

Pensar la datificación en el marco de la formación: desafíos actuales

José Texier

Universidad Nacional de Chilecito

jtexier@undec.edu.ar

Resumen: La datificación es el proceso de transformar el dato, por medio del análisis y la reorganización, en información susceptible de ser utilizada en cualquier área del conocimiento o disciplina, pero los diferentes cambios disruptivos tecnológicos junto con “Era de la información” que se vive en la actualidad, han centrado a los profesionales de las Ciencias de la Información y de las Ciencias de la Documentación (conocidos en inglés como *Library and Information Sciences* - LIS) en el recurso más importante para hacer disponible la información a la sociedad. Por tanto, este artículo reflexiona sobre la educación de los profesionales como el reto más importante para la adaptación al sinfín de oportunidades que aparecen hoy en día: servicios informáticos en la nube (IBM Cloud, Google Cloud Platform, Azure), aprendizaje automático o machine learning, la web semántica, los MOOC, la ciencia abierta y acceso abierto, bases de datos bibliográficas de acceso libre, entre otras. En conclusión, el objetivo para los profesionales del área es analizar antes de recolectar datos y esto se logra con educación y no por un golpe de suerte o por el simple manejo de una herramienta.

Palabras clave: datificación, LIS, retos, oportunidades, informática, educación.

1. Introducción

Pensar en los desafíos para la formación de los profesionales de las Ciencias de la Información y de las Ciencias de la Documentación (conocidas en inglés como *Library and Information Sciences* - LIS), significa problematizar y visibilizar el lugar de los datos en esta “Era de la información” (Cf. Castells, 2009; De Souza-Silva et al., 2001; Echeverría, 2015; Shannon & Weaver, 1949). Para ilustrar esto, basta con revisar algunas estadísticas, por ejemplo, el 5 de enero de 2020 hubo 4.437.215.927 de usuarios en internet, lo que representa el 54.6% de la población mundial (InternetLiveStats.com, 2020). Se generaron por día un aproximado de tres mil millones de búsquedas, 500 millones de *tweets*, cinco mil millones de GB de datos. Se abrieron 600 millones de blogs y se contabilizaron unos 230 millones de registros en bases de datos bibliográficas elaborados por unos 32 millones de autores (principalmente, de artículos científicos y/o académicos) en 65 millones de revistas (Cambia.org, 2020; InternetLiveStats.com, 2020; WebsiteHostingRating.com, 2020). Esta información que sorprende, obliga a preguntarse y repreguntarse ¿Qué pasa con esta multitud de datos?, ¿qué se debería hacer?, ¿a quién(es) beneficia? y ¿cómo deberían gestionarse?

Este trabajo reflexiona y esboza algunos argumentos sobre la necesidad de que la formación de los profesionales del dominio LIS, interpele y analice el uso de los datos para que sean útiles en cualquier disciplina y/o oficio de la sociedad, sobre todo, por la imperiosa exigencia de afrontar con las herramientas conceptuales, técnicas y éticas la vorágine de información que nos rodea.

La evolución de internet desde 1990 hasta el momento, el surgimiento de las redes sociales, la migración de gran parte de las actividades académicas-científicas al mundo digital y el descenso de los costos para adquirir los recursos y servicios informáticos (Griffiths et al., 2016; Suber, 2012) han marcado una agenda que no puede ser rechazada por los profesionales que directa o indirectamente están ligados al mundo de los datos y la información. El dilema actual no está solo en la cantidad de datos, sino en qué se debe hacer con todos ellos; cómo se gestionarán grandes volúmenes de información digital estructurada, semiestructurada y no estructurada, generada por dispositivos móviles, sensores, aplicaciones de informática, entre otros y, creados por personas, organismos y/o empresas públicas y privadas en todos los ámbitos de la sociedad.

Así, por la necesidad de recolectar, analizar y usar los datos, aparece la *Big Data* (BD), que se caracteriza como la tecnología que permite procesar grandes cantidades de datos en poco tiempo (Cf. Rajaraman & Ullman, 2011), ya que el procesamiento y posterior análisis de los datos y/o información, en muchos casos, debe ser realizado en tiempo real para incidir en oportunas tomas de decisiones. La caracterización típica de un problema de *Big Data* es si cumple con: un gran volumen (grandes colecciones de datos), con variedad de fuentes tanto estructuradas o no (ausencia de estructura) y con velocidad en relación con la generación, acceso y análisis de los datos para ser explotados, además de poder dar veracidad a la información y valor a los datos (Agarwal et al., 2011; Tascón, 2013; Yu & Wang, 2015).

Ahora bien, cómo se debería encarar la recolección, análisis y uso de los datos también es inherente a la datificación. Se entiende que *Big Data* y datificación están vinculados. No obstante, esta última tiene que ver con la acción de convertir y medir en datos una información, un proceso o una actividad gracias a la infraestructura tecnológica existente (Markus, 2017; Mayer-Schönberger & Cukier, 2013). A su vez, se relaciona datificación con digitalización, sin embargo, Mayer y Cukier (2013) plantean que la datificación ocurrió antes de las Tecnologías de Información, hecho que permitió el surgimiento de la digitalización. Estos autores explican que la datificación hace alusión al dato transformado, por medio del análisis y la reorganización, en información susceptible de ser utilizada en cualquier área del conocimiento o disciplina. En cambio, la digitalización (Mayer-Schönberger & Cukier, 2013) es el proceso mediante el cual la información analógica se convierte en digital (códigos binarios de ceros y unos) para que sea legible y procesada por un computador. En síntesis, con la datificación los datos se pueden cuantificar de manera que pueden ser tabulados y analizados, luego, se almacenan en diferentes formatos para poder aplicar *Big Data* (Marín & Lisbona, 2018).

La recuperación, transformación y visualización se pueden pensar como los procesos de la datificación. Así, el proceso de recuperación consistirá en obtener datos confiables, ya sea de medios y/o infraestructuras específicas o a través de la aplicación de patrones usando algoritmos y/o softwares existentes. Las fuentes de datos para realizar la recuperación pueden ser muy variadas, pero casi siempre van a estar relacionadas con soportes informáticos. Según Soares (2012), se pueden extraer datos para hacer datificación desde: *Web and social media* (Facebook, Twitter, LinkedIn, blogs, etc.), *Machine-to-machine-data* (sensores, internet de las cosas, teléfonos inteligentes), *Big transaction data* (reclamos de atención médica, registros de telecomunicaciones, registros de facturación de servicios públicos), *Biometrics* (reconocimiento biométrico o biometría, se refiere a la identificación automática de una persona basada en sus características o rasgos anatómicos o de comportamiento) y *Human-generated data* (seres humanos generan grandes cantidades de datos: notas de los agentes del centro de llamadas, grabaciones de voz, correo electrónico, fotos, videos, documentos en papel, encuestas, registros médicos electrónicos).

Por su parte, el proceso de transformación tiene como objetivo realizar la limpieza y consolidar los datos entrantes de diferentes fuentes, de tal manera que cuente con datos precisos que sean correctos, completos, consistentes y sin ambigüedad. Y, por último, el proceso de visualización se entiende como la monitorización de los datos producidos por el objeto de la datificación (Lope Salvador et al., 2020; Markus, 2017; Mayer-Schönberger & Cukier, 2013).

Los procesos de datificación brindan un espacio de oportunidades a los profesionales de la información y documentación mediante la intervención de técnicas y teorías computacionales que deberían conocerse o promover en trabajos junto con los informáticos. Por ejemplo, esos procesos tienen una gran similitud con el patrón de arquitectura de software conocido como ETL (extracción, transformación y carga/load) usado para la integración de datos en las Ciencias de la Computación (ampliar en Ashima et al., 2015). Este patrón consta de tres etapas funcionales consecutivas: la extracción de datos de diferentes fuentes, las

transformaciones necesarias por la heterogeneidad de la información a través de reglas, y la carga final en un sistema de base de datos (Awad et al., 2011; El-Sappagh et al., 2011; Texier et al., 2017). Esta arquitectura se utiliza para unificar la información en los procesos de inteligencia de negocios que conducen a la toma de decisiones.

Por ende, se puede establecer una relación entre parte de los procesos ETL y la datificación, pero sacando la fase de carga, ya que la carga sería un proceso informático relacionado con el manejo de diferentes paradigmas de bases de datos que respetan criterios como valores únicos, integridad referencial, campos obligatorios y chequeo de rangos de valores (Texier et al., 2017).

Sin duda, el área disciplinar que más oportunidades tiene ante estos procesos de datificación es la de las Ciencias de la Información y Ciencias Documentales, porque son ellos los llamados a hacer la implementación de los procesos de acuerdo con las habilidades desarrolladas en la formación académica y la experiencia laboral.

2. Reflexión

Luego de entender el concepto de datificación, se puede afirmar que los informáticos han desarrollado un conjunto de competencias necesarias para comprender un proceso ETL parecido a la datificación, pero con una debilidad en el tema de la catalogación de los datos. Varios autores confirman (por ejemplo, Chowdhury & Chowdhury, 1999; Liew, 2009; Nguyen & Chowdhury, 2013; Pomerantz et al., 2006; Texier et al., 2012) la relación entre profesionales del dominio LIS y profesionales de las Ciencias de la Computación, tomando en cuenta los cambios paradigmáticos de adquisición de datos/información y acceso al conocimiento. También lo muestran las diversas temáticas y trabajos presentados en conferencias y revistas del área: *Joint Conference on Digital Libraries, JCDL*; *Theory and Practice of Digital Libraries, TPDL*; *la International Conference on Asian Digital Libraries, ICADL*; y *la IEEE/TCDL, Technical Committee on Digital Libraries* (algunas temáticas presentadas: digitalización, almacenamiento e intercambio; objetos digitales, composición y paquetes; metadatos, catalogación, gestión de autores; repositorios, archivos; interoperabilidad, mediadores; servicios de búsqueda, linking, navegación; procesamiento de lenguaje natural; integridad de datos, preservación; inteligencia artificial).

Por otra parte, Morato et al. (2016) afirman la baja demanda laboral de profesionales de las Ciencias de la Información y Ciencias Documentales a pesar del aumento de especialización tecnológica. Los autores predicen un incremento en la demanda y necesidad de competencias tecnológicas/informáticas en el dominio LIS. Esto se traduce en una oportunidad porque visibiliza la importancia de desarrollar las competencias necesarias para la datificación que se viene describiendo en los perfiles profesionales LIS. Incluso, otros autores, principalmente Fox (2019) han propuesto el desarrollo de un Currículum DL que fortalezca el desarrollo de competencias LIS, y, por ende, los procesos de datificación (Franks, 2014; ILS, 2013; Oh et al., 2016; Pomerantz et al., 2006; Shuva & Audunson, 2014; Tammara, 2013; Texier et al., 2016; Texier & Zambrano, 2019). Este Currículum DL propone la capacitación en competencias computacionales para encarar diferentes tópicos desarrollados en programas del dominio LIS. Se destaca que el Currículum DL para profesionales del dominio LIS puede estar dirigido a través de programas académicos de posgrados, en cambio, para profesionales de las Ciencias de la Computación puede incorporarse en asignaturas electivas al final de sus carreras (Texier et al., 2016; Texier & Zambrano, 2019).

Los profesionales LIS, al tener participación activa en la recuperación, transformación y visualización de los datos existentes, permitirán mejorar los procesos de datificación y obtendrán como resultados datos de calidad que conduzcan a búsquedas más precisas y disponibles para una audiencia más variada. Un ideal de calidad de datos sería la guía de principios de la Unión Europea, conocida como FAIR Data (European

Commission, 2018) que si bien está orientada para datos científicos puede ser usada para considerar lo que es un dato de calidad en cualquier disciplina. El desarrollo de esta guía FAIR se logró por el apoyo de diversos sectores interesados (academia, industria, entes financiadores, editores académicos, gobiernos) en proporcionar datos que los computadores (mejor dicho los algoritmos) puedan encontrar (*findable*), acceder (*accessible*), interoperar (*interoperable*) y reutilizar (*reusable*) sin ninguna o mínima intervención humana.

En otras palabras, los principios remiten a: *Findable* la gran cantidad de datos se pueden encontrar a través de los metadatos que los identifican, describen y localizan. *Accessible* porque los grandes volúmenes de datos pueden recuperarse mediante protocolos estandarizados de comunicación, además, los metadatos son persistentes aún cuando los datos dejen de estar disponibles. *Interoperable* debido a que la multitud de datos de esta época tiene que poder intercambiarse con al menos dos sistemas o componentes de software. *Reusable* porque todos los datos recuperados y transformados pueden reutilizarse porque se conoce su procedencia y las condiciones de su reutilización (European Commission, 2018; Tanhua et al., 2019).

El hecho de tener gran cantidad de datos confiables o datos de calidad se convierte en una coyuntura que los profesionales de la información podrían aprovechar en lo que actualmente se conoce como la alfabetización informacional y/o digital. Con el apalancamiento de los avances tecnológicos actuales se han promovido nuevas formas de relación entre las personas y han cambiado muchos los procesos cotidianos, los cuales están cada vez más controlados por dispositivos tecnológicos (Pino, 2010). Por ello, a través de la alfabetización se democratizan los procesos de formación y se consigue la inclusión social, laboral y una mejora en la calidad de vida (Ortega, 2009). En esta búsqueda de nuevas formas de acercarse y apropiarse de las tecnologías de la información, las universidades juegan un rol primordial porque sirven como espacio para la creación de soportes (software, hardware), la educación, la investigación (sobre el funcionamiento de esos soportes y la manera de usarlos) y la vinculación (extensión) con la comunidad y otros sectores de la sociedad además de fomentar la interdisciplinariedad a través de la colaboración en equipos de trabajo dirigidos por bibliotecarios, archivistas y/o documentalistas, con tareas propias de sus conocimientos: recuperación, transformación, preservación de datos y construcción de comunidades virtuales.

Otra gran oportunidad para tomar en cuenta en los procesos de datificación en la “Era de la Información” tiene que ver con el uso de desarrollos e implementaciones tecnológicas, que Vidal et al. (2019) plantean como cambios disruptivos que no pueden ser ignorados, entre ellos, los servicios en la nube donde los datos se pueden recuperar, procesar y visualizar con tecnologías informáticas disponibles en la internet (no están instalados en una computadora local); los blockchains (mejoran la gestión de los datos, gracias al tracking de los procesos de los datos recuperados y/o creados); la lingüística computacional que habilita a las computadoras la tarea de procesar y entender el lenguaje natural (ampliar en Vallez & Pedraza, 2007); el aprendizaje automático (*Machine learning*) por la búsqueda de patrones dentro de una gran cantidad de datos a través de algoritmos y procesamiento computarizado de altas prestaciones; la web semántica para que los buscadores puedan comprender los datos y la información, es decir, que los datos publicados (datificados) puedan ser legibles (interoperables) por computadoras a través de sus metadatos semánticos y ontológicos; los MOOCs: existe una oportunidad (poca oferta) para la disciplina LIS por ser cursos abiertos y masivos (ampliar en Zade & Babare, 2020); la ciencia abierta debido a la apertura de todo el proceso de investigación y la generación de productos (Alonso-Arévalo & Lopes, 2019); y, las bases de datos bibliográficas de acceso libre, por ejemplo: Lens.org, el nuevo Microsoft Academic, Google Scholar, Dimensions, 1findr.

Lope Salvador et al. (2020) señalan que las oportunidades de un nuevo paradigma digital para los profesionales de la información están en la alianza entre la Ciencia de Datos y la Neurociencia proporcionando formas de integrar el sentido y la subjetividad en la Ciencia, por ello, mencionan la

conjunción de la lingüística computacional y el aprendizaje automático. Esta idea junto con las otras que se han venido esbozando, identifican, de alguna manera, muchas de las necesidades de conocimientos y competencias tecnológicas (muy relacionadas con los procesos de datificación) que están favoreciendo la inserción laboral de los profesionales del dominio LIS y que están en línea con los portales de ofertas laborales como Laboris.net y LinkedIn, donde dichas necesidades indican que los profesionales deben adaptarse a los nuevos contextos con el fin de facilitar al usuario datos e información accesible, relevante y de calidad, lo que modifica, poco a poco los roles de los empleos (Méndez & van Hooland, 2009; Morato et al., 2016).

3. Conclusiones

El gran reto para el dominio de las Ciencias de la Información y Documentación ante los procesos de datificación es poder repensar la necesidad de mejorar día tras día y aportar desde otro lugar para todas las disciplinas, de esta manera, los profesionales actuales y del futuro, podrán desarrollar las competencias necesarias para estar presentes en diferentes fases de la datificación (búsquedas, datos primarios de cualquier investigación, transformación, visualización) y aportar el conocimiento, las habilidades y las actitudes adquiridas en otros ámbitos disciplinares.

Este desafío evidentemente debe ir acompañado y pensado desde la interdisciplinariedad, puesto que estamos ante una sociedad cada vez más compleja y la aquejan hechos que no pueden encararse desde una única mirada disciplinar. Se destaca que la capacitación (desarrollo de competencias) tiene que estar orientada a conocimientos de base y no a la actualización de herramientas de software. De tal manera que los profesionales LIS podrán mantenerse al día con cursos o talleres sobre herramientas y técnicas computacionales pensando en los cambios disruptivos. Por ello, el dominio LIS tendrá que adaptarse a los nuevos contextos con el fin de facilitar al usuario una información accesible, relevante y de calidad.

Para finalizar, es una prioridad pensar propuestas de enseñanza y formación innovadoras, por ejemplo, la del Currículo DL que permitirá modificar la percepción del público sobre el perfil de los profesionales LIS, para entender que pueden ayudar a las otras disciplinas en procesos de datificación, pero tomando en cuenta el componente de la globalización, es decir, colaboración entre diferentes facultades y departamentos LIS, entre académicos, entre bibliotecas, entre empleados, todos estos en diferentes partes del mundo, maximizando recursos, experiencia e instalaciones (cf. con Rehman, 2008). En el mundo actual, no hay más que ser abiertos, creativos, innovadores y emprendedores. Se observa en un estudio reciente (Xue et al., 2019) el compromiso de transformar la educación LIS en los dos mayores países del mundo, China y Estados Unidos, que además tienen una extensión territorial enorme que hace el compromiso aún más difícil. Los autores identifican y discuten cuatro tipos de desafíos en la educación de LIS analizando cuantitativa y cualitativamente los datos recopilados de dichos países China y los Estados Unidos: 1) identidad y acreditación, 2) supervivencia y prosperidad, 3) actualización y mejora del currículo, y 4) formato y contenido de entrega del curso.

En fin, la generación de los datos por cada usuario hoy en día es enorme y lo importante es no fatigarse en datificar por datificar, sino plantear la *pregunta correcta*. El objetivo es analizar antes de recolectar datos y esto se logra con educación y no por un golpe de suerte o por el simple manejo de una herramienta. Sin duda, la formación de los profesionales de la información y documentación requiere cambios que trasciendan los datos e impacten en toma de decisiones que nos permitan comprender que un dato se resignifica con el uso y el impacto que produce en los otros. El reto, sin duda, es brindar una formación innovadora y de calidad que aproveche las oportunidades de este momento histórico hiperinformado nos ofrece.