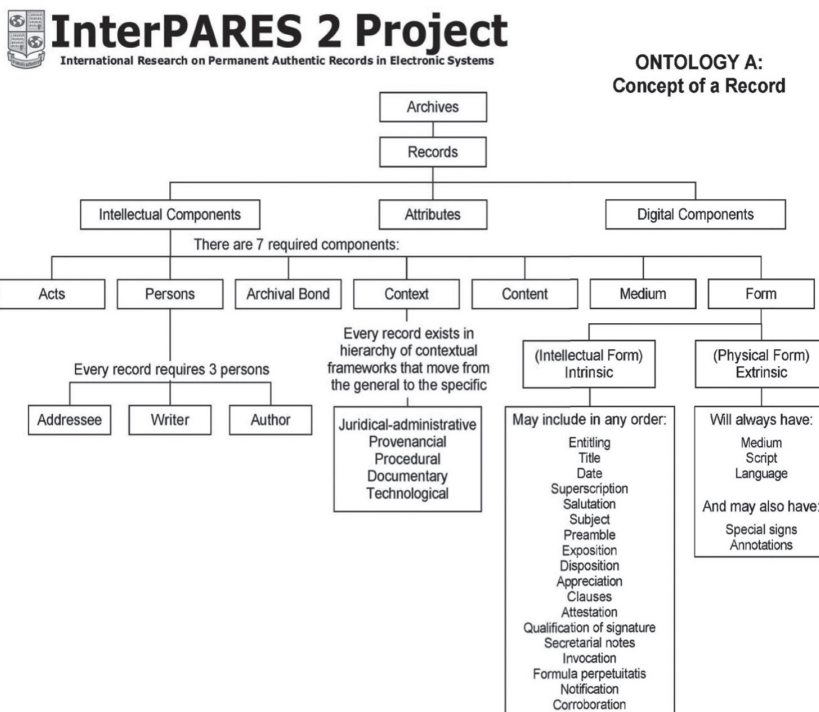
El Proyecto InterPARES (Investigación Internacional sobre Documentos de Archivo Auténticos Permanentes en Sistemas Electrónicos) o The International Research on Permanent Authentic Records in Electronic Systems lleva años estudiando los documentos de archivo digitales tratando de establecer –entre otras cosas– sus características, componentes y relaciones. Asimismo, este proyecto elaboró un glosario con terminología acerca de la preservación de los documentos de archivo digitales, con especial énfasis en las características y conceptos “modernos” derivados del desarrollo y

uso de los documentos de archivo digitales. Explorando más allá de un glosario, en 2008 el proyecto comenzó a introducir algunas ontologías básicas en relación con el conocimiento de los documentos de archivo. Esas ontologías eran muy básicas y estaban en el nivel de “taxonomías simples” antes mencionado. El conocimiento compilado entonces ha evolucionado a lo largo de los años, ya que la investigación va introduciendo nuevas ideas y refina los conceptos; por esta razón, las ontologías no pueden ser estáticas. En esta sección se presenta una propuesta personal actualizada y complementada de la original “Ontología A”, la cual analizaba el “concepto de documento de archivo”, así como sus componentes y relaciones. Esta idea comenzó

[...] siguiendo la clásica suposición de la Diplomática de que –independientemente de las diferencias de naturaleza, procedencia o fecha– todos los documentos de archivo son lo suficientemente similares como para que sea posible concebir una forma documental típica e ideal que contenga todos los elementos posibles de un documento de archivo, los investigadores plantearon la hipótesis de que, si bien pueden manifestarse de formas diferentes, los mismos elementos que están presentes en los documentos de archivo tradicionales existen explícita o implícitamente en los documentos de archivo electrónicos, y que todos los documentos de archivo electrónicos comparten los mismos elementos (Duranti 2003, 10).

Éste es el diagrama en su versión simple original:

Imagen 1



Fuente: Terminology Cross-domain Task Force. Appendix 22: InterPARES 2 Project Ontologies. En: *InterPARES 2: Experiential, Interactive and Dynamic Records*. Luciana Duranti y Randy Preston (Eds.). 2008, pp. 729-731. http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_appendix_22.pdf.

No obstante lo sencilla que era esta primera forma, al observarla pueden deducirse algunos hechos interesantes: Por supuesto, conjuntos de documentos de archivo digitales conforman archivos digitales. Un documento de archivo estaba formado por entidades denominadas componentes intelectuales, atributos y componentes digitales. En esa versión, los componentes digitales solo se introdujeron como una idea global, pero no fueron desmenuzados o detallados. Siguiendo la construcción y en relación con la visión

intelectual del documento de archivo, los componentes principales estaban ahí: cada documento de archivo tenía en la versión de entonces siete componentes intelectuales, y todos ellos debían existir siempre para cada documento de archivo. Desde el punto de vista de la Diplomática, cada documento de archivo requiere individualmente de cinco personas durante su producción: autor, escritor, destinatario, emisor y productor. Los tres primeros se convierten en componentes intelectuales del documento de archivo, y los otros dos en atributos. Se establecieron entonces cinco “contextos” o marcos referenciales para cada documento de archivo. También se estableció que debía haber un contenido –debe ser estable– y una forma documental –debe ser fija–; esta forma tenía dos manifestaciones: la forma intelectual o intrínseca, y la forma física o extrínseca. En análisis posteriores de este diagrama se han hecho más inferencias. A pesar de su simplicidad, esta tabla permitía comprender mejor el concepto de documento de archivo que una simple definición. Los principales objetivos de la nueva versión propuesta con respecto a la original son 1) analizar la evolución de los conceptos mostrados desde que se publicaron en 2008, 2) completarlos en la medida de lo posible y 3) desarrollar la parte que faltaba en el original; es decir, los componentes digitales. Todo ello desde un punto de vista personal.

Para llevar a cabo esta tarea, se consultaron los principales glosarios sobre documentos de archivo digitales, así como numerosos libros y artículos sobre conceptos y términos relacionados con el fin de extraer, analizar y comparar diferentes visiones de conceptos y términos. Se utilizó el enfoque de análisis ontológico con el fin de estudiar, comprender y analizar los contenidos como tales, independientemente de su forma de representación. Por lo tanto, se prestó especial atención a cada concepto encontrado tratando de establecer la identidad, las partes, los conjuntos, la composición, las relaciones y las propiedades. Tratando de no sobrecargar los conceptos con demasiados elementos, solo se incluyeron en el gráfico las características básicas. La nueva versión se elaboró específicamente para los documentos de archivo digitales, con dos dimensiones principales, tratando de establecer los componentes

tanto intelectuales como digitales de un documento de archivo digital y el posible mapeo entre ambos. Así, un documento de archivo digital puede visualizarse tanto a partir de los elementos intelectuales, como de los elementos digitales que lo componen. A los *componentes intelectuales* también se les conoce como *componentes conceptuales*, pero en este contexto de una ontología, donde prácticamente todo es conceptual, este término podría dar lugar a algunas ambigüedades y confusiones; por lo tanto, en este documento se prefirió utilizar el término *componentes intelectuales*.

Para empezar, la información recopilada se dibujó en forma gráfica mediante un mapa mental, uno de los cuadros más sencillos para ordenar el conocimiento. A medida que los conceptos fluyeron, se añadieron nuevos elementos y se diseñaron diferentes y más complejos esquemas de gráficos. Tras muchas reordenaciones, el mapa actual está formado por 165 partes o piezas; se consiguió una versión estable. El mapa evolucionó gradualmente desde su forma más simple hasta la actual con el fin de prepararlo para darle forma de ontología con los mínimos cambios posibles en el diseño y la forma; por ejemplo, las partes consideradas como *clases* principales se representaron con un óvalo; a las otras *clases* secundarias se les dio forma de caja cuadrada y a los elementos se les dio forma de rectángulo. Se seleccionó la notación UML (Unified Modeling Language o Lenguaje Unificado de Modelado). Como resultado, se obtuvieron versiones estables de un mapa mental y una ontología del “concepto de documento de archivo digital”. En este documento, se presenta la “versión para imprimir” de la ontología, numerada como “versión 2”:

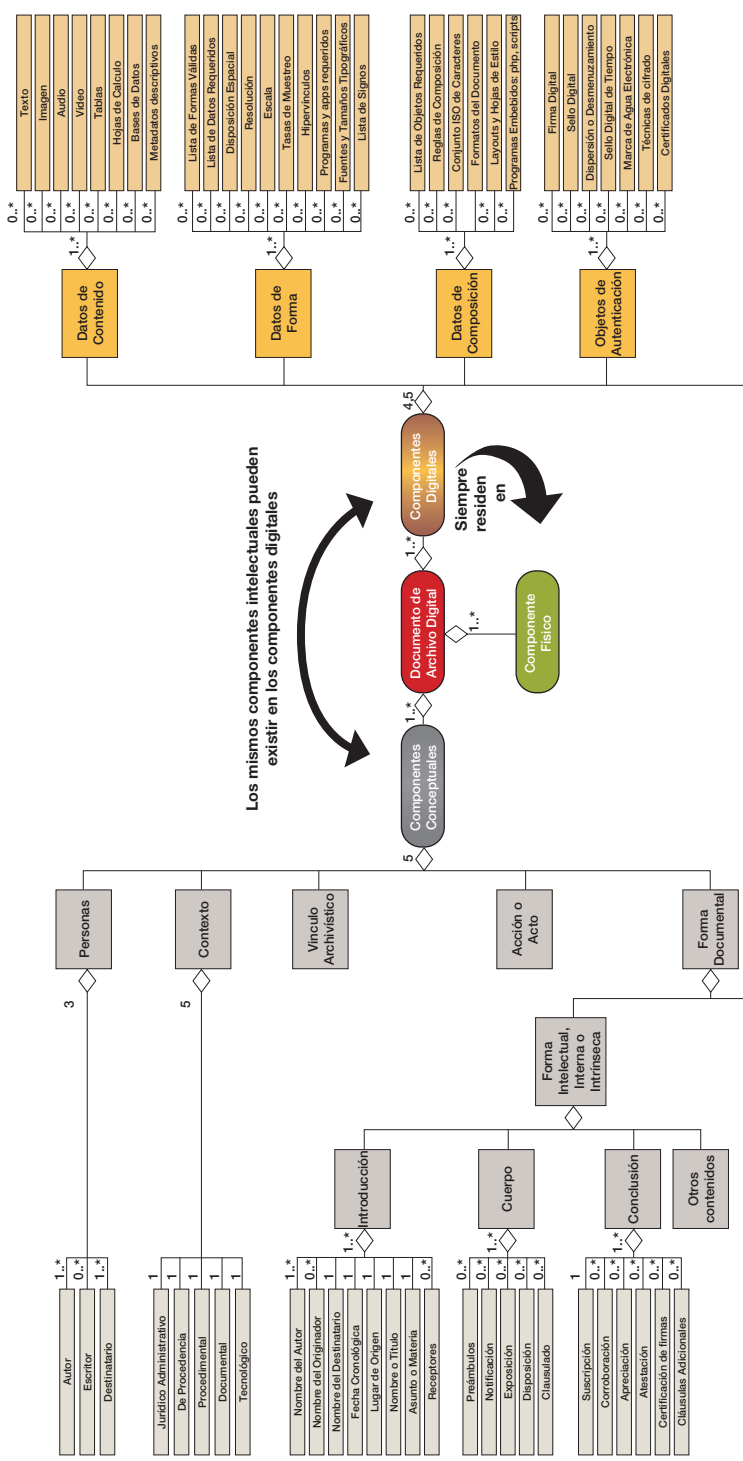
La parte izquierda del gráfico representa la visión de los componentes intelectuales del documento de archivo digital, que son las partes conceptuales y de agrupación de los elementos existentes en una forma ideal de documento de archivo. Desde el punto de vista de la Diplomática, se trata de una idealización conceptual de los rasgos o características definitorias que comprenden un documento de archivo, lo identifican de forma única y lo distinguen de otros; una determinada combinación de expresiones de

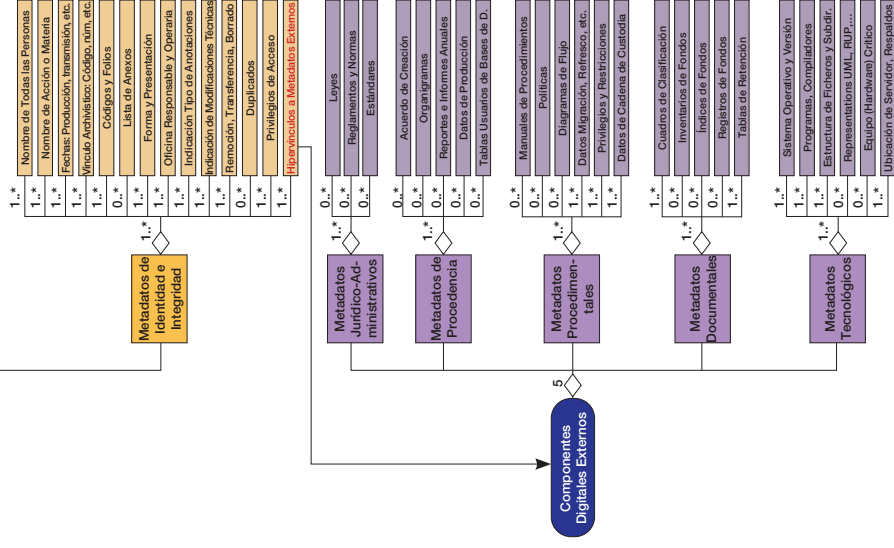
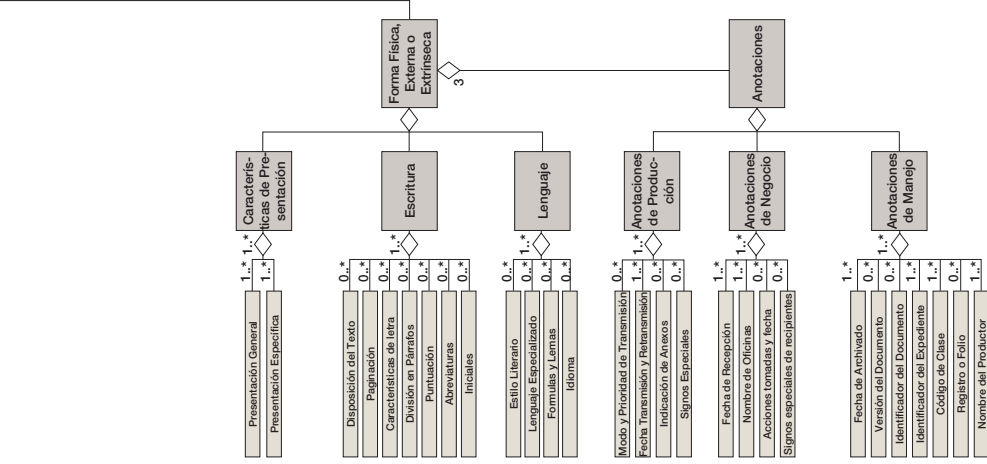
componentes intelectuales de un documento de archivo es única e irrepetible para cualquier otro.

La parte derecha del gráfico representa los componentes digitales del documento de archivo digital, el otro enfoque para entender su concepto. Los componentes intelectuales mencionados previamente son conceptos que, idealmente, dan forma a un documento de archivo y se manifiestan en diferentes elementos conceptuales que se encuentran en diversas partes del documento. Eventualmente, todos estos elementos conceptuales deben estar contenidos o expresados en un conjunto de componentes digitales; es decir, objetos digitales que no son un documento de archivo por sí mismos, sino que forman parte de un documento de archivo o de un conjunto de documentos de archivo digitales y que en conjunto contienen los datos primarios del documento de archivo, así como los metadatos y elementos adicionales necesarios para ordenar, estructurar o manifestar su contenido, todo ello contenido en ficheros informáticos. Esto es importante: muchas personas tienden a pensar que todo documento de archivo digital está auto-contenido en un único fichero informático que comprende todo el documento en un sistema informático. Ciertamente, es habitual que haya texto dentro de un documento de archivo, y por esta razón casi siempre se conceptualiza así, pero en realidad un documento de archivo puede estar conformado además por otras entidades u objetos de información: puede tener gráficos, imágenes fijas y en movimiento, sonido, fuentes, programas incrustados, metadatos, firmas digitales, hojas de estilo, etcétera.

A veces, estos elementos están embebidos en el propio documento de archivo y, por lo tanto, se convierten en componentes digitales internos del mismo (cuadros amarillos), pero otras veces existen en diferentes ficheros informáticos fuera del documento (cuadros morados), que se convierten en componentes digitales externos del mismo (Duranti y Thibodeau 2006, 29). Normalmente, estos últimos contienen una parte importante de los contextos del expediente digital. Suelen ser grandes conjuntos de información, ya que cada contexto suele generar todo un conjunto de metadatos de gran volumen; por lo tanto, no se almacenan dentro

Imagen 2. Ontología del “concepto de documento de archivo digital”





Fuente: Juan Voutsas 2017.

de cada documento de archivo individualmente. Se construyen de manera global para toda una serie o fondo y se guardan en ficheros externos al documento de archivo; por supuesto, con el correspondiente hipervínculo desde cada documento de archivo primario a esos registros externos. No obstante, todos los componentes digitales juntos –internos y externos– conforman una única unidad lógica, un “documento de archivo digital”.

Aun siendo enfoques diferentes, tanto los componentes intelectuales como los digitales no son entidades aisladas entre sí: los componentes intelectuales –los conceptos– deben estar contenidos en uno o varios componentes digitales –los objetos–. En otras palabras, los componentes digitales contienen objetos de información que, a su vez, contienen elementos conceptuales y, en conjunto, conforman un documento de archivo digital; simplemente, la relación no es biunívoca;² es decir, no es uno a uno. Cada componente intelectual no necesariamente se corresponde con un componente digital y viceversa. Evidentemente, no todos los posibles componentes digitales existen obligatoriamente en todos y cada uno de los documentos de archivo, pero la lista es bastante exhaustiva para tratar de abarcar todo lo que podría encontrarse como componente digital dentro del universo de estos documentos de archivo. Es esencial comprender plenamente esos componentes para evitar asumir que al preservar un fichero informático que contiene el cuerpo principal de un documento de archivo, se preservan todos sus componentes digitales. Por ejemplo, es un error común omitir la inclusión de los conjuntos de fuentes tipográficas (Arial, Courier, etcétera) al preservar un fondo de archivo digital, ya que normalmente los directorios del computador que contienen las fuentes están en el directorio de su sistema operativo, y no en el directorio donde se guardan los documentos de archivo digitales. Éste es un ejemplo de un error típico que se suele

2 Biunívoca: correspondencia que asocia cada uno de los elementos de un primer conjunto con uno, y solo uno, de los elementos de un segundo conjunto, y cada elemento de éste con uno, y sólo uno, del primer conjunto.

cometer por ignorancia al identificar todos los componentes digitales de un documento de archivo.

Existe también en la gráfica un tercer componente principal –el componente físico– que en la ontología original se consideraba como uno de los componentes intelectuales, pero que actualmente se considera por separado como un componente especial, ya que en la realidad no es ni un componente intelectual ni digital. Es el medio o soporte físico en el que se guardan o almacenan los componentes digitales de un documento de archivo. Implícitamente, el medio no forma realmente parte de ese documento, pero los componentes digitales no constituyen un documento de archivo hasta que se fijan en un soporte físico y estable, y la destrucción del soporte puede implicar la destrucción del documento, por lo que se necesita una clase especial para él.

Los dos tipos de componentes principales, los intelectuales y los digitales, incluyeron en lo posible las cuatro partes constitutivas de un documento de archivo consignadas en las guías del foro DLM y asumidas en el modelo Moreq2: contenido, estructura, contexto y presentación (DLM Forum 1997, 13; European Commission 2008, 18).

La “versión para imprimir” de la ontología aquí presentada es una versión breve: solo muestra las partes de identidad, composición y relación de todos los componentes. Existen además versiones interactivas completas del mapa mental y de las ontologías construidas en html con toda la información compilada sobre cada elemento, con una particularidad inherente: todas y cada una de las 165 casillas tienen una definición resultante para el término y un buen número de ellas tienen una explicación adicional de características, relaciones, mapeo, etcétera. Esta información en las versiones html se despliega en la pantalla con solo posicionar el puntero del ratón sobre cualquier casilla. Muchas de ellas enlazan también con los textos completos de un conjunto de documentos con explicaciones adicionales sobre sus conceptos. El total de la bibliografía hipervinculada de este modo con el gráfico asciende a 21 documentos. Por lo tanto, el mapa completo y la ontología con definiciones, relaciones, mapeo, etcétera se despliegan mejor y se aprecian plenamente en las versiones html. Ambas versiones, la

completa y para imprimir, están disponibles en inglés y en español en el sitio web del Instituto de Investigaciones Bibliotecológicas y de la Información (IIBI) de la Universidad Nacional Autónoma de México (UNAM): <http://iibi.unam.mx/archivistica> utilizando el botón “ontologías”.

Los gráficos intentan reflejar una visión personal de los cambios y la evolución encontrados en la literatura sobre el concepto de documento de archivo digital y sus elementos relacionados; por ejemplo: los siete componentes intelectuales originales se han reducido a cinco; la disposición de algunos elementos ha cambiado de ubicación; la separación del medio o soporte físico, es decir, el componente físico, etcétera.

La traducción de términos del inglés a otros idiomas suele introducir otra gran fuente de confusión de conceptos como consecuencia de las traducciones ambiguas desde y hacia esos idiomas, ya que los términos archivísticos no pueden traducirse simplemente a partir de los glosarios habituales. Se han realizado muchos esfuerzos en este sentido, tratando de establecer las equivalencias correctas entre los términos en varios idiomas; por ejemplo, la base de datos terminológica internacional del Consejo Internacional de Archivos (ICA), la base de datos terminológica multilingüe InterPARES, etcétera. Para esta propuesta, se construyeron versiones tanto en español como en inglés del mapa y de la ontología; un botón en las versiones html conmuta toda la página de español a inglés y viceversa. Se procuró una cuidadosa traducción de términos, definiciones, etcétera, al español a partir de las definiciones originales, utilizando las bases de datos terminológicas archivísticas actualizadas (ICA e InterPARES) en esos idiomas, tratando de resolver los problemas de ambigüedad de los términos. Cuando los documentos de los hipervínculos de la ontología hacia textos completos existían en versiones en español, los hipervínculos pasaban también a ellos. Así, el mapa y la ontología son cuidadosamente bilingües. Esto ayudará en caso de una eventual programación informática de la ontología con fines de traducción correcta de los términos entre estos dos idiomas y, obviamente, se podrían hacer fácilmente intentos similares con otros idiomas.

Se ha mencionado que esta “versión 2” es una versión estable de una ontología sobre este tema, pero no se considera en absoluto definitiva ni cerrada: es susceptible de ser perfeccionada debido a las dos siguientes razones: 1) se trata de una interpretación personal de las múltiples disposiciones posibles de los elementos que requieren de un mayor debate. 2) Como mapa, puede considerarse muy completo; como ontología, aún carece de algunos elementos (restricciones, relaciones, etcétera) necesarios para estar totalmente completa.

Como primera conclusión del análisis de los componentes intelectuales y digitales, se puede afirmar que estos últimos contienen en su interior objetos de información que albergan a los primeros y que, en conjunto, conforman un documento de archivo digital, no necesariamente en una relación uno a uno, y normalmente dispuestos de forma diferente en cada fondo. Esto es de capital importancia a la hora de diseñar un documento de archivo digital de un determinado fondo, o a la hora de evaluar si un determinado sistema de gestión documental puede producir y/o administrar un documento de archivo de una determinada organización. El diseñador o evaluador debe ser capaz de identificar y establecer *todos y cada uno* de los componentes conceptuales del documento de archivo y, a continuación, ser capaz de realizar el mapeo con todos los componentes digitales relevantes que los contendrán, con el fin de evitar omisiones, ambigüedades, confusiones y otros errores mientras se producen, utilizan o conservan dichos documentos de archivo. Obviamente, este mapeo puede realizarse de forma inversa: el análisis de todos los componentes digitales de un documento de archivo debe ser capaz de establecer una correspondencia o mapeo con todos sus componentes intelectuales.

Asimismo, puede establecerse que, aunque no es realmente una parte del documento de archivo, debe existir un componente físico junto con los demás; es decir, el soporte informático o medio físico sobre el que se guardan o almacenan los componentes digitales de ese documento. El componente físico debe existir necesariamente ya que la definición de documento ha establecido que su esencia es “la información registrada en un soporte”. Así,

un documento no puede existir si no ha sido “fijado” en un soporte físico, ya sea tradicional o digital. Por lo tanto, si un documento de archivo es un tipo especial de documento, requiere estar fijado en un soporte físico. El hecho de ser físico implica también que se requiere algún tipo de acción de conservación.

Como ha podido verse, la expresión de un conjunto de ideas en esta forma de representación permite observar exhaustivamente niveles, relaciones, subcomponentes, jerarquías, ascendientes, descendientes, condiciones, etcétera, sobre el concepto analizado, lo que lo esclarece y permite una mejor comprensión integral. Al tener esta representación a la cual se han añadido también las definiciones de los términos adyacentes o embebidos, se consigue una representación estructural más completa del concepto de “documento de archivo digital”. A través de este tipo de representaciones, se comparte el conocimiento común sobre este dominio o tema de forma más completa que las simples definiciones del concepto introducidas al principio de este documento, las cuales por supuesto están contenidas en esta representación.

Un análisis más detallado permite hacer surgir cada vez más información, contexto y respuestas según lo estudiado. En un análisis superficial es fácil entremezclar “intelectual” con “digital”, “componentes” con “elementos” de un documento de archivo como sinónimos o –sabiendo que no lo son– describir la diferencia puede convertirse en una tarea difícil. Gracias a la ontología se puede ver que, siendo similares, no son lo mismo, son entidades distintas; puede verse por qué son diferentes, y por tanto se hacen más fáciles de entender. En la ontología hay definiciones precisas extraídas de múltiples textos, pero lo más importante es su contexto y sus interrelaciones. Más allá de la utilidad de este tipo de representaciones para la comprensión de los términos, se observa que pueden ayudar en gran medida a establecer cuestiones a tener en cuenta al diseñar o adquirir, por ejemplo, un sistema de gestión de documentos de archivo, o al establecer los atributos de un determinado tipo de documento digital dentro de una organización, así como de todos y cada uno de sus componentes asociados. Con la ontología se pueden establecer requisitos más claros para un

correcto proceso, gestión, clasificación, etcétera, así como la identificación de todos los componentes digitales para salvaguardar su futura integridad, identidad, exactitud o completitud. Esto facilitará su manejo y posterior preservación a largo plazo, y al mismo tiempo mantendrá su autenticidad.

Si bien las técnicas semánticas proporcionan la metodología para hacer mejores representaciones, será fundamental que las comunidades archivísticas se encarguen de formalizar la descripción de estos atributos relevantes en una representación adecuada de los documentos de archivo, ya que son ellas quienes tienen la experiencia y el conocimiento sobre el tema. Por ello es importante que los archivistas profesionales conozcan mejor dichas técnicas. Bajo los supuestos anteriores, las características fundamentales de los documentos de archivo digitales deben ser identificadas y representadas adecuadamente. Ciertamente, esto se hace hoy en día, pero no a menudo bajo este enfoque “semántico”.

Los sistemas actuales de gestión archivística están contruidos casi en su totalidad sobre bases de datos con estructuras “relacionales” para el almacenamiento y la recuperación de los documentos que contienen. Típicamente, las búsquedas se hacen por medio de “palabras clave” que recuperan los documentos vinculados a ellas, pero la casi totalidad de los datos entre documentos están desasociados unos de otros, y poco pueden asociarse más allá de los clásicos operadores booleanos. Por lo mismo, estos sistemas tienen una comprensión muy limitada del significado de los datos que ellos guardan. De forma semejante, en la Web actual, los buscadores realizan la mayor parte de la tarea de recuperación intentando localizar las palabras solicitadas dentro de las páginas Web, encontrando palabras sueltas aquí y allá en muchas de ellas. Al disponer de miles de millones de sitios web, esos motores de búsqueda recuperan un número incontable de páginas de las que la gran mayoría son irrelevantes, por lo que el acceso real a la información significativa se convierte en una tarea difícil y frustrante. Dado que la recuperación en la Web se basa principalmente en palabras prácticamente sin contexto, lo recuperado es poco con respecto al gran total. El ingrediente primordial que falta tanto en

los sistemas de gestión archivística como en los buscadores Web es información con un mejor contexto predefinido, con semántica inherente. Esta información predefinida, formal, estandarizada y universal es precisamente la información construida con buenas “representaciones semánticas”. Gracias a ellas, podemos dotar de máxima significación a los datos, tanto a aquellos internos de los sistemas de gestión archivística, como a los de la propia Web, y así optimizar los buscadores utilizados en ellos. Los nuevos modelos conceptuales derivados de las teorías del Modelo de Datos Semántico permiten ahora definir el significado de los datos a través de sus relaciones con otros.

Éstas no son sólo reflexiones académicas. Debido a las nuevas tendencias mundiales de la *revolución de los datos*, los *datos masivos*, los *datos abiertos*, los *datos vinculados*, etcétera, se han estado desarrollando numerosos proyectos al respecto: nubes virtuales con grandes cantidades de datos en las cuales cualquier persona puede acceder a cualquier dato autorizado así como agregar nuevos, con lo que se conforma un entorno abierto, estructurado e interoperable que favorece que los datos puedan ser creados, interconectados y consumidos a escala global. Estos proyectos originalmente se asociaban a bibliotecas, repositorios y colecciones afines, pero cada vez más se acercan a los archivos. Ejemplos representativos de ello son los proyectos del Banco Mundial de Datos Abiertos (*World Bank of Open Data*), la Red semántica de datos Abiertos del Gobierno de Corea del Sur (*South Korea's Open Public Data Semantic Network*) o las recomendaciones del Grupo de Revisión de Datos Vinculados de Archivos de Investigación y Colecciones Especiales de OCLC (*OCLC Research Archives and Special Collections Linked Data Review Group*), por citar algunos. Todos ellos están en el campo de los archivos e incluyen representaciones documentales formales.

Por todo lo anterior, son necesarias mejores herramientas para una comprensión más profunda y rápida de los conceptos principales de la disciplina archivística. Lo ideal es una visión holística, entendiendo por ello que cada conjunto debe ser analizado como un todo, y no a través de las partes que lo componen consideradas

por separado. Es el conjunto como un todo integrado el que determina definitivamente cómo se comportan las partes, mientras que un mero análisis de éstas no puede explicar completamente el rendimiento del conjunto; es decir, el conjunto es más complejo que la simple suma de sus elementos constitutivos.

Pero eso es sólo el inicio. Como se ha establecido, utilizando representaciones semánticas y aprovechando los eventuales niveles de clase, subclase, elemento compuesto o simple, etcétera, dados en esquemas como el propuesto, una vez completada esta representación ontológica es factible de ser programada en un sistema informático en términos que puedan ser recuperados de manera más eficiente a través de un sistema de gestión de documentos de archivo y/o de la Web para un mejor acceso y recuperación de los mismos. Intentar establecer las relaciones, interdependencias, subcomponentes, etcétera, en un programa a partir de un simple glosario de definición de términos sería extremadamente difícil, si no imposible; programarlo todo partiendo de una ontología es mucho más sencillo. Y tiene otra gran ventaja: si la ontología es analizada por expertos en la materia, es relativamente fácil hacer equivalencias de ella en otros lenguajes. Las versiones completas en español e inglés ya realizadas son un ejercicio completo de esto. Lo cual significa que ambas versiones son compatibles para diferentes comunidades con un solo conjunto de conocimientos. Si las diferentes versiones en distintos idiomas son factibles de ser programadas en un sistema informático, o en dos o más sistemas informáticos interconectados, los motores de búsqueda podrían recuperar lo mismo, independientemente del idioma introducido, con la exacta equivalencia a los términos asignados y sin ambigüedad.

Esto es particularmente notorio hoy en día cuando se utiliza cualquier traductor o diccionario general en línea con algunas palabras específicas –por ejemplo, “*record*”– de las que no existe un equivalente exacto entre el inglés y otros idiomas. Formalmente, en muchos idiomas se ha tenido que acuñar un término compuesto específico; esto es, de más de una palabra, para ello: así, en español, portugués, italiano y francés, entre otros idiomas, se

ha acordado el término compuesto “documento de archivo”³ para marcar la diferencia con las definiciones de un simple “documento”. En esos idiomas, no existe una palabra única para representar la palabra “*record*” en su acepción archivística; por ello el término compuesto. Sin embargo, los traductores y diccionarios en línea actuales suelen traducir erróneamente este término del inglés a esas lenguas y viceversa. De forma semejante, el término “*creator*” tiene siempre una traducción automatizada al español “creador”, que no es correcta; no puede ser una equivalencia automática de diccionario. En Archivística, la equivalencia correcta de “*creator*” en español es “productor”, que en traducción inversa al inglés se volvería erróneamente “*producer*”. En portugués ocurre exactamente lo mismo: “*creator*” se traduce siempre en línea por “criador”, mientras que el término archivístico correcto en esa lengua es “produtor”, y así sucesivamente.

Con las ontologías no hay confusión entre los conceptos por un simple error del sistema de traducción, ya que la ontología al programarse establece los términos exactos que son equivalentes y los que no lo son. Como otro ejemplo, con la ontología nunca habrá una confusión sobre qué significa exactamente el término “título” desde el punto de vista de los documentos de archivo digitales, a pesar de lo que puedan decir todos los diccionarios generales al respecto. Las ontologías bilingües permiten, por un lado, entender los términos que se explican y contextualizan en su propia lengua, y por otro lado permiten que los motores de búsqueda o los sistemas internacionales tengan los equivalentes correctos en la segunda lengua, por supuesto, lo bilingüe puede convertirse en multilingüe. Una vez más, no se trata de un simple proceso de traducción de diccionario: es un proceso cognitivo, con un orden coherente, definición, experiencia, jerarquía y estructura. Los términos de la ontología no son meras traducciones literales

3 Inglés = “record”. Español = “documento de archivo”. Portugués = “documento arquivístico”. Italiano = “documento archivistico”. Francés = “document d’archiv”. InterPARES Terminology Database (2012). http://www.interpares.org/ip3/ip3_terminology_db.cfm?letter=r&term=41.

y, sobre todo, no pueden delegarse a un computador mientras se construyen. Además, en la archivística –como en muchas otras disciplinas–, se crean diariamente neologismos en muchas lenguas, y como todas las palabras nuevas o los nuevos significados de las existentes, deben ser debidamente ponderados antes de ser traducidos e introducidos con éxito en esa lengua en este proceso. La simple traducción, hecha al azar y sin semántica, conduce a la aparición de múltiples versiones de términos que luego serán difíciles de estandarizar. Las representaciones semánticas como las ontologías permiten tanto a los individuos como a los programas informáticos realizar traducciones correctas, y facilita establecer cuándo ciertos términos son equivalentes o sinónimos y cuándo no.

La ontología propuesta aquí es sólo un ejemplo. Al ser una representación arbitraria del conocimiento, no es necesariamente la única posible. Evidentemente, puede haber más de una ontología para representar un mismo conocimiento: depende de quién la haga, del enfoque que se le quiera dar, de la profundidad, el énfasis, el interés, el agrupamiento, el alcance, etcétera. No se trata de una representación ecuménica de una determinada realidad; es solo una de las posibles representaciones de dicha realidad bajo una determinada visión. De hecho, el énfasis o “modelo de aproximación” empieza a definir desde el inicio cada posible representación del conocimiento. Éste ha sido un ejemplo personal desarrollado a partir del mencionado enfoque original de InterPARES con algún análisis ontológico. Por supuesto no es el único posible.

Por lo tanto, una gran ventaja de las ontologías así desarrolladas es que cada comunidad puede construirlas según sus intereses, o puede tomar cualquiera de las existentes y adaptarla a su contexto y perfeccionarla a su conveniencia. Esta última opción, además de ahorrar mucho esfuerzo, ayudaría a que las ontologías ya creadas sobre archivística vayan normalizando el conocimiento a través de las comunidades de esta disciplina a la vez que se estabilizan los términos y conceptos, lo cual es especialmente útil en estos tiempos de cambio e innovación, y a largo plazo acabará facilitando el intercambio de datos e información digital en este campo. No es deseable ir a los extremos y tener “una y sólo una”

ontología para un concepto como definición universal, pero ciertamente es deseable intentar generalizar el uso de las existentes a través de las comunidades y organizaciones archivísticas con el propósito de estandarizar ciertos conceptos fundamentales, como ya se ha hecho en otras disciplinas.

Éste es sólo un ejemplo de lo que las representaciones semánticas pueden lograr con una mayor calidad de construcción. Al ser semánticamente ricos y utilizar metadatos adecuados y convenientes, los documentos de archivo pueden ser mejor definidos y enlazados con otros documentos similares para optimizar su búsqueda, integración, reutilización, preservación, minería de datos e interoperabilidad a través de mejores descripciones funcionales de sus contenidos, impulsando interacciones con aplicaciones externas para explotarlas adecuadamente. Así, esas aplicaciones podrían resolver de mejor forma problemas bien definidos a través de operaciones bien definidas a realizar sobre datos existentes bien definidos. Esto es especialmente útil en la actualidad, ya que muchas instituciones gubernamentales, de patrimonio cultural, organizaciones privadas, etcétera, que están produciendo enormes cantidades de documentos de archivo digitales están muy preocupadas por el uso óptimo y la preservación a largo plazo de sus documentos para el patrimonio cultural y científico; los fines legales o la competitividad económica.

Los principios y las recomendaciones de las propuestas semánticas aquí contenidas aún no están totalmente descritas ni terminadas; al contrario, acaban de comenzar su desarrollo y evolucionan cada día más. Esto es solo un ejemplo de lo que se puede hacer y por qué. Actualmente existen buenos fundamentos sobre los caminos y las acciones a seguir, pero la mayor parte queda por investigar, descubrir y aplicar. Determinar qué tipo de conocimiento contextual se requiere para el desarrollo de los documentos de archivo y digitales, sus programas, *hardware*, formatos, estándares, metadatos, etcétera, es un paso crucial que apenas comienza. Las decisiones que se tomen a corto plazo por parte de los responsables de la comunidad archivística serán de suma importancia en este sentido a largo plazo. Como puede deducirse, los

aspectos semánticos y pragmáticos de los materiales digitales no son solo un capricho académico, sino que son esenciales para que los usuarios puedan acceder, comprender y utilizar los documentos de archivo digitales en un futuro próximo. El *hardware*, los programas, los formatos y las redes pueden estar estrechamente interconectados con una mejor representación del conocimiento y con modelos de gestión más sofisticados para permitir a los usuarios un acceso más rápido, oportuno y relevante a la información, así como su correcta presentación.

En la medida en que podamos comprender y contribuir a estas nuevas representaciones, desarrollaremos conceptos, normas, reglas y tecnologías que ayuden a los sistemas a “entender” mejor la información digital para que puedan tratarla de una forma más conveniente y conseguir así mejores recuperaciones, y aún más: utilizando representaciones semánticas, no solo pueden conseguirse resultados más precisos en la búsqueda de información, sino también una mejor integración de la información procedente de diferentes fuentes, así como garantizar una preservación óptima de esa información para aspirar a todo un nuevo y moderno tipo de servicios automatizados relacionados con aquellas organizaciones con responsabilidades archivísticas.

Referencias bibliográficas:

(Todas las referencias electrónicas han sido verificadas como exactas y existentes al 24 de julio de 2019).

Arakawa, Naoya. 2008. "Semantic Analysis based on Ontologies with Semantic Web Standards". *Computer-aided Acquisition of Semantic Knowledge (CASK)*: 2. <http://www.celta.paris-sorbonne.fr/anasem/papers/Theory/Arakawa-CASK2008.pdf>.

Arano, Silvia. 2005. "Thesauruses and Ontologies", *Hipertext.net - Anuario Académico Sobre Documentación Digital y Comunicación Interactiva*, núm. 3: 3. <http://eprints.rclis.org/8972/2/12.pdf>.

Baader, Franz; Horrocks, Ian; Sattler, Ulrike. 2004. "Description logics". *Handbook on Ontologies*: 3-28.

Berners-Lee, Tim. 2001. Conceptual Architecture of the Semantic Web. W3C (World Wide Web Consortium) Website <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slides10-0.html>.

Borst, Willem. 1997. *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. Enschede, The Netherlands: W.N. Borst, (Ed.) <http://doc.utwente.nl/17864/1/t0000004.pdf>.

Chandrasekaran, Balakrishnan; Josephson, John; Benjamins, Richard. 1999. What are Ontologies? Why do we need them?". *IEEE Intelligent Systems*, vol. 14, num. 1: 20-26 <http://www.csee.umbc.edu/courses/771/papers/chandrasekaranetal99.pdf>.

Document Lifecycle Management Forum (DLM Forum). 1997. *Guidelines on best practices for using electronic information*, section 2.4. Luxembourg: European Commission. <https://dlmforum.typepad.com/gdlines.pdf>.

- Duranti, Luciana. 2003. "More than Information, Other than Knowledge: The Nature of Archives in the Digital Era": *Cadernos Bad*, Lisboa, Portugal, num. 2: 6-16. ISBN: 007-9421 <https://www.bad.pt/publicacoes/index.php/cadernos/article/view/845>.
- Duranti, Luciana; Thibodeau, Kenneth. 2006. "The Concept of Record in Interactive, Experiential and Dynamic Environments the View or InterPARES": *Archival Science*, num. 6, 13-68. DOI: 10.1007/s10502-006-9021-7.
- European Commission. 2008. *MoReq2 - Model Requirements for the Management of Electronic Records*, 2nd version, Glossary Section. Brussels: European Commission. https://cdn.ymaws.com/irms.org.uk/resource/resmgr/moreq2_type/set_version.pdf.
- Green, Peter; Rosemann, Michael. 2004. "Editorial Preface: Ontological Analysis, Evaluation and Engineering of Business Systems Analysis Methods". *Journal of Database Management*, num. 15, April/June (2004): i-ii.
- Gruber, Thomas. 1995. "Toward Principles for the Design of Ontologies Used for Knowledge Sharing". *International Journal of Human-Computer Studies*, vol. 43, num. 5-6, 1995: 907-928. <https://doi.org/10.1006/ijhc.1995.1081>.
- Guarino, Nicola; Giaretta, Pier-Daniele. 1995. *Ontologies and Knowledge Bases: Towards a Terminological Clarification*. Trento, Roma: Laboratory for Applied Ontology. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.320.8006&rep=rep1&type=pdf>.
- Guarino, Nicola. 1998. *Formal Ontology in Information Systems*. Amsterdam, The Netherlands: IOS Press: N. Guarino (Ed.): 3-15. Amended version of a paper appeared in N. Guarino (Ed.), Formal Ontology in Information Systems, *Proceedings of FOIS'98*, Trento, Italy, June 6-8, 1998 <http://dl.acm.org/citation.cfm?id=521669>.

- International Council on Archives (ICA). 2014. Multilingual International Terminology Database, Entrada por: "record" <http://www.ciscra.org/mat/mat/term/60>.
- International Standards Office (ISO). 2001. Standard ISO 15489-1:2001. Information and Documentation – Records Management – Part 1: General. Geneva, Definition 3.15.
- InterPARES Terminology Database. 2012. Entradas por 'record' y 'digital record': http://www.interpares.org/ip3/ip3_terminology_db.cfm?letter=r&term=41 and http://www.interpares.org/ip3/ip3_terminology_db.cfm?letter=d&term=222.
- Milton, Simon. 1998. "Top-Level Ontology: The problem with Naturalism". *Formal Ontology in Information Systems*, N. Guarino (Ed.), (1998). Amsterdam: IOS Press: 85-94.
- Neches, Robert; Fikes, Richard; Finin, Timothy; Gruber, Thomas; Patil, Ramesh; Senator, Edward; Swartout, William. 1991. "Enabling Technology for Knowledge Sharing". *Artificial Intelligence Magazine*, vol. 12, num. 3. Menlo Park, CA: American Association for Artificial Intelligence Press: 36-56.
- Noy, Natalia; McGuinness, Deborah. 2001. *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, (March 2001). <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf>.
- Pearce-Moses, Richard. 2005. *A Glossary of Archival and Records Terminology*. Chicago: The Society of American Archivists. <https://dictionary.archivists.org/entry/record.html>.

- Souza, Renato; Tudhope, Douglas; Almeida, Mauricio. 2011. Towards a taxonomy of KOS: Dimensions for classifying Knowledge Organization Systems. International Society for Knowledge Organization (ISKO). From 11th ISKO International Conference, Rome (Italy): 23-26 February 2010.
- Steve, Geri; Gangemi, Aldo; Pisanelli, Doménico. 1998. Integrating Medical Terminologies with ONIONS Methodology. *Reperto Informatica Medica*, Istituto Tecnologie Biomediche, CNR, Italy <http://www.loa.istc.cnr.it/old/Papers/onions97.pdf>.
- Swartout, William; Patil, Ramesh; Knight, Kevin; Russ, Thomas. 1996. *Toward Distributed Use of Large-Scale Ontologies*. University of Southern California <https://www.aaai.org/Papers/Symposia/Spring/1997/SS-9706/SS97-06-018.pdf>.
- Uschold, Michael; Grüninger, Michael. 1996. "Ontologies: Principles, Methods, and Applications". *Knowledge Engineering Review*, vol. 11, num. 2, Cambridge Journals On-line: 93-155.
- Wand, Yair; Weber, Ron. 2004. Reflection: Ontology in information systems, *Journal of database management*, vol. 15, num. 2: iii-vi.
- World Wide Web Consortium (W3C). 2010. Web Ontology Language Overview (OWL). Official Consortium Website <http://www.w3.org/TR/owl-features/>.

THE CONCEPT OF DIGITAL RECORD:
AN INTENDED ONTOLOGY

*I would like to express my deepest gratitude to
Kenneth Thibodeau, Luciana Duranti, Alicia
Barnard, Claudia Lacombe Rocha, Giovanni
Michetti and Rosely Rondinelli from InterPARES
Teams, who generously provided their knowledge
and expertise which greatly contributed to
the charts, although they may not always agree
with all their interpretations or conclusions.*

*Things exist when they have name; the substance
of things stems primarily from the name, and even
disappearing, its essence remains in the name.*

UMBERTO ECO

Postscript to *"The Name of the Rose"*

Section 1: Ontologies

We have common knowledge acquired about the concept of “record” in Archival Science, which is obviously an essential and indispensable requirement for any responsible person for its management and/or preservation. For a long time, both Archival Science and Diplomatics have studied, disassembled and analyzed this concept. The resulting definitions about it have been mainly presented through dictionaries or glossaries, or as “definition pages” in some standards, like ISO-15489-1, or in projects related to the subject, as MoReq. For many years, these definitions were about records on “*traditional supports*”, but in the light of digital records at certain moment it was necessary to reconsider this knowledge given that the prior one had become obsolete with the advent of this new type of records. Global increase of digitally created information –especially digital records and their growing importance required the development of new tools for its knowledge, management and preservation. It was thus necessary to work with contemporary concepts developed in contexts of this digital version and not only on those based just on traditional support definitions; not because they were wrong; they were simply outdated.

It would seem that the concept of “record” in its archival meaning has been more than widely defined, in fact the concept is more complex than what it seems at first glance. Modern archival theory makes a clear difference between “document” and “record”, as well as among the different types of the latter, but still there are environments where these concepts are mixed up and differences are not quite well understood. Furthermore, the term is now more complex since “records” have become increasingly digital. This has not been only a change of medium; it has meant a profound change of the inherent nature of records. For years, definitions treated digital records in the same manner they did with *traditional* records, as if they were the same, now written in digital media. Nowadays we know that due to their nature, digital records are quite different from prior ones, especially in regard to their management and preservation. Because of that, numerous definitions of ‘record’ are outdated, even many of those who have tried to define ‘digital record’.

As an example of this we can use some common, interesting and long-time shared definitions about ‘record’ and ‘digital record’:

- The *Glossary of Archival and Records Terminology* of the Society of American Archivists presents seven definitions for ‘record’; the two of them closer to Archival Science are: 2) Data or information that has been fixed on some medium that has content, context, and structure; and that is used as an extension of human memory or to demonstrate accountability. 3) Data or information in a fixed form that is created or received in the course of individual or institutional activity and set aside (preserved) as evidence of that activity for future reference”. In regard to ‘digital record’, the glossary refers to ‘electronic record’: “Data or information that has been captured and fixed for storage and manipulation in an automated system and that requires the use of the system to render it intelligible by a person” Pearce-Moses (2005).

- The *InterPARES Terminology Database* defines ‘record’ as: “A document made or received in the course of a practical activity as an instrument or a by-product of such activity, and set aside for action or reference”. In regard to ‘digital record’ the project defines: “A digital document that is treated and managed as a record” (Interpares Terminology Database 2012). This definition was later perfected as: “[...] document created (made or received and set aside -i.e. kept, saved- for action or reference) by a physical or juridical person in the course of a practical activity as an instrument and by-product of such activity. When the record is set aside and used in digital form then is a digital record irrespective of the original form in which it may have been made or received” (Duranti & Thibodeau 2006, 3).
- In international standard ISO-15489 ‘record’ is defined as: “[...] information created, received, and retained as evidence and information by an organization or person in pursuance of legal obligations or in the transaction of business” (ISO-15489 2001). No specific definition is provided there for digital or electronic record.
- MoReq2 specification defines ‘record’ as: “[...] document(s) prepared or received by a person or organization in the course of business, and retained by that person or organization”. And for ‘electronic record’ it defines: “[...] a record in electronic form as a result of having been created by application software or as a result of digitization, e.g. by scanning.” (European Commission 2002).
- The ICA –International Council on Archives– *Multilingual International Terminology Database* refers to ‘record’ almost exactly as ISO-15489: “A document created or received and maintained by an agency, organization, or individual in pursuance of legal obligations or in the transaction of business” (ICA 2014). For ‘digital record’ refers exactly to the two definitions of the Glossary of Archival and Records Terminology of the Society of American Archivists and the InterPARES Project aforementioned.

Looking at the definitions presented –current as they are– and even adding some more, it is easy to conclude that all of them resemble each other and indeed give us a rather good idea of the term, some of them highlighting several elements that seem particularly relevant to their authors. But beyond simple definitions we could raise some additional questions: Which are the components of a digital record? Does it have subcomponents? In how many levels? Which are the relationships and/or conditions for these subcomponents? How is the whole perceived? How it is disassembled? Trying to answer these questions from glossaries would be possible but to do so, many additional documents should be read, analyzed and pondered; and many charts, tables or sketches should be done to achieve a logical and consistent result. It would be necessary to read over and over again lists of terms to establish sets, relations, rules, equivalences, restrictions, ascendant and descendant terms, etc., among all these concepts. Further readings would be required to put them in context, which would take a long time. Even then, we could not answer directly from glossaries deeper questions without extra background about the concept. This is because they are only meant to define loose terms, but not within a further context or structure which could define not only the terms but their interrelationships, their levels and the conditions that must be satisfied to combine these terms; in short, to understand them thoroughly. Definitions from different sources are similar, but not identical, and some of them include elements not found in others. Organizations and projects which have undertaken the task of conceptualization did it in various ways, trying to define as much as possible the elements which allow them to establish as clearly and simply the concept of “digital record”. In order to find out the resulting concepts, dictionaries or glossaries are usually consulted, and sometimes books or articles, but rarely ontologies; and the main reason for doing that is because these are scarce in the field of archival records, and even scarcer on digital records. This raises the question: ¿Why ontologies for defining “digital record”?

Because of these new global trends, government information and archives are increasingly being required to be available to the public, and more people want to access them through the Web, so this information must be “Web prepared”. This global increase of digitally created information –especially digital records– and its growing importance, have required the development of new tools for its knowledge, management and preservation. The international archival community has crafted in recent years valuable theoretical and practical conceptualizations and contributions for modern management of huge amounts of this kind of data and information. Archival Science, combined with other emerging disciplines have made fundamental contributions in late years to knowledge innovation in order to improve continued and better access to digital documents and records.

One of the emerging global proposals with innovative technical solutions for knowledge representation and management has been the “*Semantic Web*”, based on a series of recommendations made by the World Wide Web Consortium (W3C), whose objective has been to improve access to all kinds of information on the web by a large number of users and institutions through an optimized use of the global network. One of the basic premises of this proposal has been the use of documentary “*semantic representations*”. In short, it is a set of principles and tools to improve the formal description of documents, which in turn will favorably affect document search and retrieval. The underlying principle is that documents formally described under a set of rules and structures become much more accessible through interoperable systems in the World Wide Web environment. Applying this in combination with the modern treatment of digital records promises to be a better solution for the description and management of such records, improving their access both in local records management systems and at the web level. Records built upon these principles would thus become “*semantic records*” which will form “*semantic archives*”.

In addition, better formal representations of the attributes of digital records within the new preservation systems, will help to keep them more stable over several generations. These representations

can definitely assist in the future to design and implement better strategies for recordkeeping, migration, transfer, etc., of digital records. It is known that the form of records must be fixed and their content must be stable, but technology will not. It is known too –paradoxical as it sounds– that to preserve digital records they should not remain in their native format forever; otherwise this very fact will make them inaccessible in time. It is already clear nowadays and in the future that one of the main challenges for preservation of digital records is how to cope with the ever-changing technology: hardware, programs, operating systems, formats, etc. All these elements have a direct impact and affect the use, management, authenticity and preservation of these records. Better formal documentary representations will undoubtedly help them to move more smoothly and safely through all these changes.

Moreover, users' communities have changed in recent years; previously, documentary records were almost exclusively for the use of a few officials within their organizations, but nowadays it is increasingly required that public users can access them, since the number of documents is growing exponentially. The principles of access to public information, transparency, accountability, open government, participatory democracy, etc., are becoming more universal every day, and the public is demanding more and more information from archives. These users need better tools to access, combine, analyze and reuse this information. But for really making this information meaningful for users, it is essential to link it with more similar information. For this reason, it is essential to combine current archival knowledge with adequate information management under these new formal representations in the short term, as a solution for improved access to information. Therefore, semantic [Semantic] approaches promise to improve access to the information thus represented.

The Semantic Web has introduced a whole conceptual structure made up of certain components, some of them already existing and others new. There are technological and conceptual components, accessories, and so on. The structural scheme of the Semantic Web introduced by Berners-Lee (2001) is already well known;

it contains technological layers which have more to do with hardware, software, networks, etc., as well as conceptual layers such as XML, RDF2 and OWL standards which are intended to turn the web into this global logical infrastructure where data and documents can be shared and reused among different types of systems and users; it is a complex structure. The “OWL” layer –Ontology Web Language– is one of these tools aimed to develop specific vocabularies in order to associate documentary resources formally (W3C 2010). OWL provided a Web standard formal structure intended to explicitly represent the meaning of terms and the relationships between these terms in vocabularies. OWL is based on the Description Logics framework –DL– which forms the basis of several well-known ontology languages (Baader *et al.* 2004, 3-28). OWL never really established itself as a universal methodology for this, but its principle of using an ontology to perform the task has become quite relevant and is the basis for many other endeavors in this area. Current approaches of knowledge representation with ontologies are being increasingly studied as an interesting tool since in theory, knowledge must be represented in a way that is formal, consensual, reusable, and eventually, computer readable. Ontologies provide a way to represent knowledge in such a manner. But, ¿what exactly are these “ontologies”? And, ¿what is their potential utility in modern Archival Science?

Originally, the term “ontology” has been used for a long time in Philosophy, particularly in Metaphysics, as a study aiming to define which things exists in reality and what can be said about them. It has to do with the science of what “is”, the “entity”, and its transcendent properties; its essence is the study of “being”, regarding what “is” and “how is” in order to achieve a systematic explanation of it. In this philosophical sense an ontology is therefore a particular category structure representing a certain view of the world about the nature and organization of reality, i.e. that which “exists”. In more recent times, the term and some of its original techniques have been adopted by various disciplines which aim –beyond defining “being”– to establish the fundamental categories of things from the study of their properties, attributes and

structures; consequently becoming a new and diverse concept. Thus, in short, a distinction can be made between “Ontology” as a philosophical consciousness of *being*, and “ontologies” as conceptualizations built on certain knowledge “[...] in a broad sense, ontologies are sets of inference rules for interpreting a system of symbols... inferences based on ontologies can be used to draw implicit information from a set of semantic representations in a knowledge base” (Arakawa 2008, 2).

Ontology as a conceptualization is a complex concept. Guarino & Giarretta (1995) conducted a study and found seven different approaches to the concept with seven corresponding interpretations; in short: 1) The aforesaid traditional approach: ontology as part of Philosophy; 2) Ontology as an informal conceptual system; 3) Ontology as a semantic formal explanation; 4) Ontology as a specification of a conceptualization; 5) Ontology as a representation of a conceptual system through a logical theory, characterized either by specific formal properties or only for specific purposes; 6) Ontology as the vocabulary used by a logical theory; 7) Ontology as a specification or a meta-level of a logical theory. Thus, there are multiple approaches to ontologies which comprise different interpretations; therefore, “ontology” is a polysemic term, and each one of its interpretations has various manifestations and utilities for different uses and disciplines: in linguistics and literature, in artificial intelligence, in theory of knowledge, in biomedical sciences, in geomatics and in computer science –especially in databases design–, in information retrieval and in computational linguistics. Other disciplines use them increasingly as well and of course, they are also useful for archival science. Due to this great disciplinary variety, it is only natural then to observe different approaches and treatments of the term and therefore is necessary to be aware of its multiple uses by different communities with different contexts, all of which could lead to confusion during its study and use.

To understand the concept in its basic essence, it is then convenient to depart from the most general and common definitions. In its simplest form, an ontology *is a theory that explains knowledge*.

- A widely used definition is the basic proposal prepared by Gruber (1995): “[...] an ontology is an explicit specification of a conceptualization”. Borst (1997, 20) modified it slightly by stating that “[...] ontology is a formal specification of a shared conceptualization”. According to them, in these definitions ‘conceptualization’ means a model or idea of some object or fact in the world whose relevant concepts are identified: objects, concepts and other entities that are assumed to exist in some area of interest and the relationships containing them; it is an abstract, simplified view of the world to be represented for some purpose; ‘*explicit*’ means that the type of concepts used and restrictions for their use are defined in a clear and manifested manner; ‘*formal*’ refers to the fact that ontology is machine readable or at least feasible to be translated into it; ‘*shared*’ means it describes consensual knowledge accepted by a certain group.
- Swartout *et al.* (1996) stated that an ontology is “[...] a set of structured terms describing a topic or ‘domain’. It is intended to provide a basic structure for a knowledge base”.
- According to Neches *et al.* (1991) “[...] an ontology defines the basic terms and the relations comprising the vocabulary of a certain subject area, and the rules for combining terms and relations to define extensions to such vocabulary”.
- For Uschold & Grüninger (1996) “[...] an ontology is a set of clusters or ‘classes’ with different types of relationships among them. The main ‘class’ is the one that describes the taxonomy¹ of the ontology. Relationships have properties whose utmost usefulness is the ability to make inferences

1 “A taxonomy is a controlled vocabulary organized in a hierarchy. Each term refers to a category, type or class. There is only one type of bond, which means ‘is a variety of’ and corresponds to a subclass relationship. Strictly speaking, each node in a taxonomy has exactly one ‘father’, but the term ‘taxonomy’ often refers to hierarchies with multiple ‘parents’. It is also sometimes used to refer to networks with more than one type of bond” –Uschold & Grüninger. *Op. Cit.*

from them about classes... ontology is the term used to refer to the shared understanding and can be used as a unifying framework to solve problems... An ontology necessarily entails or represents some sort of world view with respect to a particular domain or topic. This view is often conceived as a set of concepts (e.g. entities, attributes, and processes), their definitions and their relationships, which is known as a conceptualization”.

- Noy & McGuinness (2001) stated that “[...] ontology is a formal and explicit description of concepts in a domain of action consisting of classes, where each of them has properties –which describe various characteristics, roles and attributes of the concept– as well as restrictions”.
- Arano (2005, 7) defines ontology as “[...] a formal, explicit representation of the conceptual structure of a field of knowledge. Ontology is a semantic support for words that are described as linguistic objects in a lexical or terminological database. The conceptual relations represented in an ontology have a wide variety which depends on the field of knowledge to structure. An ontology is built with the aim to share and reuse stored information which, after being formalized, can be interpreted both by persons and by computer programs”.
- To Chandrasekaran *et al.* (1999) “[...] an ontology is a theory of content about the classes, properties and relationships of objects that are possible in a certain domain of knowledge and are explained by a representation language”.
- Wand & Weber (2004) stated: “[...] although most ontologies are a taxonomy rather than a theory, still have predictive and explanatory capabilities”. They also added that “if a phenomenon is correctly classified according to one theory, people can better understand and predict that phenomenon and consequently work more efficiently with it”.

Many of the information sciences –especially Computer Science and Artificial Intelligence– have adopted the use of modern

versions of “ontology”. In the field of these sciences “what exists is that what can be represented”. Therefore in this disciplinary approach “[...] ontology is an artifact of engineering, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions about the intended meaning of the vocabulary words” (Guarino 1998). It can be summarized briefly and simply that in ICT, Information and Communication Technologies, the idea of “formal description” or “formal specification” means essentially that what is described in such a manner could be eventually programmable in a computer system. Hence in these disciplines the term *ontology* is basically a programmable formulation of a specific vocabulary to describe the common knowledge about a certain reality from a given domain. It should also be noted that in these disciplines the term *artifacts* is used to name the concepts and the specific vocabulary, and *domain* to name a certain area or field of knowledge that is being defined. More specifically, the *artifact* usually consists of a thorough and comprehensive conceptual structure within a field of knowledge or *domain*, and is constructed in order to facilitate communication and information sharing among users, databases and computer systems. Its construction and use aims to go beyond an organization or region; and rather aims to an entire community interested in that field of knowledge. In this paper there is no intention to reach formal descriptions up to the level of describing items in such a manner that they can be programmable in a computer system; at least, not immediately.

From the numerous existing definitions can be inferred that the concepts of *ontology* –in their modern form– do not diverge substantially among them, but there are different “emphasis” or “focusing” depending both on the approach of each discipline and the main purpose of each ontology. Nevertheless, this tends to produce many specific definitions which confuse those who study them. Green & Rosemann (2004) affirmed that “[...] not only the type of research work that is conducted under the umbrella term ‘ontology’ varies significantly, but also the understanding of a certain ontology varies significantly in terms of scope, detail and purpose”.

In this paper other diverse disciplinary approaches are discarded, and therefore only the Archival Science approach will be considered here. Under this particular disciplinary focus of interest and for the scope and purposes of this paper: “[...] ontologies are conceptualizations built to formally describe –with a vocabulary of terms, attributes and relationships– common and consensual knowledge related to a certain topic, in a way that such knowledge can be understood, shared and reused mainly by different groups of people and eventually by computer applications”. ‘Conceptualizations’ are formal representations of a body of knowledge which comprises objects, concepts and other entities which are assumed to exist in some area of interest as well as having existing relations among them; they are simplified abstract views of the world that is intended to be represented for some purpose. Every body of knowledge or system based on this is committed explicitly or implicitly to some conceptualization in a certain field of interest or ‘domain’ where there are entities or ‘objects’ to be described in an explicit, formal, consistent, consensual and semantic way. The vocabulary of representation provides a set of terms to describe objects and their relationships in a certain field of interest. ‘Formally’ means that the ontology is described in such a language that it can be processed by a computer or at least, that it is feasible to be transferred into a program.

Beyond definitions, and trying to better understand ontologies, a ‘characterization’ can be extracted from all the aforementioned definitions; i.e., a list of all relevant and/or common characteristic elements which shape them. Thereupon ontologies:

- Aim to make explicit the assumptions, views, structure and content about a specific topic or field.
- Describe some knowledge accumulated and shared by a community about a subject or field.
- Define the concepts, relationships, and other important features which model a certain subject or field.

- Allow sharing knowledge about a common subject between persons, between computers, and eventually between both of them.
- May have different levels of generality or specificity. Since no ontology is able to achieve a complete description of a certain topic, a set of ontologies allows construction from the general to the particular.
- Provide a vocabulary for a certain field through a set of terms.
- Allow new terms from the reuse of old terms or their combination, and permit to cast new concepts for old terms.
- Permit reuse of the formal descriptions of their terms in other entities.
- Must specify or describe in a formal way the meaning or semantics of its terms, establish their restrictions and disambiguate their interpretation.
- Allow “mapping”; i.e., specify equivalences, interrelationships and combinations among terms within the same ontology or among other ontologies.
- Allow to add rules or conditions required by concepts (identity, rigidity, unity, dependence, obligatory, etc.).
- Are not unique or absolute; there can be multiple representations of the same concept, as it can be represented in many ways and for different purposes.
- Some ontologies foster the construction and use of a set of terms to represent facts within a specific instance of a field of knowledge or domain, while others are built making an emphasis on the set of facts that are shared.

In addition to the various possible approaches derived from the use or discipline, (Souza *et al.* 2011) have established different levels or depths of “ontological precision”; from the simplest to the most complex they have distinguished:

- Lexicon or Catalog (vocabulary with definitions in natural language).

The concept of digital record:...

- Simple taxonomy (vocabulary, data dictionaries).
- Complex taxonomy (vocabulary, classification, data dictionaries, hierarchies).
- Thesaurus (taxonomy with related terms).
- Relational model/Scheme (type constraints and arbitrary relations among entities).
- Complete axiomatic theory.

Some other authors also make distinctions on levels or depth of ontologies. They generally name “lightweight ontologies” those which only cover the stages of vocabulary, classification or thesaurus, while they name “complex ontologies” those which include axioms, restrictions, etc. Milton (1998) distinguishes between ontologies focused on theory and those pragmatically oriented. For him, the ones focused on theory are those created from some scientific, humanistic or social theory and make emphasis on it, while pragmatically oriented ontologies are those stemming from the consensual practice of a discipline and constructed usually having in mind that eventually they would be used by computer systems. The latter are therefore common in the practice of information science and are directed to specific practical areas as in the case of recordkeeping.

Depending on the discipline, focus and purpose, ontologies can be of many types. There are several proposals by different authors, e.g., Steve *et al.* (1998) proposed this typology:

- Generic ontologies –those who represent broad theories or general concepts of a field fundamental to a field of knowledge.
- Domain ontologies –those who represent the specialized and relevant knowledge about a certain domain or sub-domain, such as Medicine, Obstetrics, Library or Archival Sciences.
- Representation ontologies –those who specify the entities with general representation underlying on formalisms of knowledge representation, without defining what

is represented; also called meta-ontologies or ontologies of meta-level categories.

- Task ontologies or method ontologies –those created for an activity, procedure, task or specific method.

Among the numerous variants in ontologies, some components can be distinguished and used frequently varying widely among different authors and purposes, a resultant list can be stated:

- Classes –sets, collections, types or categories of objects or concepts, whether abstract or concrete.
- Attributes –aspects, properties, features or characteristics that classes and objects may have.
- Relationships –types of interaction between objects or concepts (*consists of, is subclass of, is connected to, etc.*).
- Rules –propositions of the form “*if-then*” (antecedent-consequent) describing the logical inferences to be drawn from a statement expressed in a particular way.
- Restrictions –formally stated descriptions of what must be true for a statement to be accepted as valid input.
- Functions –calculations or operations performed and added to an object or class.
- Cardinality –The number of *at least, at most* or *exactly* relationships of an item with other entities.
- Axioms –complex statements and their rules presented in a logical form that together comprise the overall theory which the ontology describes in its domain of application;
- Instances –each individual member of a class.

The current importance of ontologies lies in the fact that all knowledge must be expressed in conceptualizations that inevitably require a specific vocabulary for a certain field. Ontologies provide a way to facilitate these conceptualizations and specific vocabulary associated while allowing making explicit the assumptions and structure of the subject to define. They facilitate the analysis of knowledge, provide order and structure and increase reuse. Their

future importance lies in the fact that they can effectively help to achieve in a shorter term that information be shared among individuals, among systems and among both of them in an unified and coherent way, eventually leading to a more accurate, relevant and timely retrieval of such information. Like thesauri, ontologies have to be built, one by one, for a certain “domain”; i.e., a certain field or area of knowledge: Computer Science, Linguistics, Microbiology, Astronomy, Library Science, Archival Science, etc. As each one of these fields is very large by itself, more and more detailed ontologies are to be required for subfields or specific areas of knowledge. It is a deeply intellectual work, not delegable to machines—at least not for now—. Therefore a whole set of ontologies is still to be built, particularly for Archival Science, and it is desirable that they are built by professional archivists. Much of the problem is that not only it is needed to compile and systematize common knowledge already accumulated about this area, but also to integrate new knowledge, as new theories arise from research, as well as changes and other emerging characteristics imposed by technological advances, meaning specially the development and increase of digital records.

Section 2

The intended ontology

Under the current particular focus of interest –Archival Science– and for the scope and purposes of this paper, the ontology constructed and proposed is a personal conceptualization built to formally describe –with a vocabulary of terms, attributes and relationships– common and consensual knowledge related to the topic ‘digital record’, in a way that such knowledge can be understood, shared and reused mainly by different groups of people and eventually in a future by computer applications. From all the aforementioned about ontologies, it can be stated that for this particular case, the intended ontology:

- Is a *domain ontology*, representing the specialized and relevant knowledge shared about the domain “digital record”, its components, terms, and relationships;
- Gathers many definitions of the concepts, relationships, and other important features which shape the concept of “digital record”, making explicit the assumptions, views, structure and content about this topic;
- Is a *lightweight ontology*, since it pretends simply to define the basic terms and the relations comprising the vocabulary

- related to “digital record”, and the rules for combining terms and relations useful to eventually extend such vocabulary;
- Has an *ontological precision* at the complex taxonomy level, since it only covers the stages of vocabulary, classification, data dictionary, and hierarchies; gathers a shared vocabulary for the field of interest through a common set of terms;
 - Has two dimensions inside, since it analyzes the concept of “digital record” from its intellectual components and from its digital components, and performs the mapping (equivalences, interrelationships and combinations) of terms among them;
 - Has no intention to reach formal descriptions up to the level of describing items in such a manner that can be immediately programmable in a computer system;
 - It is not unique or absolute; there can be multiple representations of the same concept, as it can be represented in many ways and for different purposes;
 - It allows new terms from the reuse of old terms or their combination, and permits to cast new concepts for old terms;
 - It allows reuse of the formal descriptions of its terms in other entities.

The InterPARES Project –The International Research on Permanent Authentic Records in Electronic Systems– has been studying digital records for years trying to establish –among other things– their characteristics, components and relationships. At the same time this project developed a glossary of terminology about digital records preservation, with special emphasis on the features and “modern” concepts derived from the development and use of digital records. To explore beyond a glossary, in 2008 the project began to introduce some basic ontologies in relation to records’ knowledge. Those ontologies were very basic, since they were based on the “simple taxonomies” level aforementioned. Since knowledge thus compiled has evolved over the years as research introduces new ideas and refines concepts; ontologies cannot be static. This section

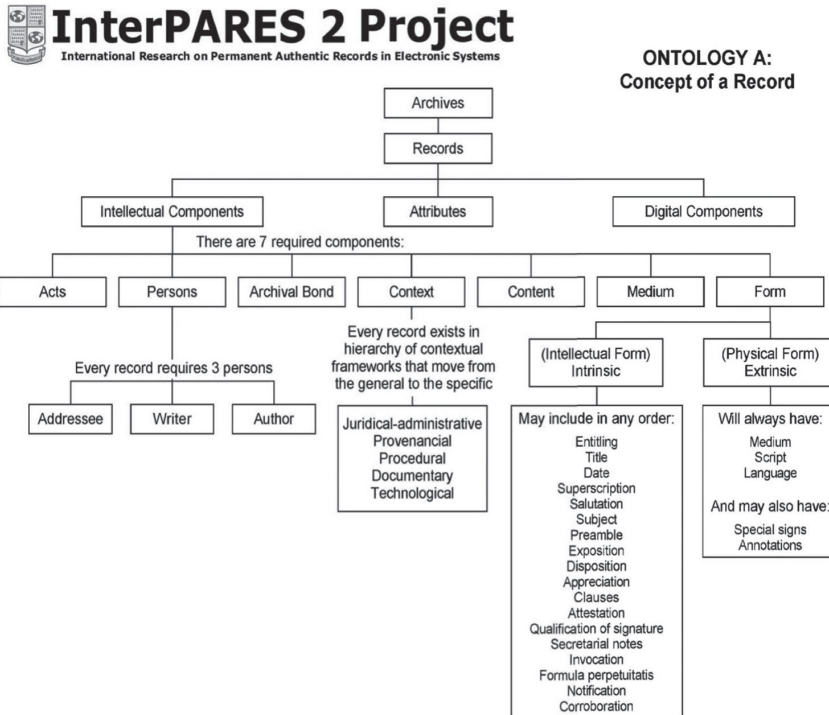
presents a personal updated and supplemented proposal of the original “Ontology A”, which analyzed the “concept of a record” as well as its components and relationships. This idea began

[...] following the classic Diplomatic assumption that –regardless of differences in nature, provenance or date– all records are similar enough to make it possible to conceive of one typical, ideal documentary form containing all possible elements of a record, the researchers hypothesized that, while they may manifest themselves in different ways, the same elements that are present in traditional records exist either explicitly or implicitly in electronic records, and that all electronic records share the same elements (Duranti 2003, 10).

Diagram 1 show as it was in its original simple version:

Simple as this early form was, while observing this representation some interesting facts could be inferred: Obviously, sets or groups of digital records make up digital archives. A record is made by entities called intellectual components, attributes and digital components. In that version digital components were only introduced as a whole, but not unpacked. Following the construction and in regard with the intellectual vision of the record, the main components were there: each record required then seven intellectual components, and all of them must always exist for each record. Under the point of view of Diplomatics, individual records require five persons during their creation: author, writer, addressee, originator and creator. The first three become intellectual components of the record, and the other two become attributes. There were already five “contexts” or frameworks for each record. It was established too that there should be a content –which must be stable– and a documentary form –that should be fixed–; such form had two manifestations: the intellectual or intrinsic form, and the physical or extrinsic form. More inferences have been made in further analysis of this diagram. Notwithstanding its simplicity, the chart allowed a better understanding about the concept of a record than a simple definition. The main goals of the current

Image 1.



Source: Terminology Cross-domain Task Force. Appendix 22: *InterPARES 2 Project Ontologies*. In: *InterPARES 2: Experiential, Interactive and Dynamic Records*. Luciana Duranti & Randy Preston (Eds.). 2008, 729-731. http://www.interpares.org/ip2/display_file.cfm?doc=ip2_book_appendix_22.pdf

proposed version in regard with the original one are 1) To analyze the evolution of the shown concepts since they were published in 2008, 2) To complete them as much as possible, and 3) To develop the missing part in the original, i.e. –digital components–. All of this from a personal point of view.

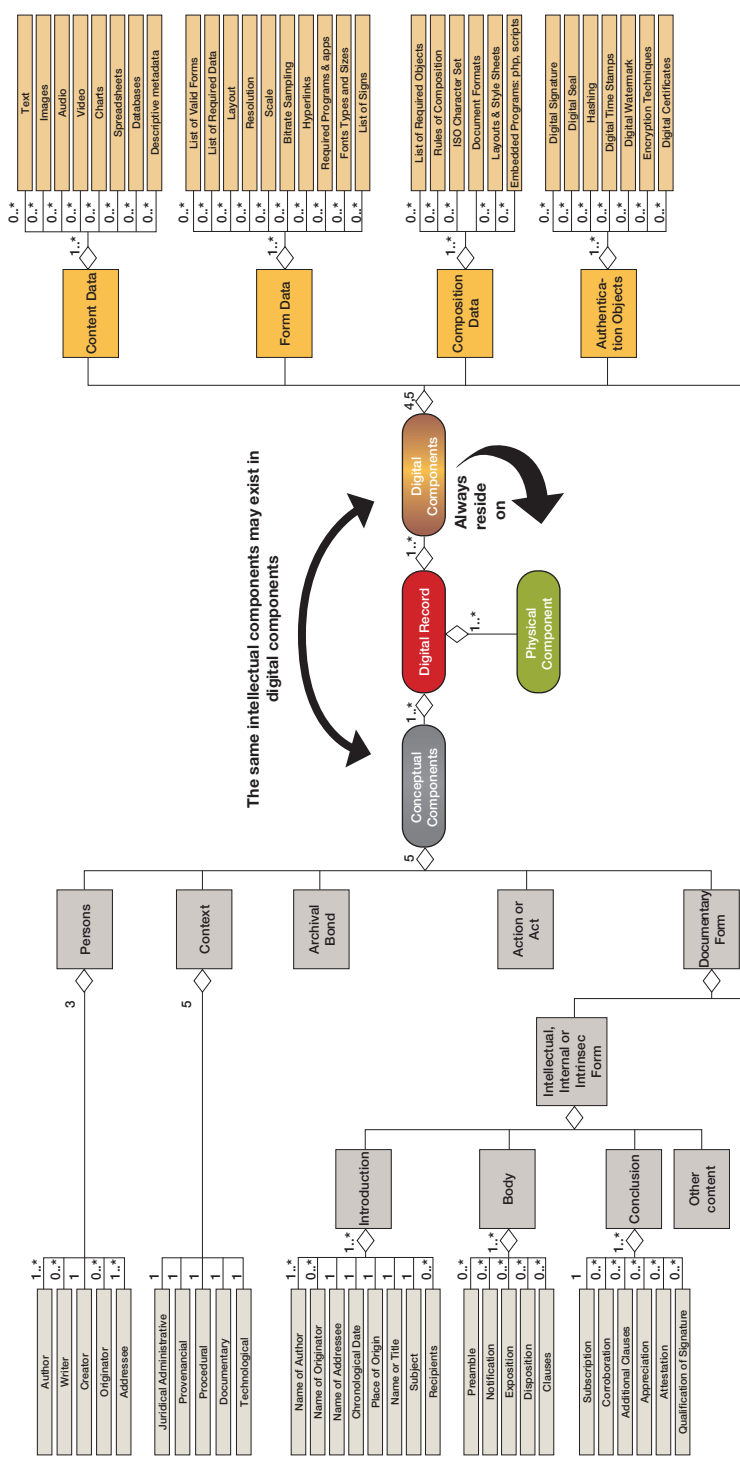
The main glossaries about digital records as well as numerous books and papers regarding related concepts and terms were consulted to perform this task, in order to extract, analyze and

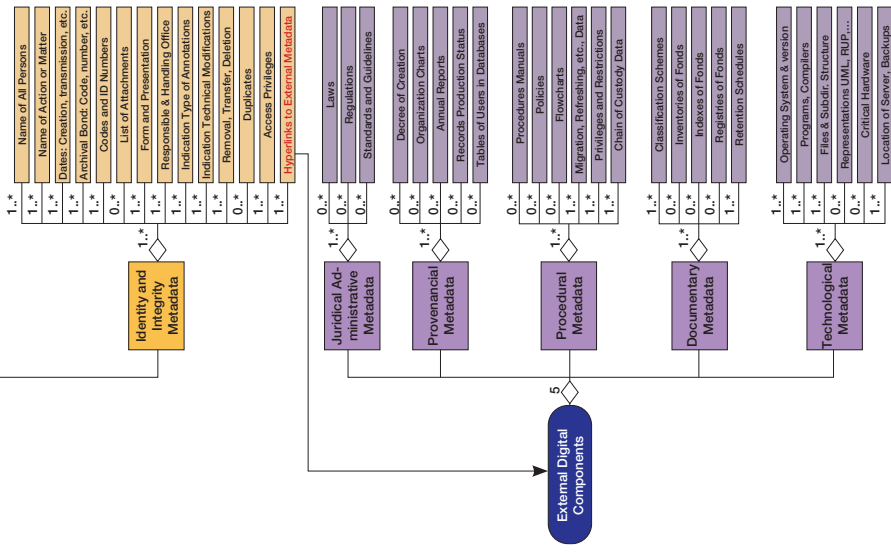
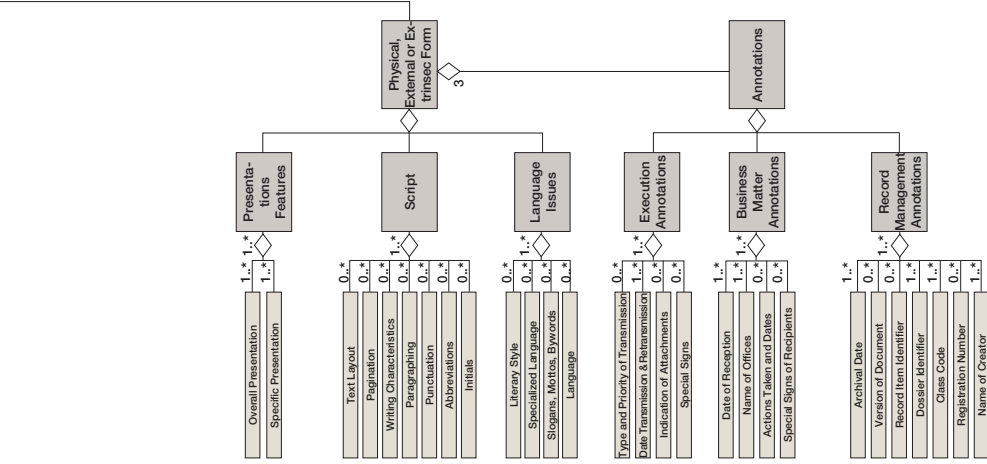
compare different visions of concepts and terms. The *ontological analysis* approach was used, in order to study, understand and analyze contents as such, independently of the way it was represented. Therefore, special attention was paid to each concept found trying to establish its identity, parts, wholes, composition, relations, and properties. Trying not to overload the concepts with too many elements, only basic features were included in the chart. This new version was casted specifically for digital records, with two main dimensions trying to establish the intellectual as well as the digital components of a digital record and the possible mapping between them. Thus, a digital record can be visualized either from its intellectual composing elements or from its digital composing elements. Intellectual components are also known as *conceptual components*, but in an ontology, where almost everything is conceptual, this term could lead to some ambiguities and confusions; therefore *intellectual components* will be used in this paper.

To begin with, the gathered information was drawn in graphic form by means of a mind map, one of the simplest charts to order knowledge. As concepts flowed, new elements were added, and different and more complex chart layouts were designed. After many rearrangements, the current map is formed by 165 parts or pieces. A stable version was achieved. The map gradually evolved from its simplest form to its present one in order to prepare it to shape it as an ontology with the minimum possible changes in design and shape; for instance, the parts considered as main classes were shaped as ovals; other secondary classes were shaped as square boxes; elements were shaped as wide rectangles. Unified Modeling Language –UML– notation was selected. The results of these are stable versions of a map and an ontology of the concept of a digital record. Here it is the “printable version” of the ontology, numbered as “version 2”:

The left part of the chart (grey boxes) represents the view of the intellectual components of the digital record, which are the conceptual and grouping parts of the elements existing in an ideal record form. From the point of view of Diplomatics, this is a conceptual idealization of the features or defining characteristics

Image 2. Ontology of the Concept of Digital Record.





which comprehend a record, identifying it uniquely and distinguishing it from others; a certain combination of expressions of intellectual components of a record is unique and unrepeatable for any other.

The right side of the chart (yellow and purple boxes) represents the digital components of the digital record, the other approach to understanding its concept. The intellectual components mentioned previously are concepts that, ideally, shape a record and are manifested in different conceptual elements found in various parts of the record. Eventually, all of these conceptual elements must be contained or expressed in a set of digital components; i.e., digital objects that are not a record by themselves, but rather a part of a record or a set of digital records, and together contain the primary data of the record, as well as the metadata and the additional elements necessary for structuring, or manifesting its content; all contained in computer files. This is important because many people tend to think that every digital record is self-contained within a single computer file comprising the entire document in one computer system. Certainly it is common that there should be text within a record, and it is almost always conceptualized like that, but actually a record may be further shaped by other entities or information objects: it can have graphs, still and moving images, sound, fonts, embedded programs, metadata, digital signatures, style sheets, etc. Sometimes those elements are embedded within the record itself, and thus become the internal digital components of the record (yellow boxes), but some other times they exist in different computer files outside of the record (purple boxes), becoming external digital components of the record (Duranti & Thibodeau 2006, 29). Usually, the latter contain a significant part of the contexts of the digital record. Typically, they are large sets of information, since each context tends to generate a whole set of metadata of large volume; therefore, they are not stored inside each specific record. They are constructed globally for a whole series or fond and saved in computer files outside the record, obviously with the appropriate hyperlink from every primary record to those external items. Nevertheless, all digital

components together –internal and external– make up a single logical unit, a “digital record”.

Being different approaches, both intellectual and digital components, are not isolated entities from each other: the intellectual components –the concepts– are to be contained in one or more digital components –the objects–. In other words, digital components contain information objects which in turn contain conceptual items and together make up a digital record; simply the relationship is not biunivocal,² i.e., not one-to-one. Each intellectual component does not necessarily corresponds to one digital component and vice versa. Obviously not all of the possible digital components necessarily exist in each one of the records, but the list pretends to be fairly exhaustive about those who could be found as a digital component within the universe of these records. It is essential to fully understand those components to avoid assuming that by preserving a computer file holding the main body of a record, all its digital components are thus preserved. For example, it is a common mistake to omit the inclusion of the font sets (Arial, Courier, etc.) when preserving a digital archival fond, since usually the computer directories containing the fonts are in the computer’s operating system directory, and not in the directory where the digital records are kept. This is an example of a typical mistake often made just by ignorance while identifying all the digital components of a record.

There is also a third main component –the physical component (green box)– which was considered as one of the intellectual components in the original ontology, but currently is considered separately as a special component since it is neither an intellectual nor a digital component. It is the medium or physical support in which the digital components of a record are saved or stored. Implicitly, the medium does not really form part of the record, but digital components do not make a record until it is affixed to

2 *Biunivocal*: correspondence that associates each of the elements of a first set with one, and only one, of the elements of a second set, and each element of the latter with one, and only one, element of the first set.

a physical and stable support, and the destruction of the support may imply the destruction of the document, thus a special class is needed for it.

Both types of main components, intellectual and digital, included as far as possible the four constituent parts of an archival document as set out in DLM Forum guidelines and assumed in the Moreq2 model: content, structure, context, and presentation (DLM Forum 1997, 13), and (European Commission 2008, 18).

The “printable version” of the ontology presented here is a brief one: it shows only the identity, composition and relation parts of all components. There are full interactive versions of the map and the ontology built in html with all the information gathered about each element, with a special feature within: each and every one of the 165 boxes has a resultant definition of the term and most of them have an additional explanation of characteristics, relationships, mapping, etc. This information in the html versions displays on the screen just sliding the mouse pointer on any box. Many of them link also to full texts of a set of documents with further explanation about their concepts. The total amount of the bibliography hyperlinked in this way with the chart raises to 21 documents. Hence, the full map and ontology with definitions, relationships, mapping, etc. are better displayed and fully appreciated in the html versions. Both full and printable versions are available in English and in Spanish versions at the Institute of Library and Information Sciences Research (IIBI) Website at the National Autonomous University of Mexico (UNAM): http://iibi.unam.mx/archivistica/index_ingles.html using the “ontologies” button.

The charts try to reflect a personal view of the changes and evolution found in the literature about the concept of digital records and its related elements, e.g.,: the original seven intellectual components have been reduced to five; the arrangements of some elements have changed; the separation of the medium or physical support, i.e., physical component, etc.

The translation of terms from English into other languages usually introduces another big source of confusion of concepts as a result of ambiguous translations from and into those languages,

since archival terms cannot be translated simply from regular glossaries. Many efforts have been made in this regard trying to establish the correct equivalences between terms in several languages, i.e., ICA International Terminology Database, InterPARES Multilingual Terminology Database, etc. In this case, Spanish versions of the map and the ontology were made in such language; a button in the html versions toggles the whole page from English to Spanish and vice versa. A careful translation of terms, definitions, etc. into Spanish was intended, using those updated archival terminology databases (ICA and InterPARES) in that language, trying to solve problems of ambiguity of terms. When documents in the hyperlinks of the ontology existed in Spanish versions, hyperlinks toggle to them too. So, the map and the ontology are carefully bilingual. This will help in case of an eventual computer programming of the ontology with purposes of correct translation of terms between these two languages, and obviously, similar attempts could be easily made with other languages.

It has been mentioned that this “version 2” is a stable version of an ontology on this subject, but it is by no means considered definitive or closed: it is susceptible to further refining, due to the following two reasons: 1) This is a personal interpretation of the many possible arrangements of elements, which needs further discussion. 2) As a map it can be considered highly complete; as an ontology it still lacks of some elements, (restrictions, relationships, etc.) needed to be fully completed.

As a first conclusion from the analysis of the intellectual and digital components, it can be stated that the latter contain within them information objects that include the former and which, together, constitute a digital record, not necessarily in a one-to-one relationship, and usually arranged differently in each asset. This is of capital importance while designing a digital record of a certain fond, or when assessing whether a given recordkeeping system can create and/or manage a record of a given organization. The designer or evaluator must be able to identify and establish *each and every one* of the conceptual components of the record and then be able to do the mapping with all relevant digital

components which will contain them, in order to avoid omissions, ambiguities, confusion and other errors while creating, using or preserving those records. Obviously, this mapping can be reverse performed: the analysis of all the digital components of a record must be able to establish a correspondence or mapping with all of its intellectual components.

It can also be established that –although it is not really a part of the record– a *physical component* must exist along with the others, meaning the computer support or physical medium over which digital components of a record are saved or stored. The physical component must necessarily exist since the definition of document has established that its essence is “information recorded on a support”. Thus, a document cannot exist unless it has been “fixed” on a physical medium, either traditional or digital. Therefore, if a record is a special kind of document, it requires to be fixed on a physical medium. Being physical implies too that some kind of conservation action is required.

As it has been seen, the expression of a set of ideas in this form of representation enables us to comprehensively observe levels, relationships, subcomponents, hierarchies, ascendants, descendants, conditions, etc., on the analyzed concept, which enlightens it and allows a better overall understanding. Having this representation also being added with the definitions of the adjacent or embedded terms, a more complete structural representation about the concept of “digital record” is achieved. Through this kind of representations, common knowledge on this domain or theme is shared in a more comprehensive form than just the mere definitions of the concept introduced at the beginning of this document, which are of course contained in this representation.

A more detailed analysis can arise more and more information, context and answers, as we have studied. Without a careful analysis, it is easy to intermingle “intellectual” with “digital”; “components” with “elements” of a record as synonyms; or –knowing they are not– can become a hard task to describe the difference. Thanks to the ontology it can be seen that –being similar– they are not the same that they are different entities and why they are different, and

therefore everything becomes easier to understand. Precise definitions extracted from many documents are here, but their context and interrelationships are more important. Beyond the utility of this type of representations for the understanding of the terms, it can be seen that such terms can greatly help to establish issues to consider, while designing or acquiring, for example, a records management system, or while establishing the attributes of a certain type of digital record within an organization, as well as each and every one of its associated components. With the ontology, clearer requirements for proper process, management, classification, etc., can be established, as well as the identification of all digital components to safeguard their future integrity, identity, accuracy or completeness. This will facilitate their handling and subsequent long-term preservation while keeping their authenticity.

While semantic techniques provide the methodology for doing better representations, it will be essential that archival communities be responsible to formalize the description of these relevant attributes in an appropriate representation of the records, since they are the ones who have the experience and knowledge on the subject. Therefore it is important that professional archivists know better such techniques. Under the above assumptions, fundamental characteristics of digital records must be properly identified and represented. Certainly this is done today, but unfortunately not often under this “semantic” approach.

Current records management systems are built almost entirely on databases with “*relational*” structures for storing and retrieving the documents they contain. Typically, searches are done through “keywords” retrieving the documents linked to them, but almost all of the data among documents are disassociated from each other, and few can be associated beyond the classic Boolean operators. For the same reason, these systems have a very limited understanding of the meaning of the data they store. Similarly, on today’s Web, search engines perform most of the retrieval task by trying to locate the requested words within Web pages, finding single words here and there in many of them. With billions of Web sites at their disposal, these search engines retrieve countless pages of which

the vast majority are irrelevant, making actual access to meaningful information a difficult and frustrating task. Since Web retrieval is based primarily on words with virtually no context, what is retrieved is little relative to the grand total. The primary ingredient missing from both records management systems and Web search engines is information with a better predefined context, with inherent semantics. This predefined, formal, standardized and universal information is precisely the information built with good “semantic representations”. Thanks to them, we can give utmost significance to data, both those internal to records management systems and those on the Web itself, and therefore optimize the search engines used in them. The new conceptual models derived from the theories of the “Semantic Data Model” now make it possible to define the meaning of data through their relationships with other data.

These are not just academic musings. Due to the new global trends of the *data revolution*, *big data*, *open data*, *linked data*, etc., numerous projects have been developed in this regard: virtual clouds with large amounts of data to which anyone can access any authorized data as well as add new ones, thus forming an open, structured and interoperable environment favoring data to be created, interconnected and consumed on a global scale. These projects were originally associated with libraries, repositories, and related collections, but are increasingly coming closer to archives. Representative examples are the projects of the “World Bank of Open Data”, “South Korea’s Open Public Data Semantic Network”, or the recommendations of the “OCLC Research Archives and Special Collections Linked Data Review Group”, to name a few. All of them are in the archival field and include formal documentary representations.

For all these reasons, better tools are needed for a deeper and faster understanding of the main concepts of the archival discipline. Ideally a holistic view, meaning by this that each set should be analyzed as a whole and not through its component parts considered separately. It is the set as an integrated whole which definitely determines how the parts behave, while a mere analysis of these parts cannot fully explain the performance of the whole; i.e.,

the whole is more complex than the simple sum of its constituent elements.

But this it is not all. Using semantic representations and taking advantage of the eventual levels of class, subclass, compound or simple element, etc. that are given in charts like this one, once completed this ontological representation would be more feasible to be programmed in a computer system in terms that can be recovered in more efficient ways through a records management system and/or the Web, for improved access and recovery of records. Trying to set the relationships, interdependencies, sub-components, etc., in a program based on a simple term definition glossary this would be extremely difficult, if not impossible; to program it all; departing from an ontology is much simpler. And, there would be another great advantage: if the ontology is analyzed by experts in the field, since it is relatively easy to make its equivalences in other languages. The full alternative Spanish version already accomplished is a complete exercise of this. This means that both versions are shareable for different communities with a single set of knowledge. If different language versions are feasible to be programmed in one computer system, or in two or more interconnected computer systems, search engines could retrieve the same items, regardless of the language prompted, with exact equivalence to the terms assigned unambiguously.

This is particularly notorious nowadays when using any general on-line translator or dictionary with some specific words –for example archival “record”– where there is no exact term to translate it into many non-English speaking countries. Formally, they had to cast a specific composite term for that purpose: therefore in Spanish, Portuguese, Italian and French –among other languages– the term has been agreed as a composed term close to “archival document”,³ in order to make a difference with definitions of

3 Spanish = “documento de archivo”. Portuguese = “documento arquivístico”. Italian = “documento archivistico”. French = “document d’archiv”. InterPARES Terminology Database (2012). http://www.interpares.org/ip3/ip3_terminology_db.cfm?letter=r&term=41

a simple “document”. Current on-line translators and dictionaries usually translate wrongly these terms from English to those languages and vice versa. Equally, the term “creator” has always an automatic translation into Spanish “creador”, which is not correct; it cannot be an automatic dictionary equivalence. In Archival Science the right equivalence of “creator” in Spanish is “productor”, which in English should be “producer”. Exactly the same happens in Portuguese: “creator” is always on-line translated into “criador”, while the right archival term in that language is “produtor”; and so on.

With ontologies, there is no confusion between concepts due to a simple translation program mistake, as the ontology, when programmed, establishes the exact terms which are equivalent and those who are not. As another example, with the ontology there will never be a confusion about what exactly the term “title” does mean from the point of view of digital records, despite what all the general dictionaries can say about it. The bilingual ontologies allow on one hand to understand the terms to be explained and contextualized in its own language, and on the other hand enables search engines or international systems to have the correct equivalents in the second language. And of course, bilingual can be converted into multilingual. Again, this is not a simple process of dictionary translation: it is a cognitive process, with coherent order, definition, experience, hierarchy and structure. The terms of the ontology are not mere literal translations, and above all they are not to be delegated to a computer while being constructed. Furthermore, in archival science –as in many other disciplines– neologisms are created daily in many languages, and like all new words or new meanings of existing ones, they must be properly weighted before being translated and introduced successfully in another language in this process. The simple translation, randomly made without semantics leads to the appearance of multiple versions of terms which latter will be hard to standardize. Semantic representations as ontologies allow both to individuals and computer programs correct translations, and facilitates establishing when certain terms are equivalent or synonymous and when they are not.

The ontology proposed here is only one example. Since it is an arbitrary representation of knowledge, not necessarily is the only one which is possible. Obviously, there may be more than one ontology to represent the same knowledge: it depends on who makes it, the intended approach, the depth, the emphasis, interest, grouping, scope, etc. This is not an ecumenical representation of a certain reality; it is only one of the possible representations of such reality under a certain approach. In fact, emphasis or “approach model” starts to define each possible knowledge representation. This is a personal example developed departing from the aforementioned InterPARES original approach, with some ontological analysis. It is obviously not the only possible one.

Therefore, a great advantage of ontologies thus developed is that each community can build them according to their interests, or can take any of the existing ones and adapt it to their context and perfect it at their convenience. This last option, in addition to saving a lot of effort, would help the ontologies already created on archival science to standardize knowledge throughout the communities of this discipline while stabilizing terms and concepts, which is particularly useful in these times of change and innovation, and in the long-term will eventually facilitate the exchange of data and digital information in this field. It is not desirable to go to extremes and have “one and only one” ontology for a concept as a universal definition, but certainly it is desirable to try to generalize the use of existing ones along archival communities and organizations for the purpose of standardizing certain fundamental concepts as it has already been done in other disciplines.

This is just an example of what semantic representations can achieve with a higher quality of construction. Being semantically rich and using also suitable and better metadata, records can be better defined and linked to other similar documents to optimize its search, integration, reuse, preservation, data mining and interoperability through better functional descriptions of their contents, boosting better interactions with external applications to exploit them properly. Then, those applications could solve better well-defined problems through well-defined operations to be

carried out on existing well-defined data. This is particularly useful currently, as many governmental, cultural heritage institutions, private organizations, etc., who are producing huge amounts of digital records are highly concerned about optimal use and long-term preservation of their documents for cultural and scientific heritage, legal purposes or economic competitiveness.

The principles and recommendations from the semantic proposals contained herein are not yet fully described or closed; on the contrary, they have only just begun their development and they are evolving increasingly every day. This is only one example of what can be done and why. Currently there are good foundations about the ways and actions to follow, but most of them remain to be researched, discovered and implemented. To determine which type of contextual knowledge is required to develop on digital records and documents, as well as their hardware, programs, formats, standards, metadata, etc., is a crucial step that is just beginning. Decisions taken in the short term from those responsible for the archival community will be of the utmost importance in this regard in the long term. As it can be inferred, semantic and pragmatic aspects of digital materials are not just an academic whim; they are rather essential elements to enable users to access, understand and use of digital records in the near future. Hardware, software, formats and networks can be closely interconnected with a better representation of knowledge and more sophisticated management models to enable users with faster, timely and more relevant access to information and its proper presentation.

To the extent that we can comprehend and contribute to these new representations we will develop concepts, standards, rules and technologies which help systems to better “understand” digital information so they can better deal with it for achieving improved retrievals, and even more: using semantic representations and not only more accurate results in the search for information can be obtained, but also better integrated information from different sources, as well as ensuring optimal preservation of that information, and thus aspire to a whole new and modern kind of automated services related with organizations with ampler archival responsibilities.

Bibliographic references

(All electronic references have been verified as accurate and extant as of July 24, 2019).

Arakawa, Naoya (2008). Semantic Analysis based on Ontologies with Semantic Web Standards. In: *Computer-aided Acquisition of Semantic Knowledge (CASK)*, 2. <http://www.celta.paris-sorbonne.fr/anasem/papers/Theory/Arakawa-CASK2008.pdf>

Arano, Silvia (2005). Thesauruses and Ontologies, 3. In: *Hipertext.net - Anuario Académico Sobre Documentación Digital y Comunicación Interactiva*, num. 3 <http://eprints.rclis.org/8972/2/12.pdf>

Baader, Franz; Horrocks, Ian; Sattler, Ulrike (2004). Description logics. In: *Handbook on Ontologies*, 3-28.

Berners-Lee, Tim (2001). Conceptual Architecture of the Semantic Web. W3C – World Wide Web Consortium Website <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>

Borst, Willem (1997). *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. Enschede, The Netherlands: W.N. Borst, (Ed.), 20. <http://doc.utwente.nl/17864/1/t00000004.pdf>

Chandrasekaran, Balakrishnan; Josephson, John; Benjamins, Richard (1999). What are Ontologies? Why do we need them? In: *IEEE Intelligent Systems*, vol. 14, num. 1, 20-26 <http://www.csee.umbc.edu/courses/771/papers/chandrasekaranetal99.pdf>

DLM Forum - Document Lifecycle Management Forum (1997). *Guidelines on best practices for using electronic information*, section 2.4. Luxembourg: European Commission. <https://dmlforum.typepad.com/gdlines.pdf>

- Duranti, Luciana (2003). More than Information, Other than Knowledge: The Nature of Archives in the Digital Era, 10. In: *Cadernos Bad*, Lisboa, Portugal, num. 2, 6-16, <https://www.bad.pt/publicacoes/index.php/cadernos/article/view/845>
- Duranti, Luciana; Thibodeau, Kenneth (2006). The Concept of Record in Interactive, Experiential and Dynamic Environments, In: *Archival Science*, num. 6, (2006), 13-68. DOI: 10.1007/s10502-006-9021-7.
- European Commission (2008). *MoReq2 - Model Requirements for the Management of Electronic Records*, 2nd version, Glossary Section. Brussels: European Commission. https://cdn.ymaws.com/irms.org.uk/resource/resmgr/moreq/moreq2_typeset_version.pdf
- Green, Peter; Rosemann, Michael (2004). Editorial Preface: Ontological Analysis, Evaluation and Engineering of Business Systems Analysis Methods, In: *Journal of Database Management*, num. 15, April/June (2004), i-ii.
- Gruber, Thomas (1995). Toward Principles for the Design of Ontologies Used for Knowledge Sharing, In: *International Journal of Human-Computer Studies*, vol. 43, num. 5-6, 1995, 907-928 <https://doi.org/10.1006/ijhc.1995.1081>
- Guarino, Nicola (1998). *Formal Ontology in Information Systems*. Amsterdam, The Netherlands: IOS Press: N. Guarino (Ed.), 3-15. Amended version of a paper appeared in N. Guarino (Ed.), *Formal Ontology in Information Systems*, "Proceedings of FOIS'98", Trento, Italy, June 6-8, 1998 <http://dl.acm.org/citation.cfm?id=521669>
- Guarino, Nicola; Giarretta, Pier-Daniele (1995). *Ontologies and Knowledge Bases: Towards a Terminological Clarification*. Trento, Roma: Laboratory for Applied Ontology. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.320.8006&rep=rep1&type=pdf>

- ICA – International Council on Archives (2014). Multilingual International Terminology Database, Entry by: “record” <http://www.ciscra.org/mat/mat/term/60>
- InterPARES Terminology Database (2012). Entries by ‘record’ and ‘digital record’: http://www.interpares.org/ip3/ip3_terminology_db.cfm?letter=r&term=41 and http://www.interpares.org/ip3/ip3_terminology_db.cfm?letter=d&term=222
- ISO – International Standards Office (2001). Standard ISO 15489-1:2001. Information and Documentation – Records Management – Part 1: General. Geneva, Definition 3.15.
- Milton, Simon (1998). Top-Level Ontology: The problem with Naturalism. In: *Formal Ontology in Information Systems*, N. Guarino (Ed.) (1998). Amsterdam: IOS Press, pp. 85-94.
- Neches, Robert; Fikes, Richard; Finin, Timothy; Gruber, Thomas; Patil, Ramesh; Senator, Edward; Swartout, William (1991). Enabling Technology for Knowledge Sharing, In: *Artificial Intelligence Magazine*, vol. 12, num. 3. Menlo Park, CA: American Association for Artificial Intelligence Press, 36-56.
- Noy, Natalia; McGuinness, Deborah (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, (March 2001). <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf>
- Pearce-Moses, Richard (2005). *A Glossary of Archival and Records Terminology*. Chicago: The Society of American Archivists. <https://dictionary.archivists.org/entry/record.html>

- Souza, Renato; Tudhope, Douglas; Almeida, Mauricio (2011). Towards a taxonomy of KOS: Dimensions for classifying Knowledge Organization Systems. International Society for Knowledge Organization (ISKO), From *11th ISKO International Conference*, Rome (Italy): 23-26 February 2010.
- Steve, Geri; Gangemi, Aldo; Pisanelli, Doménico (1998). Integrating Medical Terminologies with ONIONS Methodology. *Reperto Informatica Medica*, Istituto Tecnologie Biomediche, CNR, Italy. <http://www.loa.istc.cnr.it/old/Papers/onions97.pdf>
- Swartout, William; Patil, Ramesh; Knight, Kevin; Russ, Thomas (1996). *Toward Distributed Use of Large-Scale Ontologies*. University of Southern California <https://www.aaai.org/Papers/Symposia/Spring/1997/SS-97-06/SS97-06-018.pdf>
- Uschold, Michael; Grüninger, Michael (1996). Ontologies: Principles, Methods, and Applications. In: *Knowledge Engineering Review*, vol. 11, num. 2, Cambridge Journals On-line, 93-155.
- Wand, Yair; Weber, Ron (2004). Reflection: Ontology in information systems, In: *Journal of database management*, vol. 15, num. 2, iii-vi.
- W3C – World Wide Web Consortium (2010). OWL - Web Ontology Language Overview. Official Consortium Website <http://www.w3.org/TR/owl-features/>

El concepto de documento de archivo digital: una propuesta de ontología / The concept of digital record: an intended ontology. Instituto de Investigaciones Bibliotecológicas y de la Información/UNAM. La edición consta de 100 ejemplares. Coordinación editorial, Anabel Olivares Chávez, revisión especializada: Valeria Guzmán González y Francisco González y Ortiz; corrección de pruebas: Carlos Ceballos Sosa; formación editorial, Mario Ocampo Chávez. Fue impreso en papel bond blanco de 90 g en los talleres de Litográfica Ingramex, Centeno 162-1, Col. Granjas Esmeralda, Alcaldía Iztapalapa, CDMX, C. P. 09810. Se terminó de imprimir en mayo de 2022