

LA “AUTENTICIDAD” EN DOCUMENTOS DE ARCHIVO DIGITALES: UNA ONTOLOGÍA

Juan Voutssas Márquez*

En los últimos años, y con el advenimiento mundial de la enorme expansión de información digital, las comunidades de archivos y bibliotecas digitales han hecho valiosas aportaciones a la gestión de enormes cantidades de datos e información, cada una desde sus diferentes perspectivas y enfoques. Sus correspondientes disciplinas combinadas con otras disciplinas emergentes aportan además una contribución fundamental a la innovación del conocimiento para garantizar el acceso permanente a los documentos digitales. Una de estas disciplinas emergentes que propone innovadoras soluciones técnicas para la representación y gestión del conocimiento es el área de la *web semántica*, que se apoya en las recomendaciones del *World Wide Web Consortium –W3C–* y cuyo propósito es lograr un mejor acceso a la información por parte de una gran cantidad de instituciones y usuarios a través de las Tecnologías de Información y Comunicaciones –TIC– con la ayuda de la *web*. El uso de *representaciones semánticas*, establecidas por esa disciplina en combinación con los principios archivísticos modernos para el tratamiento de los documentos de archivo digitales de un cierto patrimonio documental puede ser una gran solución para mejorar la descripción documental y el acceso en los sistemas de información archivística. A este nuevo conjunto de principios y propuestas para la descripción formal documental se le ha empezado ya a denominar *representaciones semánticas* y a los archivos con ellas construidos *archivos semánticos*.

La idea central es que con mejores *representaciones semánticas* de los documentos de archivo digitales se pueden crear dentro de los sistemas de gestión archivística *nuevas representaciones formales* de los atributos de esos documentos que se desea mantener sin cambios durante diferentes generaciones, lo cual ayudaría en un futuro a diseñar e implementar mejores

* Instituto de Investigaciones Bibliotecológicas y de la Información, UNAM.

estrategias de migración y/o transferencia. En los archivos, la forma debe ser fija y el contenido estable, pero las tecnologías no lo serán. Sabemos que uno de los principales problemas en la actualidad y en el futuro para el mantenimiento de los documentos de archivo digitales consiste en cómo hacer frente a las tecnologías cambiantes: equipos y dispositivos, programas, sistemas operativos, formatos, etcétera. Todos estos elementos tienen un impacto directo y modifican el uso, gestión y preservación de esos documentos de archivo; por lo tanto, es necesario combinar los conocimientos archivísticos actuales con la formalización adecuada de la información bajo estos nuevos enfoques semánticos, no sólo como una solución para una mejor gestión de ellos, sino para su mejor preservación ulterior.

Las *representaciones semánticas* permitirán también a la larga una mejor preservación de los documentos de archivo ya que con ellas se asegura de mejor forma la autenticidad de los contenidos a lo largo de generaciones tecnológicas de equipos, programas y formatos. Considerando además que los documentos de archivo digitales que sean transferidos a archivos históricos raramente serán descartados de sus sistemas de información, el número de ellos crecerá de manera muy significativa dificultando con el tiempo su adecuada y oportuna recuperación. Estas representaciones podrían ayudar a un acceso eficiente a pesar de su creciente volumen.

Si bien la disciplina semántica aporta la metodología para hacerlo, será muy necesario que sean las propias comunidades archivísticas quienes formalicen con su ayuda la descripción de esos atributos relevantes en una representación apropiada de los documentos de archivo, ya que tienen la experiencia y el conocimiento sobre el tema. Por ello es importante que los profesionales de los archivos comiencen a conocer estas propuestas. Bajo las anteriores premisas, los atributos de los documentos de archivos digitales deben ser apropiadamente identificados y representados. Obviamente esto se hace hoy en día, pero no bajo este enfoque “semántico”.

Hemos mencionado que los *archivos semánticos* se basan en los principios de la “*Web Semántica*”. En la *Web* actual, básicamente, la recuperación se hace mayormente por parte del programa buscador el cual trata de ubicar dentro de las páginas *web* las palabras que hemos introducido en él, –tanto palabras contiguas como sueltas–, y va encontrando una palabra en una

página y otra palabra en distinta página; si acaso, palabras juntas en una misma página. Por ello entrega las innumerables cantidades de páginas recuperadas de las cuales la inmensa mayoría es irrelevante y de donde el acceso a la información se convierte en una tarea difícil y frustrante. La recuperación actual se hace por medio de palabras, pero prácticamente sin contexto, sin significado para los buscadores: en resumen, sin *semántica*.

En una explicación sencilla, la *web semántica* no es una nueva marca o modelo de *navegador*; no es un producto que se compre: es un concepto más amplio que se construye colectivamente. Pretende ser una *web* extendida, dotada obviamente de mayor significado para que cualquier usuario dentro de ella pueda encontrar respuestas más certeras y pertinentes a sus preguntas de forma más rápida y sencilla gracias a un conjunto de “componentes”, pero en especial, a una información mejor pre-definida, con contexto, con semántica. Esa información pre-definida de manera formal, normalizada y universal, es precisamente la que conforma las *representaciones semánticas* a las que ya hemos hecho alusión. Gracias a ellas, podremos dotar a la *web* de más significado para sí misma, para sus programas y para sus buscadores; tendrá por lo tanto más eficiencia, y debido a ello se podrán obtener mejores soluciones a problemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual será posible compartir, procesar y transferir información de forma sencilla.

Para lograr ese propósito, la *web semántica* ha introducido toda una estructura conceptual conformada por ciertos componentes, algunos que ya existían y otras nuevas “herramientas” cuyas primeras versiones de prueba ya se han empezado a definir y utilizar. Hay componentes tecnológicos, conceptuales, accesorios, etcétera. Véase el diagrama del anexo 1 para observar su estructura conceptual.

Podemos ver que existen capas “tecnológicas” que tienen que ver más con equipos, programas, redes, etcétera, y podemos ver que hay capas “conceptuales”, tales como las del *XML*, el *RDF* y el *OWL*, estándares que pretenden ayudar a convertir la *web* en esa infraestructura global en la que es posible compartir y reutilizar datos y documentos entre diferentes tipos de sistemas y usuarios. Como puede observarse, es una estructura compleja de la cual hoy queremos analizar exclusivamente la derivada de “*OWL*” –*Web Ontology Language*–, una herramienta para desarrollar vocabularios específicos

para asociar esos recursos formalmente. Lo que hace OWL es proporcionar un lenguaje para definir *ontologías* estandarizadas que pueden ser utilizadas a través de diferentes sistemas. ¿Qué son estas *ontologías*? ¿Cuál es su utilidad en la ciencia archivística moderna?

El término *ontología* ha sido utilizado tradicionalmente durante siglos por la filosofía, –en particular por la metafísica–, y tiene que ver con la ciencia de lo que “es”, del “ente”, así como de sus propiedades trascendentales; su esencia es el estudio del “ser” en lo que “es” y en “cómo es”. En tiempos más recientes, el término ha sido adoptado por diversas disciplinas y pretende además de definir al ser, establecer las categorías fundamentales de las cosas a partir del estudio de sus propiedades, sistemas y estructuras. Por tanto ha devenido en nuevos y diversos conceptos.

Nicola Guarino y Pierdaniele Giaretta¹ llevaron a cabo en 1995 un estudio al respecto del cual se desprenden siete enfoques de los que el concepto de *ontología* tiene sendas interpretaciones: 1) El “tradicional”, ontología como disciplina filosófica. 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 un teoría lógica, caracterizada ya sea por: las propiedades específicas formales, o bien sólo por sus fines específicos. 6) Ontología como el vocabulario utilizado por una teoría lógica. 7) Ontología como una especificación o meta-nivel de una teoría lógica.

Estas interpretaciones tienen diversas utilidades para distintas disciplinas, además de la filosofía: la lingüística y la literatura, la inteligencia artificial, la informática, la teoría del conocimiento, etcétera; y como veremos, también puede tener utilidad para la ciencia archivística.

Robert Neches² estableció que: “una ontología define los términos básicos y las relaciones que componen el vocabulario de una cierta área temática, así como las reglas para la combinación de términos y relaciones para definir extensiones para ese vocabulario”.

Según Swartout una ontología es: “un conjunto de términos estructurados

1 Guarino, Nicola; Giaretta, Pierdaniele, *Ontologies and Knowledge Bases: Towards a Terminological Clarification*, en: <http://www.loa-cnr.it/Papers/KBKS95.pdf>

2 Neches, Robert *et al.*, en: *Artificial Intelligence Magazine*, pp. 36-56.

que describen un tema o ‘dominio’. Su propósito es proporcionar una estructura básica para una base de conocimientos”.³

Para Uschold:

[...] una ontología es un conjunto de agrupamientos o ‘clases’ con diversos tipos de relaciones entre ellas. La principal ‘clase’ es aquella que describe la taxonomía⁴ de la ontología y es su columna vertebral. Las relaciones tienen propiedades cuya mayor utilidad es la de poder hacer inferencias a partir de ellas acerca de las clases... La ontología es el término utilizado para referirse a la comprensión compartida y puede ser utilizada como un marco unificador para resolver los problemas... Una ontología necesariamente implica o representa algún tipo de visión del mundo con respecto a un determinado dominio o tema. El mundo visto a menudo se concibe como un conjunto de conceptos (por ejemplo, entidades, atributos, procesos), sus definiciones y sus interrelaciones, lo que se conoce como una conceptualización.⁵

Para Silvia Arano

[...] una ontología es una representación formal y explícita de la estructura conceptual de un campo de conocimiento. La ontología es el soporte semántico de las palabras que se describen como objetos lingüísticos en una base de datos léxica o terminológica. Las relaciones conceptuales representadas en una ontología son muy variadas y dependen del campo de conocimiento a estructurar. Una ontología se construye con el objetivo de compartir y reutilizar la información almacenada que, después de haber sido formalizada, puede ser interpretada tanto por personas como por programas informáticos.⁶

3 Swartout, William, *et al.*, en: http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html

4 “Una taxonomía es un vocabulario controlado que se organiza en una jerarquía. Cada término designa una categoría, tipo o clase. Sólo hay un tipo de vínculo, que significa ‘es una variedad de’ y corresponde a una relación de subclase. Hablando estrictamente cada nodo en una taxonomía tiene exactamente un ‘padre’, pero el término ‘taxonomía’ a menudo se refiere a las jerarquías con múltiples ‘padres’. A veces también se utiliza para referirse a las redes con más de un tipo de enlace”. Uschold, Michael, *Ontologies: Principles, Methods, and Applications*, p.

5 *Ibid.*, pp. 93–155.

6 Arano, Silvia. “Thesauruses and Ontologies” en: *Hipertext.net - Anuario Académico Sobre Documentación Digital y Comunicación Interactiva*.

Para la informática –de manera breve y sencilla–, básicamente el término “ontología” es la formulación programable de un vocabulario específico para describir el conocimiento común acerca de una cierta realidad en un campo del conocimiento dado. En la terminología del medio se acostumbra denominar con frecuencia “artefactos” a los conceptos y al vocabulario específico, y “dominio” al área temática o campo del conocimiento que se pretende definir. Más específicamente, el “artefacto” por lo general consiste en una estructura conceptual muy completa y exhaustiva dentro de ese campo del conocimiento o “dominio”, y se construye con la finalidad de facilitar la comunicación y la compartición de la información entre usuarios, bases de datos y sistemas informáticos dentro de ese “dominio”. Es decir, su aplicación pretende ir más allá de una organización o región: toda una comunidad específica interesada en ese campo de conocimiento.

Un uso común tecnológico actual de este concepto de ontología lo encontramos en la inteligencia artificial y en otras aplicaciones donde este tipo de estructuras conceptuales tienen particular relevancia. Por ejemplo, tienen una gran aplicación en la representación del conocimiento y un impacto directo en la construcción de algoritmos y programas para el almacenamiento y recuperación de información.

Bajo estos particulares enfoques de nuestro interés, como conclusión puede establecerse que las ontologías son entidades construidas que permiten describir formalmente con un vocabulario el conocimiento común y compartido sobre un cierto tema.

Su importancia radica en que con ellas podemos lograr en un futuro compartir información entre personas, máquinas y programas de una manera unificada y coherente, la cual llevará a una recuperación más precisa, pertinente y oportuna de esa información.

Al igual que los tesauros, las ontologías tienen que ser construidas, una por una, para un cierto campo o área del conocimiento, –un “dominio” como ya mencionamos–: la microbiología, la astronomía, la bibliotecología, la ciencia archivística, etcétera. Y como cada de uno de esos campos a su vez es muy extenso, deben construirse ontologías especiales más detalladas para sub-campos o áreas específicas de ese ámbito del conocimiento. Ésa es una labor profundamente intelectual, no delegable a máquinas, –al menos no por ahora–. Y como mencionamos, es recomendable que sean hechas

por expertos en el “dominio” o tema a desarrollar. Por lo mismo, en particular para la ciencia archivística deben ser construidas no una, sino toda una serie de ontologías en campos específicos dentro de esta gran área del conocimiento. Y es deseable que sean construidas por profesionales de la archivística.

Parte del problema de esta construcción consiste en que no sólo hay que compilar y sistematizar el conocimiento común acumulado acerca de una cierta área, sino además integrar el nuevo conocimiento y las características emergentes que impone el advenimiento de desarrollos tecnológicos, en especial el desarrollo y auge de los documentos de archivo digitales.

Así, por ejemplo, tenemos el conocimiento común adquirido alrededor de la “autenticidad” de un documento de archivo, requisito indispensable para su gestión y preservación. Durante mucho tiempo, la diplomática ha tratado de establecer esto con respecto a los documentos de archivo “tradicionales” en papel, pero es necesario replantear este conocimiento a la luz de los documentos de archivo digitales. Varios intentos se han hecho al respecto. Por ejemplo, el proyecto internacional de investigación archivística InterPARES⁷ desarrolló un glosario de terminología en preservación de archivos digitales, poniendo especial énfasis en estos atributos y características “modernos” impuestos por el desarrollo y uso de documentos de archivo digitales. Con el propósito de ir más allá de un glosario, el proyecto elaboró unas “ontologías” básicas así como documentos complementarios acerca de algunos temas de particular interés y que tienen que ver con lo que hoy nos ocupa. En especial, es de nuestro interés la denominada “Ontología C”: “confianza de un documento de archivo”.

Sabemos que en cada organización que maneja documentos de archivo tenemos que establecer sus *requisitos de autenticidad*; esto es, las especificaciones de los elementos de forma y contexto que necesitan ser preservados para mantener la autenticidad de cierto tipo de documento de archivo digitales. También debemos establecer los *requisitos básicos de autenticidad*⁸ y los *requisitos de referencia de autenticidad*.⁹ Toda organización

7 InterPARES = The International Research on Permanent Authentic Records in Electronic Systems o *Investigación Internacional Sobre Documentos de Archivo Auténticos y Permanentes en Sistemas Electrónicos*.

8 Condiciones mínimas necesarias para permitir al preservador avalar la autenticidad de las copias de los documentos de archivo digitales de un productor y que se encuentran bajo su custodia.

9 Condiciones que sirven al preservador como base para la evaluación acerca de la autenticidad

emprende esta tarea por diversos caminos, tratando de definir todos y cada uno de los elementos que le permitan asumir esa autenticidad. Las fuentes para esas definiciones son por lo general diccionarios de archivística. Si comparamos entre ellos sus definiciones de “autenticidad” veremos que son semejantes pero no idénticas; algunas de ellas incluyen elementos que no se encuentran en otros, etcétera.

Por ejemplo: “autenticidad: condición indispensable en cualquier documento de archivo de cualquier época, que supone que un documento está revestido de formalidades precisas; está redactado por quien dice ser su autor y en la fecha consignada, que el contenido responde a la voluntad de quien lo valida y no ha sufrido alteración”.¹⁰

[...] autenticidad: 1) cualidad de auténtico, referida al documento, aquel que está acreditado de cierto y positivo por los caracteres, requisitos o circunstancias que en él concurren, tales como la estructura de la información, la presencia de firmas, sellos, fórmulas de redacción y los demás elementos diplomáticos necesarios. 2) Inmunidad ante todo tipo de alteración, manipulación o falsificación; cualidad por la que se puede probar que un documento de archivo es lo que pretende ser, ha sido producido o enviado por la persona y en la fecha que se presumen.¹¹

Si analizamos estas definiciones, muy actuales, e inclusive agregáramos más, podemos concluir que son correctas, se parecen mucho y nos dan una buena idea del término. Pero podemos plantear una serie de cuestiones adicionales. ¿Cuáles son los componentes de la autenticidad? ¿Tienen éstos a su vez sub-componentes? ¿La autenticidad es a su vez parte de algo superior? ¿Cuáles son las relaciones y/o condiciones para esos sub-componentes? Contestar estas interrogantes no sería imposible, pero para hacerlo tendríamos que meditar y analizar mucho; tal vez pergeñar una y otra vez algunas páginas, cuadros o tablas hasta lograr algún resultado lógico y coherente. Es un hecho que podemos partir de lo que dicen los diccionarios, pero no podemos contestar directamente estas preguntas a partir de ellos.

de los documentos de archivo digitales durante la valoración, especialmente para aquellos que se pretende sean transferidos a un archivo histórico.

¹⁰ Heredia, Antonia. *Lenguaje y vocabulario archivísticos: algo más que un diccionario*, p. 52.

¹¹ Cruz Mundet, José Ramón, *Diccionario de Archivística*, p. 32.

Lo anterior es debido a que los glosarios sólo están hechos para definir términos sueltos, pero no dentro de un contexto o estructura que además de definir los términos, los interrelacione, los jerarquice, e indique las condiciones que deben darse para combinar o llegar a esos términos. Es decir, se consultan diccionarios, pero no una ontología. Y la razón principal de esto es simplemente porque prácticamente no existen en el campo de la archivística, y menos en la archivística digital.

Es por ello que el proyecto InterPARES decidió realizar unas ontologías básicas, con el propósito de comenzar a experimentar en el establecimiento de este tipo de herramientas –o “artefactos”, si usamos la terminología del medio– y de esta forma comenzar este camino en el desarrollo de estas herramientas más modernas para la representación del conocimiento archivístico y, –por supuesto–, desarrollado por comunidades archivísticas profesionales.

ONTOLOGÍA “C”: CONFIANZA DE UN DOCUMENTO DE ARCHIVO
(trustworthiness)



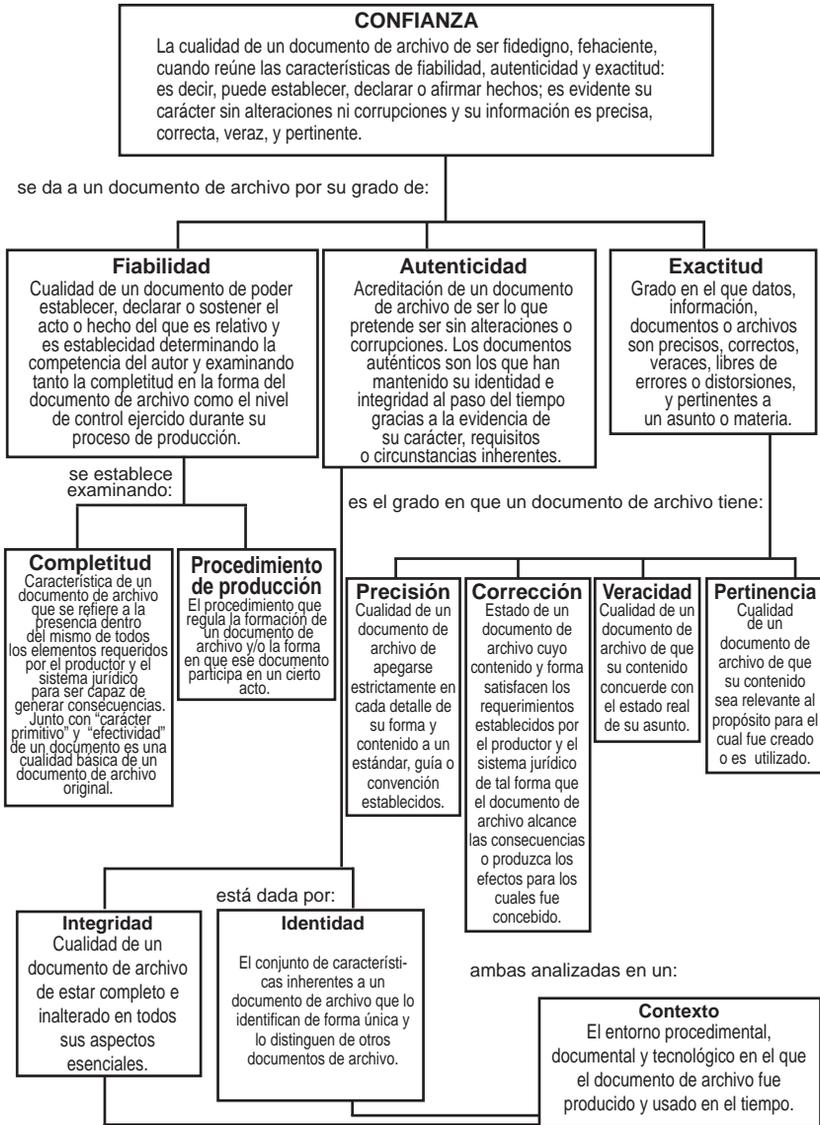
Si observamos esta representación, podemos inferir de ella algunos puntos interesantes:

En primer lugar, nosotros siempre hemos hablado de la *autenticidad* de un documento como atributo indispensable para su preservación, pero vemos que la autenticidad por sí misma no es suficiente; es sólo uno de tres elementos que conforman la *confianza* de un documento de archivo. Podemos ver de ello que en realidad debemos buscar la *confianza* de un documento, la cual “se da a un documento de archivo por su grado de” fiabilidad, autenticidad y exactitud. Dicho de otra forma, la autenticidad no es un fin en sí misma; es un medio para buscar y lograr la *confianza* en un documento de archivo. Comenzamos así a ver *componentes* y *sub-componentes*, lo cual nos empieza a dar mayor luz sobre el concepto.

Podemos ver además que cada uno de los tres componentes de la confianza tienen sub-componentes y condiciones o relaciones: la *fiabilidad* “se establece examinando” la *completitud* del documento y su *procedimiento de producción*; la *autenticidad* “está dada por” la *integridad* y la *identidad* del documento analizadas en un contexto, y finalmente la *exactitud* “es el grado en que un documento de archivo tiene” *precisión*, *pertinencia*, *veracidad* y *corrección*. Ahora sí, podemos ver todos los elementos que están en cuestión, sus relaciones, ascendientes, descendientes, condiciones, etcétera.

Si agregamos además a esta representación las definiciones de los términos, de forma adyacente o embebida, tenemos entonces una representación estructural completa de la “autenticidad”, tenemos de manera contigua e integral el conocimiento común y compartido sobre este “dominio” o tema. Una “ontología de la confianza de un documento de archivo”, la cual es más completa que sólo la definición de “autenticidad” y que por supuesto, la contiene.

ONTOLOGÍA “C”: CONFIANZA DE UN DOCUMENTO DE ARCHIVO
(*trustworthiness*)



El análisis detallado de la misma nos puede dar más y más información, contexto y respuestas conforme lo estudiemos. Por ejemplo, muchas personas piensan que “exactitud” y “precisión” de un documento de archivo son sinónimos o –sabiendo que no lo son– les cuesta mucho describir la diferencia. De esta ontología, podemos ver que no son lo mismo, que la precisión es un sub-componente de la exactitud y que ninguna tiene una relación directa con la *autenticidad*, sino con la *confianza*. Las definiciones precisas están ahí, pero más importante aún, el contexto y las interrelaciones.

Aparte de la utilidad que este tipo de representación tiene para la comprensión de los términos, podemos ver que esto nos puede ayudar mucho a establecer los *requisitos de autenticidad* para los documentos de archivo dentro de nuestra organización, así como todos y cada uno de sus componentes. Podemos entonces establecer más claramente requisitos para salvaguardar integridad, identidad, precisión o completitud de un documento. Esto facilitará a la larga su manejo y ulterior preservación, toda autenticidad guardada.

Útil como es lo anterior, no es todo; volvamos al punto inicial de nuestra reflexión: los “*archivos semánticos*”. La ventaja informática para la recuperación y acceso a unos documentos de archivo así guardados es que esta representación ontológica sí puede ser programada en un sistema para que los términos puedan ser recuperados de manera más eficiente vía un programa de *gestión archivística* y/o la *web*. Programar las relaciones, interdependencias, componentes, sub-componentes de los términos a partir de las definiciones de diccionario resulta muy difícil, por no decir imposible; programar todo eso a partir de la ontología es mucho más simple. Existe además otra gran ventaja adicional: si es hecho por expertos en el tema, es relativamente fácil hacer equivalencias de las ontologías en otros idiomas. De hecho, la ontología aquí presentada fue hecha originalmente en inglés, y luego transformada al castellano. Ello hace que puedan programarse ambas en un sistema, con lo cual los buscadores semánticos encontrarán lo mismo independientemente del idioma en el que se pregunte, con las equivalencias exactas asignadas a los términos, sin ambigüedades. Por ejemplo, *exactitud* tendrá el mismo peso que “*accuracy*”, su equivalente exacto con la ontología en inglés; no podrá ser confundida con “*precisión*” por cualquier programa traductor, ya que la ontología al ser programada hace las equivalencias

exactas de los términos que son y los que no. Las ontologías bilingües nos permiten por un lado, ayudar a la comprensión de los términos al ser explicados y contextualizados en nuestra lengua, y por otro lado mantiene a la vez la accesibilidad total por parte de buscadores o sistemas internacionales al tener los equivalentes correctos en inglés. Este no es un simple proceso de traducción de diccionario: es un proceso cognoscitivo, de orden y definición, de estructura coherente. Los términos de la ontología no son simples traducciones, y mucho menos encargadas a una máquina. Ahí está el tristemente célebre ejemplo de *record* por “registro” en vez de por “documento de archivo”. Éstas son las limitaciones de los sistemas tradicionales de representación y recuperación de información a las que me refería en un principio y por ello estas nuevas soluciones pueden ayudar a esos programas.

Es necesario subrayar en este momento que la ontología presentada aquí es sólo un ejemplo; es una cierta representación de un conocimiento, pero no necesariamente es única. Puede haber más de una ontología para representar el mismo conocimiento; depende de quién la hace, los enfoques, el énfasis, los intereses, etcétera. No es una representación ecuménica de una cierta realidad; es sólo una de las posibles representaciones de esa realidad. De hecho, el énfasis o “modelo” de aproximación define de inicio cada posible representación del conocimiento. Continuando con el ejemplo de la “autenticidad” como “dominio” o tema a definir, podemos encontrar varios enfoques: OAI¹² hace énfasis en la *integridad a través de los metadatos*, RLG-NARA lo hace en la *confianza por medio de la certificación*¹³ e InterPARES, como ya analizamos, lo hace reforzando la autenticidad por medio del control del proceso y los atributos. Y éstos no son los únicos enfoques posibles,

Por lo anterior, una gran ventaja de las ontologías así desarrolladas consiste en que cada comunidad puede desarrollarlas de acuerdo con sus intereses, o bien tomar alguna de las ya existentes y adaptarla a su contexto. Esto último, además de que ahorra una buena cantidad de esfuerzo al

12 OAI = Open Archival Information System. Consultative Committee for Space Data Systems. 2002. *Reference Model for an Open Archival Information System (OAI)*. Washington, D.C.: CCSDS Secretariat.

13 RLG-NARA = Research library Group-National Archives and Records Administration. Disponible en: <http://www.crl.edu/PDF/trac.pdf>

respecto, ayuda a que las ontologías ya creadas sobre conceptos básicos de archivística vayan estandarizándose a lo largo de las comunidades de esta disciplina, lo que favorece la normalización y a través de ésta la estabilización de términos y conceptos, algo particularmente útil en estos tiempos de cambio e innovación, y favorecerá a la larga el intercambio de datos e información electrónica. No se trata de llegar al extremo de desear que exista “una y sólo una” ontología para un concepto, una definición universal –todos los extremos son malos– pero sí de generalizar el uso de cada una de ellas a lo largo de las comunidades y organizaciones archivísticas con el propósito de homogeneizar ciertos conceptos fundamentales, como se ha hecho ya en otras disciplinas.

Esto es sólo un ejemplo de lo que los documentos de archivo semánticos pueden lograr con una mayor calidad de construcción; al ser semánticamente ricos y con ayuda además de apropiados y mejores metadatos pueden ser definidos y ligados de tal forma con otros documentos similares que permitan su óptima búsqueda, integración, reuso, preservación, minería de datos e interoperabilidad por medio de mejores descripciones funcionales de sus contenidos, que garanticen mejores interacciones con programas propios y externos para que puedan explotarlos adecuadamente. Estos programas podrán así resolver problemas bien definidos, a través de operaciones bien definidas que se llevarán a cabo sobre datos existentes bien definidos.

Los principios y recomendaciones provenientes de la disciplina semántica aquí enunciados no están ya totalmente descritos ni cerrados. Todo lo contrario; apenas han comenzado su desarrollo y están evolucionando cada día. Hoy hemos estudiado sólo un ejemplo de lo que puede hacerse y por qué. Actualmente existen buenas bases acerca de los caminos y acciones a seguir, pero la mayor parte está aún por ser investigada, descubierta e implementada. Determinar qué tipo de conocimiento contextual sobre los archivos digitales y sus documentos, sus equipos, programas, formatos, estándares, normas, metadatos, etcétera, es un paso crucial que apenas está iniciando. Las decisiones que los responsables de las comunidades archivísticas tomen son fundamentales al respecto. Los aspectos semánticos y pragmáticos de los materiales digitales son cruciales para que los usuarios puedan ser capaces de acceder, comprender y usar los objetos digitales en el futuro.

Equipos, programas, formatos y redes pueden ser estrechamente interconectados con una mejor representación del conocimiento sofisticado y modelos de gestión para que las comunidades en el futuro se vean habilitadas con un acceso rápido, pertinente y oportuno y con su presentación más apropiada, multimedios inclusive.

En la medida en que podamos comprender y contribuir con estas nuevas representaciones podremos desarrollar conceptos, estándares, normas y tecnologías que ayuden a los sistemas a “entender” más la información digital para que puedan lograr mejores recuperaciones, y aún más: con ayuda de las representaciones semánticas podremos lograr no sólo recibir resultados más exactos en la búsqueda de información, sino también integrar mejor información de diferentes fuentes, lograr documentos integrados para comparar, asegurar una óptima preservación de los mismos y aspirar a toda una nueva clase de servicios automatizados relacionados con los quehaceres propios de las organizaciones archivísticas.

Referencias Bibliográficas:

(Todas las referencias electrónicas han sido verificadas como exactas y disponibles al 1 de junio del 2012.)

Arano, Silvia, “*Thesauruses and Ontologies*”, en: *Hipertext.net - Anuario Académico Sobre Documentación Digital y Comunicación Interactiva*, núm. 3, 2005. Disponible en: <http://www.hipertext.net/english/pag1009.htm>

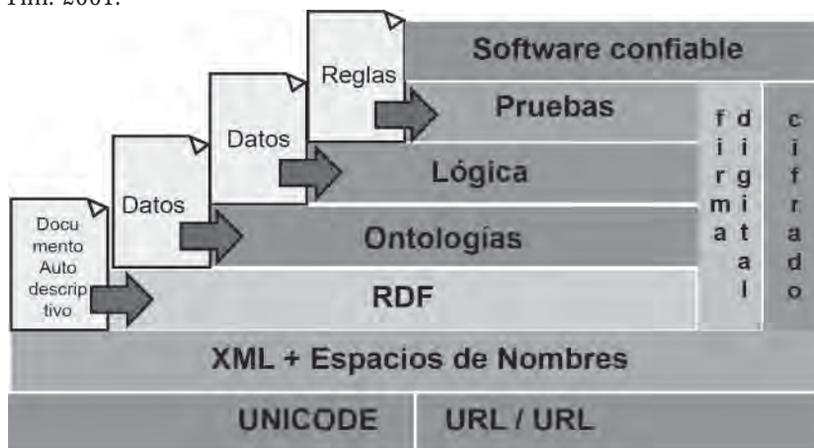
Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)*, Washington, D. C.: CCSDS Secretariat, 2002.

Guarino, Nicola; Giaretta, Pierdaniele. *Ontologies and Knowledge Bases: Towards a Terminological Clarification*. Trento, Roma: Laboratory for Applied Ontology, 1995. Disponible en: <http://www.loa-cnr.it/Papers/KBKS95.pdf>

InterPARES, “The International Research on Permanent Authentic Records in Electronic Systems” o “Investigación internacional sobre documentos de archivo auténticos y permanentes en sistemas electrónicos”, sitio web oficial del Proyecto InterPARES.

- Disponible en: <http://www.interpares.org>
- InterPARES. The International Research on Permanent Authentic Records in Electronic Systems. *Glosario de Preservación de Archivos Digitales* (traducción al castellano). Sitio *Web Proyecto Interpares México*, 2010. Disponible en: <http://www.interpares.org> y también en: <http://www.iibi.unam.mx/archivística>
- InterPARES. The International Research on Permanent Authentic Records in Electronic Systems. 2010. *Ontología 'C'* (traducción al castellano), sitio web oficial del Proyecto InterPARES. Disponible en: <http://www.interpares.org> y también en: <http://www.iibi.unam.mx/archivística>
- Neches, Robert; Fikes, Richard; Finin, Timothy; Gruber, Thomas; Patil, Ramesh; Senator, Edward & Swartout, William, "Enabling Technology for Knowledge Sharing", en: *Artificial Intelligence Magazine*, vol. 12, núm. 3, 1991. American Association for Artificial Intelligence Press, pp. 36-56. Menlo Park, CA, USA.
- RLG-NARA Digital Repository Certification Task Force, *Trustworthy repositories audit & certification: Criteria and checklist*, 2007. Disponible en: <http://www.crl.edu/PDF/trac.pdf>
- Swartout, William, *et al.* 1996. *Toward Distributed Use of Large-Scale Ontologies*, University of Alberta at Calgary. 1996. Disponible en: http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html
- Uschold, Michael, and Grüninger, Michael, "Ontologies: Principles, Methods, and Applications," en: *Knowledge Engineering Review*, vol 11, núm. 2, 1996. Cambridge Journals On-Line. pp. 93-155.
- W3C – World Wide Web Consortium. 2010. "*OWL - Web Ontology Language Overview*". Sitio Oficial del Consorcio. Disponible en: <http://www.w3.org/TR/owl-features/>

Anexo 1. Diagrama conceptual de la *web semántica*. Berners-Lee, Tim. 2001.



Unicode es una codificación de textos que permite utilizar y visualizar en pantalla los caracteres de diferentes idiomas. Esto permite expresar información para la recuperación dentro de la *web semántica* en cualquier alfabeto.

URI son las siglas de *Uniform Resource Identifier*, o *Identificador Uniforme de Recursos*. Es un sistema de identificación única para algún tipo de documento, no necesariamente computacional; por ejemplo el "ISBN" para libros o el "DOI" para artículos son URI's. Cuando se trata de recursos que pueden ser accedidos vía internet como un sitio web se le denomina URL o *Universal Resource Locator*, *Localizador Universal de Recursos*.

XML + Espacio de Nombres. La capa más técnica de la *web semántica*. En esta capa se agrupan las diferentes tecnologías que hacen posible que datos, sistemas y programas puedan entenderse entre ellos, por ejemplo los descriptors XML de los documentos. "XML es *eXtended Markup Language*" o *Lenguaje Extendido de Marcado*"; un estándar para describir todo tipo de documentos de acuerdo con una forma normalizada y entendible por máquinas. Espacio de nombres: Un conjunto de nombres que es usado para denominar tipos de elementos y nombres de atributos y en el cual cuando se reúnen objetos de diferentes orígenes que tienen nombres iguales se procede a eliminar ambigüedades ya que éstas no deben existir; esto se logra añadiendo el URL u otro origen del objeto a su nombre. Por ejemplo, una copia del mismo documento que existe en dos sitios Web simultáneamente.

RDF. Marco de Descripción de Recursos" o *Resource Description Framework*. Esta capa define un lenguaje universal con el cual podemos representar metadatos de nuestro documento de manera tal que puedan ser leídos por diferentes programas de variadas plataformas.

Ontología o vocabulario ontológico - Especificaciones explícitas de conceptualizaciones del conocimiento, al agregar nuevas clases y propiedades para describir los recursos.

Lógica Lógica asociada a las búsquedas o recuperaciones del documento y programable en un sistema.

Pruebas Pruebas aplicables al documento en su caso, para establecer autenticidad, propiedad, legalidad, etcétera.

Software confiable - *software* de confianza asociado para explotar el documento.

Firma digital Firma electrónica que puede agregarse al documento para autenticarlo si se desea.

Cifrado Tipo y llaves de cifrado o encriptado del documento si se desea que exista en esta forma.

Anexo 2. Ontología “C” Original en inglés. Tomada de: InterPARES. The International Research on Permanent Authentic Records in Electronic Systems. Inter pares 2 Terminology Database and Ontologies. Disponible en: <http://www.interpares.org> http://www.interpares.org/ip2/display_file.cfm?doc=ip2_ontology.pdf 

ONTOLOGY “C”:

Trustworthiness of a Record

