**Periodismo de datos y las nuevas necesidades de sistematización de información: una mirada desde la Bibliotecología**

**Lourdes Feria Basurto**

**Resumen**

La evolución de los medios de comunicación en el entorno digital ha derivado en nuevas especialidades, siendo una de ellas el Periodismo de Datos; donde, al igual que en todo proyecto basado en el uso intensivo de datos, el riesgo de vulnerabilidad de éstos está presente de no haber una estrategia de sistematización desde el inicio. Estas iniciativas se beneficiarían positivamente al manejarse en congruencia con los principios FAIR que promueven la localización, accesibilidad, interoperabilidad y reutilización de los datos mediante buenas prácticas como la construcción de bases de conocimiento, repositorios normalizados, asignación pertinente de metadatos e identificadores globales, desarrollo de vocabularios, ontologías y elección de plataformas estandarizadas, que permitan alimentar iniciativas de acceso abierto y favorezcan la reproducibilidad. El periodismo de datos suele estar conformado por equipos multidisciplinarios en los cuales participan reporteros, científicos de datos, e informáticos, cuya fortaleza se vería incrementada al incorporar la participación de un profesional de la información, cuyo perfil contemple las características necesarias para la sistematización de información que requiere dicha vertiente informativa. Con base en todo ello, este trabajo busca responder las siguientes preguntas: ¿Por qué tendrían que integrarse los profesionales de la información a las actividades relacionadas con el periodismo de datos? ¿Cómo serían y cuáles, las buenas prácticas del bibliotecario (documentalista) de datos en las salas de prensa digitales? ¿Cuáles tendrían que ser los contenidos necesarios para su formación?

**Descriptores**

Periodismo de datos, Bibliotecario de datos, Manejo de datos, Herramientas de gestión de datos, Habilidades relacionadas con datos

**Introducción**

*It was the best of times, it was the worst of times,*

*it was the age of wisdom,*

*it was the age of foolishness,*

*it was the epoch of belief…*

C. Dickens

Los escenarios del mundo digital y sus capacidades para el manejo de volúmenes de datos cada vez mayores impulsan a repensar metodologías, estructuras, formas de acreditar el saber y a la creación de redes multidisciplinarias. La Bibliotecología y las profesiones de la información están analizando su papel para adaptarse al futuro abriéndose a una mayor vinculación con las tecnologías, a la integración en equipos multiprofesionales y multihabilitados y a la comprensión permanente de las necesidades de los usuarios y sus expectativas.

La web es muy compleja, abundante, ágil, pareciera infinita en datos (palabras, números, metadatos, preguntas, respuestas—que oscilan entre la falsedad y la veracidad), y formatos (videos, podcasts, blogs y tantos otros). En ese contexto, abordar el término *datos* no es sencillo ya que se observa una tendencia a abusar del mismo en el ámbito mercadológico y en redes sociales; de ahí la necesidad de revisar previamente el concepto. Para los fines de este documento se ha partido de la definición que, desde las Ciencias de la Documentación, propone José López Yépez (2004, p. 411):

“[...] testimonio o expresión mínima de un hecho mensurable. Es la representación de algo que aparece, que está presente en el sujeto cognoscente y que es el punto de partida del conocimiento. En general los datos son el punto de partida, los hechos y principios indiscutidos para una investigación científica. Es la experiencia sensible en el empirismo [...] Un conjunto de datos forman una información.*”*

En el escenario de la Sociedad Red los datos cada vez están más presentes, tan es así que a ese crecimiento desbordante se le ha conoce como *Big Data*. Este ecosistema digital ha transformado todo; se trata de un fenómeno cultural que está permeado por tres factores:

“1) Tecnológico (algoritmos que recaban, analizan, enlazan y comparan enormes cantidades de datos; 2) analítico que permite identificar patrones; 3) la mitología de que los grandes datasets permitirán percepciones precisas dando un aura de verdad, objetividad y exactitud.” Lewis & Westlund (2015) citado por Sandoval-Marín & La Rosa (2018), p.195

Ante este panorama y motivada por el interés de cómo acceder adecuadamente a los datos, cómo visualizarlos, y especialmente cómo gestionarlos de manera más efectiva desde el universo de la Documentación y la Bibliotecología para ofrecer mejores servicios de información, la Federación Internacional de Asociaciones de Bibliotecarios (IFLA) tomó la iniciativa a publicar dos números especiales de su *IFLA Journal* a fin reunir las teorías más recientes, investigaciones y prácticas que están informando e innovando en servicios de datos efectivos desde una perspectiva internacional; el primero de ellos apareció en 2016 (conteniendo siete artículos) y otro en 2017 (con diez contribuciones) abarcando cuatro grandes temas: las necesidades de los usuarios, las habilidades requeridas, los servicios a ofrecer y la alfabetización en datos.

Adicionalmente, quien esto escribe, en una búsqueda bibliográfica realizada en bases de datos científicas, localizó trescientos treinta y nueve artículos en otras fuentes relacionadas con buenas prácticas de gestión de datos en bibliotecas. Esta exploración, además de propiciar la reflexión en torno al trabajo que se hace en el mundo en torno al tema, ha permitido preparar una relación con vínculos donde se muestran más de sesenta recomendaciones y experiencias en el ámbito bibliotecológico. Entre tanto ¿qué está ocurriendo con la gestión de datos en el mundo del periodismo? ¿cómo se vive en las salas de redacción su manejo desde que se generan hasta que se publican? ¿qué medidas se están tomando para su preservación y para posteriormente seguirse consultando?

**Periodismo de datos**

La digitalización y los nuevos medios “han modificado las tareas de producción, edición y difusión en las industrias de noticias y, por tanto, también las labores del periodista” (Ferreras, 2013, p.117). Su aspiración es ser un periodismo de primicias, ir más allá del boletín, considerar no sólo cifras sino otras variables en torno a sus hipótesis, fortaleciendo los valores clásicos del periodismo: rigor, veracidad, transparencia (Sandoval-Martin, 2018, p.200).

Los especialistas que realizan este tipo de prácticas como Crucianelli (2013), Saúl (2016) y Cabra (2013), coinciden en afirmar que el periodismo de datos, consiste en usar datos masivos, realizar más investigación, profundizar en el seguimiento (entrevistas, búsqueda de evidencias, etc.), y en la verficación antes de proceder a contar una historia. Para ello emplean “hojas de datos, emails, tuits, fuentes que permiten hacer historias con mucho más detalle” (Cabra, 2013), agregando que desde luego las estadísticas son fuente primaria fundamental, pero más allá todo lo procesable por algún medio informático entraría en la categoría de “datos” (Saúl, 2016); con la peculiaridad de que los productos que se entregan son en sí mismos reportes informativos “[…] aunque no se transmitan con palabras, imágenes, audio o video, sino en código. Cuando hablamos de periodismo de datos, tenemos que hablar de tecnología, software, dispositivos y cómo contar una historia con ellos” (Matzat, 2012); e incluso entrarían en esta categoría proyectos como *apps* de noticias [y] aplicaciones interactivas de noticias” (Boss, 2017, p.151).

El insumo para esta variante periodística proviene, en buena parte, del espectro de datos masivos en el ciberespacio. Todo ello implica capturar información muchas veces de manera directa, generar tablas, descargar datos, convertir *pdfs* a *Excel*, *XML* u otros formatos, integrar los datos cuantitativos con los cualitativos, compartir archivos, colocarlos a disposición en repositorios abiertos, graficarlos y generar líneas de tiempo y visualizaciones profesionales. En gestión de datos periodísticos, al igual que en otros ámbitos llegó el momento de aprender a usar software para información cuanti y cualitativa, hacer infografías, diseño y buen manejo de redes sociales; tomando en cuenta, desde luego que por precisas que sean las herramientas éstas no resuelven todo, ya que se sigue requiriendo la revisión puntual por parte del reportero que investiga. (Saúl, 2016)

El ciclo de los datos en estas metodologías va “de los datos muertos al análisis, a la limpieza y al conocimiento [a partir de la reflexión sobre ellos] y de ahí hacer mapas interactivos, gráficas claras, historias interesantes y verídicas” (Saúl, 2016). La secuencia constaría de cuatro etapas (fases) de captación de datos de acuerdo con Heravi & McGinnis citado por García & García (2018, p..86): descubrimiento de contenido, filtrado y contextualización, confianza y verificación y, finalmente, publicación.

El equipo básico del periodista de datos lo constituye, desde luego, su computadora personal (cabe considerar que para el manejo de datos podrían ser necesarias tanto una PC como una Mac, ya que el sistema operativo compatible con la procedencia de los datos suele ser un requerimiento obligado); herramientas (hojas de cálculo, aplicaciones libres para limpieza de datos como *OpenRefine,* o para graficación y visualización); y una matriz de lo que se aspira a conseguir y obtener al ir respondiendo las preguntas de investigación. (Saúl, 2016)

Diarios de impacto mundial ya hacen periodismo de datos formalmente, algunos de ellos son *The Guardian*, *The New York Times*, *El País* (España), *La Nación* (Argentina). Es aquí donde surge otra pregunta ¿esta opción es para todos los medios?. Podría pensarse que no, por los costos y por la imperiosa necesidad de contar con un *know-how* informático importante, sin embargo empiezan a surgir iniciativas en medios locales jóvenes que han encontrado en este nicho una vocación a ser atendida.

**Principios FAIR**

Independientemente del ámbito en el que se apliquen los datos, sea en el periodismo, en la investigación científica, o en la administración y mercadotecnia, el profesional de la información debe estar consciente de la responsabilidad que implica una buena gestión de los mismos; por esa razón los datos deberán, ante todo, tener cuatro características conocidas como *FAIR*, mismas que tienen su origen en la reunión de expertos convocada en el año 2014 por el Netherlands eScience Center y el Dutch Techcentre for the Life Sciences (DTL) donde surgieron los lineamientos que establecen que los datos de investigación deben ser: *Findable*, *Accessible*, *Interoperable* and *Reusable* (FAIR), tanto para los equipos informáticos como para las personas, lo cual tiene que ver con los datos pero también con las herramientas, vocabularios e infraestructuras que han de tomarse en cuenta para lograr adecuadamente los objetivos de descubrimiento y reutilización de información. (FORCE11).

Los profesionales de la información cuentan con la preparación y la experiencia para integrarse en esta cultura y aportar al proceso las herramientas adecuadas para convertirse en gestores, facilitadores, custodios, e intermediarios de datos, de tal suerte que se hagan cargo de coordinar procesos para conservarlos, curarlos digitalmente, descubrirlos, organizarlos y correlacionarlos; con lo cual algunos de los productos y servicios que se estarían generando irían desde portales de conocimiento, aplicaciones de minería de datos, analítica de usuarios, clasificación de contenidos y orientación a favor del cuidado y acceso de y hacia los datos *FAIR* en las salas de redacción.

**El documentalista de datos en los medios de comunicación**

En el caso particular del periodismo ¿cuáles son las prácticas documentales específicas? Boss & Broussard (2017) se plantean el manejo de datos para la preservación, la recolección de información para atender necesidades de curadores y archivistas y consideran importante incluir la información descriptiva de las *apps* de noticias, la arquitectura de software, sus autores, los dueños de los derechos y la información sobre el licenciamiento, entre otros elementos; así como las preguntas que surgen acerca de cómo capturar, archivar, preservar y permitir el descubrimiento de los trabajos y las fuentes.

En gestión de datos periodísticos, al igual que en otros ámbitos es cada vez más necesario aprender a usar software para información cuantitiva y cualitativa, hacer infografías y tener buen manejo de redes sociales, no necesariamente para aplicarlos sino para entenderlos, ya que “la posibilidad de compartir datos supone un nuevo reto en las tareas básicas de selección, análisis y difusión que realiza el documentalista” García & García (2018). Por otra parte, las distintas legislaciones relacionadas con la apertura de datos y las leyes de transparencia en los países han favorecido la creación de fuentes que nutren al periodismo de datos. El profesional de la información por su formación sabe de esas fuentes, de datos abiertos de gobierno y de búsqueda y recuperación de información en ellas. Las búsquedas que hacen los periodistas no siguen una metodología sistematizada. Un estudio reciente citado por García & García (*Op. cit*) refuerza este modo de recuperación de información por parte de los periodistas, quienes reconocen basarse en la aleatoriedad aportada por el entorno digital. “Este hecho induce a pensar en una precaria especialización en la búsqueda de datos online [que] se solventa en distintos medios de comunicación por los documentalistas en su calidad de formadores.” (Ibidem).

El equipo ideal estaría conformado por el periodista, el programador y el documentalista, ya que éste sería un valioso asistente de datos en la medida en que crecen y surgen nuevas aplicaciones y proyectos que no se limitan a la sistematización de texto digital, sino que incorporan lo que Boss & Broussard (2017) denominan “piezas de software”, es decir, conjuntos de datos que no se pueden archivar del mismo modo que los documentos digitales; de ahí la necesidad de herramientas dinámicas y arquitecturas para las nuevas colecciones y necesidades de información (ver figura 1), que integren fuentes de datos heterogéneas (estructuradas, no estructuradas o semiestructuradas), con capacidades de consulta robusta. Los ejemplos que se observan en la figura son: un conjunto de tweets, una plataforma de búsqueda sobre texto completo; un conjunto de bases de datos relacionales seleccionadas mendiante curaduría y un conjunto de fuentes RDF con información georeferencial y LOD (linked open data).

La propuesta mixta también comprende: un gráfico de datos RDF (de la propia aplicación) representado aquí como nodos (puntos negros), este incluye una ontología (clases y propiedades) y/o tripletas de datos específicos de determinadas áreas de periodismo de datos que pueden exportarse fácilmente a RDF o armarse manualmente (en ocasiones es necesario hacer limpieza y transferencia de datos, incluso de imágenes en pdf a hojas Excel).

Es muy interesante el caso del software porque no resulta sencillo identificar si lo que debe preservarse es el código fuente, la versión ejecutable, o ambos junto con todo el ecosistema de hardware, sistema operativo, lenguajes y compiladores, software, archivos, etc. (Boss & Broussard, 2017, p. 152) y agregan que “gran cantidad de terabytes de contenido nacido digital está condenada a perderse en el agujero negro de la obsolescencia tecnológica” (Boss & Broussard, 2017, p. 154).

**Arquitectura para consulta de bases de datos mixtas en periodismo de datos**



Fuente: Bonaque, et al. (2016)

Hasta ahora, la realidad es que los periodistas se están autoformando como documentalistas, “el documentalista, por su parte, se ha visto obligado a abrir su círculo relacional: ya no trabaja únicamente para y con el equipo de redacción de cada medio, sino “como señala Guallar (2012) debe mantener un contacto directo con los usuarios; una novedosa faceta que es posible gracias a la proliferación de las redes. A todo ello se añade una profunda alteración de las rutinas laborales que supone la constante actualización de las informaciones; una tarea derivada de los cambiantes contenidos en la web y en las redes sociales, cada vez más acelerados y globalizadores” (García & García. *Idem*). Los argumentos que Matzat (Op. cit) esgrime para convencer a los editores para que apoyen el periodismo de datos también sustentarían la participación de un bibliotecario de datos: son proyectos de largo aliento, se apoyan en trabajo previo (incluso esos datos pueden ser reutilizados), cuesta menos que desarrollar un motor de búsqueda y llega a más usuarios.

Como puede observarse, el tratamiento de información, entraña importantes derivaciones documentales; entre ellas cabe reseñar cambios en la recopilación de datos para la elaboración de informaciones que implica, en la labor documental, una mejor contextualización de los temas tratados como noticias con el empleo de recursos que facilitan la interpretación. Como caso particular destacan los llamados content-curators y los lenguajes documentales (García & García. Idem). En la tabla 1 se han integrado siete competencias adicionales, señaladas por las mismas autoras, que deberían formar parte de las habilidades del profesional de la información que aspire a participar en los proyectos de periodismo de datos. Si bien en esta relación se incluye el contar con capacidades de programación en JavaScript esto podría obviarse si en el equipo se cuenta con un profesional informático.

|  |
| --- |
| Herramientas para gestión de datos en periodismo |
| **Web scraping.** Técnica por la que se extrae información de la web y que se vincula con fórmulas de tratamiento y reconocimiento de información semántica. Ejemplo: ETL (extract, transform and load) para extracción, transformación y cargado de datos; se manejan en entornos de grandes volúmenes de datos a partir de diferentes fuentes y están destinadas al refinado y difusión de nuevos documentos |
| **Herramientas de refinamiento.** Ejemplo: Open Refine o Google Refine, que permiten entender la estructura y calidad de unos datos, corrigiendo los tipos de errores comunes en ellos |
| **Hojas de cálculo**. Como opción para la gestión y organización de la información estructurada |
| **Programación** Con lenguajes de programación como JavaScript y lenguajes de marcado como HTML o CSS (hoja de estilo en cascada) |
| **Prácticas y herramientas de infografía y visualización**. Mejoran el acceso a los usuarios a los datos organizados. Uno de los ejemplos de esta clase de utilidades es el de la generación de mapas interactivos |
| **Prácticas y herramientas para la comunicación visual.** Videos y animaciones |
| **Manejo de bases de datos georeferenciados**. Hacen posible la localización de personas, instituciones, objetos o eventos en mapas |

Fuente. Adaptado de García & García. (2018)

Finalmente, un documento que se ha convertido en referente en esta área es el *Data Journalism Handbook* que iniciativa de escritura colaborativa apareció primero en inglés, surgiendo posteriores versiones (que no traducciones) en español, de las cuales cabe destacar el titulado *Manual de Periodismo de Datos Iberoamericano* (2012-). Éste ofrece explicaciones detalladas sobre cómo buscar noticias geolocalizadas, usar buscadores comerciales y académicos; cómo ingresar a la Web profunda, la importancia de la sistematicidad, la extracción de datos, las herramientas del internet invisible, el rastreo digital sistematizado y herramientas digitales para la búsqueda de personas. También revisa los temas de la web semántica, redes sociales y sus búsquedas avanzadas, mashups y *Google Plus*. Otros capítulos se dedican al tema de los datos abiertos, bases de datos, visualización de datos y sistemas de información geográfica, herramientas para convertir audio y video, elementos gráficos gratuitos, software para extracción de datos, creación de libros de fotos y murales colaborativos, así como páginas con recursos y listas sobre herramientas digitales. Su consulta tendría que ser imprescindible para los bibliotecarios y documentalistas que planean incursionar en este ámbito.

**Conclusión**

La cultura de datos es un aprendizaje permanente que constituye un área de interés y de responsabilidad para las profesiones de la información. El conocer buenas prácticas de datos que han sido probadas en otros ámbitos, tales como el periodismo, fortalecerán la conciencia del rol del bibliotecario/documentalista como actor importante en el mundo académico y serán fuente de inspiración para nuevos servicios encaminados a apoyar al usuario especializado y a coadyuvar a la mayor calidad e integridad en los trabajos de investigación.

Las tecnologías han transformado los formatos, procesos, servicios y la relación misma entre el profesional de la información y las comunidades a quienes sirve—cada vez más preparadas digitalmente. Periodismo de datos es ir más allá de la atención pasajera, es periodismo sostenible y de calidad. Dadas las características de quienes lo desarrollan, la integración de equipos más sistémicos y colaborativos, donde el profesional de la información aporte sus destrezas y saberes, contribuirá a generar condiciones favorables al desarrollo de su entorno y al mismo tiempo una mayor capacidad de evolucionar en el acceso, uso, aprovechamiento y preservación de las fuentes.

**Referencias**

Ávila, Ana María. (2012). “Periodismo de datos: historia y momento actual”. En Manual de Periodismo de datos Iberoamericano. [Recurso Electrónico] <http://manual.periodismodedatos.org/ana-maria-avila.php>

Bonaque, Raphaël; Cao, Tien Duc; Cautis, Bogdan; Goasdoué, François; Letelier, Javier; et al. (2016). Mixed- instance querying: a lightweight integration architecture for data journalism. VLDB, New Delhi, India. hal-01321201v2. [Recurso Electrónico] <https://hal.inria.fr/hal-01321201/document>

Boss, Katherine & Meredith Broussard. (2017). Challenges of archiving and preserving born-digital news applications. IFLA Journal. Volume: 43 issue: 2, page(s): 150-157.

Cabra, Mar. (2013). Periodismo de datos ¿qué es el periodismo de datos? [Recurso Electrónico] (<https://www.cccb.org/es/multimedia/videos/periodismo-de-datos-que-es-el-periodismo-de-datos/211075>)

Crucianelli, Sandra. (2013). Herramientas digitales para periodistas. 2ª. Edición. Texas, Knight International Journalism Center. [Recurso Electrónico] https://knightcenter.utexas.edu/books/HDPP.pdf

European Journalism Center. Data journalism handbook (2012-). <https://datajournalism.com>

Ferreras Rodríguez, Eva María (2013). Aproximación teórica al perfil profesional del ‘Periodista de datos’. ICONO 14, Revista de comunicación y tecnologías emergentes, 11(2),115-140.[fecha de Consulta 3 de Febrero de 2020]. [Recurso Electrónico] <https://www.redalyc.org/articulo.oa?id=5525/552556576006>

FORCE11. (s. f.). Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version B1.0. [Recurso Electrónico] https://www.force11.org/fairprinciples

García Jiménez, Antonio, & Catalina García, Beatriz. (2018). Una perspectiva documental y bibliotecológica sobre el big data y el periodismo de datos. *Investigación bibliotecológica*, *32*(74), 77-99. [Recurso Electrónico] <https://dx.doi.org/10.22201/iibi.24488321xe.2018.74.57910>

López Yepez, José (ed.) (2004). *Diccionario Enciclopédico de Ciencias de la Documentación*. Madrid: Síntesis, 2 vol. pp. 59, 411-412.

Perry, Felipe & Miguel Paz. (2012-). Manual de periodismo de datos iberoamericano. Fundación PoderOMedia, Universidad Alberto Hurtado. [Recurso Electrónico] http://manual.periodismodedatos.org/sobre.php

Matzat, Lorenz. (2012). Nuestras historias aparecen en forma de código. En. Manual de periodismo de datos 1.0. Liliana Bounegru, Lucy Chambers, Jonathan Gray, March 2012. [Recurso Electrónico]

<http://interactivos.lanacion.com.ar/manual-data/en_la_redacción_8.html>

Sandoval-Martín, Teresa & Leonardo La-Rosa. (2018). Big data as a differentiating sociocultural element of data journalism: the perception of data journalists and experts. Communication & Society. 31 (4), 193-209.

Saúl, Lilia. (2016). Cómo sobrevivir al periodismo de datos sin un desarrollador [Entrevista en video realizada por Suchit Chávez de Connectas]. [Recurso Electrónico] https://www.youtube.com/watch?v=Eg0b4mlKabU