



## Del Big Data al library data

Nora Ledis Quiroz Gil  
Jefe. Biblioteca Fundadores Universidad CES  
[nquiroz@ces.edu.co](mailto:nquiroz@ces.edu.co)

### Resumen

La toma de decisiones acertadas basada en el adecuado análisis de datos es el proceso ideal de toda organización, las bibliotecas como las demás organizaciones producen un gran volumen de datos, de variado tipo y a velocidades exponenciales (Big Data), los cuales necesita gestionar y valorar.

La Biblioteca Fundadores, de la Universidad CES en Medellín - Colombia, presenta como una experiencia exitosa en unidades de información, la aplicación del modelo de descubrimiento del conocimiento en bases de datos – KDD – en el contexto de minería de datos o bibliomining, a través del diseño, desarrollo e implementación de ELOGIM, sistema integral que permite la producción y administración de datos estadísticos sobre el uso de los recursos electrónicos, elabora informes y provee información determinante para la toma de decisiones y la gestión en las unidades de información. Actualmente más de 30 bibliotecas y unidades de información de instituciones de educación superior de Colombia usan esta herramienta.

**Palabras clave:** Big data; indicadores de gestión; minería de datos; toma de decisiones; gestión del conocimiento; bibliotecas; unidades de información.

Big Data es un concepto lleno de significados, pues ahora no es exclusivo del mundo de la tecnología sino que se ha trasladado a todo tipo de organizaciones. Según el Institute for Business Value de IBM en su estudio Analytics: el uso del Big Data en el mundo real, (IBM Global Business Services, 2013) “Esta palabra se ha utilizado para trasladar al público todo tipo de conceptos entre los que se incluyen grandes cantidades de datos, analítica de redes sociales, herramientas de última generación para gestionar datos en tiempo real y mucho más.”

La compañía Gartner (Gartner, 2016) líder mundial en investigación tecnológica, le asigna al Big Data tres características claves iniciales “el Big Data es información de alto volumen, alta velocidad y gran variedad, que exigen formas rentables e innovadoras en su procesamiento y que permiten mejorar la comprensión de los datos, tomar decisiones y automatizar los procesos.” En este contexto el “volumen” corresponde a la cantidad de los datos, la “velocidad” al ritmo en que se crean, procesan y analizan los datos y la “variedad” a las diferentes fuentes y tipos de datos. Estas características se conocen como las “V”, del Big Data.

Por su parte, en el estudio antes mencionado IBM propone una “V” adicional correspondiente a la “veracidad de los datos” que hace referencia a la confiabilidad de los mismos. De igual forma, en su libro Big Data (Marr, 2015), Bernard Marr propone otra característica que sin duda es el elemento más importante: el “Valor”, entendido como el significado de los datos y como el costo – beneficio que estos representan para la organización, desde la generación de nuevos productos o servicios, la disminución de costos o la transparencia en la administración y en los procesos.

Otras teorías del Big Data proponen nuevas “V” como, Veracidad, Visualización, Volatilidad, Validez y Viabilidad. (Mitchell, Locke, Wilson, & Fuller, 2013)

## **Herramientas ¿Cómo se hace?**

Existen diferentes tipos de herramientas susceptibles a ser usadas pensando en dinamizar el concepto Big Data dentro de las bibliotecas, para ello se necesitan

obviamente los datos, el almacenamiento de los mismos, y una interface analítica que también permita su visualización. Todo esto hace necesario conocer la estructura de datos que propone el Big Data la cual se divide en:

1. **Datos estructurados:** que son producidos a través de máquinas como sensores, GPS, medidores clínicos, contadores eléctricos; web logs data que incluyen redes, aplicaciones, servidores depósitos en donde se generan grandes cantidades de datos estructurados; datos financieros y de ventas, etc.
2. **Datos no estructurados:** sin formato específicos como videos, audios, fotografías, imágenes satelitales, los generados en redes sociales, blogs, etc.
3. **Datos semiestructurados:** como los correos electrónicos, páginas web, etc.

Como una herramienta de Big Data basada en datos estructurados particularmente en los web log data, la Biblioteca Fundadores de la Universidad CES diseñó, desarrolló e implementó ELOGIM, un sistema web integral que permite la producción y administración de datos estadísticos de uso de los recursos electrónicos ofrecidos por las bibliotecas y unidades de información, el cual provee indicadores contundentes para la gestión académica, administrativa y la toma de decisiones.

Concebido a través de recursos de acceso abierto, permite obtener estadísticas en tiempo real por criterios y variables tales como: facultad, programa, tipo de usuario o un usuario en particular, y recurso electrónico. Es posible también definir el tipo de resultado deseado como: comparativos anuales, mensuales, diarios o por hora en el uso, y por consulta de los recursos electrónicos; ranking de usuarios, reportes generales, independientes o ranking de IP entre otros, lo que facilita la obtención de indicadores de costo/beneficio.

ELOGIM surgió como respuesta a la constante pregunta de los pares académicos del Ministerio de Educación Nacional sobre el comportamiento del uso de los recursos electrónicos en la universidad; de la necesidad de presentar

informes confiables a las directivas y del interés de gestionar adecuadamente los recursos económicos que invierte la biblioteca en este aspecto.

## ¿Cómo funciona ELOGIM?

Basado en la estructura del proceso de descubrimiento en bases de datos (KDD), en donde todos los datos transitan las etapas de recolección, depuración y análisis, apoyado en un modelo descriptivo para dar paso a un modelo deductivo, ELOGIM se desarrolló como una herramienta web que basa su arquitectura cliente - servidor en recursos de acceso abierto. El lenguaje de programación usado es PHP con HTML y Java Scrip, el servidor web es APACHE y el gestor de bases de datos MySql.

La captura de la información se realiza a través de los datos que se almacenan en el archivo (.LOG) generados por el tráfico de los usuarios al ingresar a cada recurso electrónico. La figura 1 presenta el modelo relacional del proceso de ELOGIM. (Quiroz, 2012)



Figura 1: Modelo Relacional del proceso de ELOGIM. (Construcción propia)

Este archivo (.LOG) puede ser configurado y parametrizado de acuerdo al requerimiento del sistema en la organización y a la necesidad de información y es a partir de él que se toman los datos, se generan los registros y posteriores

reportes estadísticos, convirtiéndolos a través del cruce de variables en información concreta como:

- Ranking de recursos electrónicos: Listado de los recursos electrónicos de mayor uso.
- Ranking de uso de los recursos electrónicos por usuarios, facultad y programa: Listado de personas que más usan los recursos electrónicos y que puede ser discriminado por tipo de usuario, facultad o programa.
- Efectividad en el recurso electrónico: Listado comparativo entre las sesiones iniciadas, búsquedas efectivas y los documentos descargados.
- Reconocimiento de direcciones IP de donde se accede a los recursos: El rastreo hace posible el seguimiento y control de las claves de acceso a los recursos electrónicos.
- Periodos de acceso: Definición de horas o periodos de tiempo con mayor tráfico de transacciones por parte de los usuarios de los recursos electrónicos. (Quiroz, 2012)

Teniendo en cuenta lo anterior, lo que hace ELOGIM entonces, es combinar variables pre establecidas sobre las cuales se quiere obtener la información de uso de los recursos electrónicos; estas variables están dadas por elementos como tipo de usuarios (estudiantes, docentes, investigadores); facultades o programas académicos (medicina, administración, derecho), sedes institucionales, recursos electrónicos (bases de datos, repositorios, libros electrónicos, etc.); y permite definir el rango de fechas entre las cuales se quiere hacer la consulta. La figura 2 presenta la interface gráfica de ELOGIM.

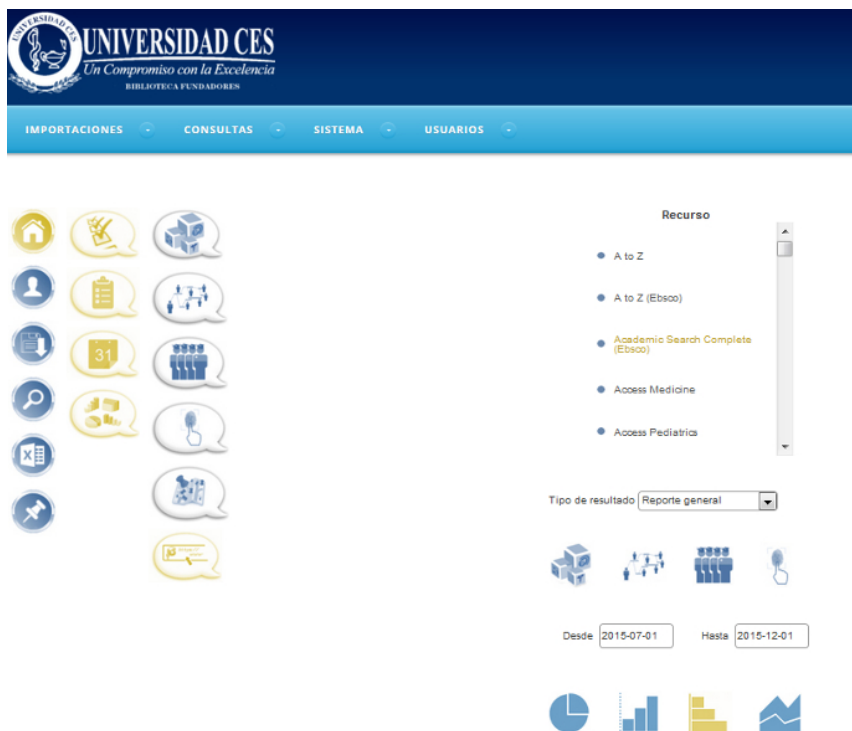


Figura 2. Interface de ELOGIM

También el tipo de resultado deseado (totales por recurso, independientes, comparativos de meses, días, años, horas, ranking de IP o ranking de usuarios), y descargas frente a sesiones de consulta. ELOGIM permite que estos resultados sean graficados en barras, sectores, curvas de crecimiento, así como su exportación a archivos en Excel para que sean manipulados y presentados de la forma más conveniente. La figura 3 presenta a selección de variables en ELOGIM



Figura 3. Selección de variables de ELOGIM

Resultado de las Sesiones, en el rango 2015-07-01 y 2015-12-01 (Descargas Vs Sesiones)

Facultad: Todas  
 Programa: Todos  
 Componente: Todos  
 Tipo de usuario: Todos  
 Sede: Todos  
 Recurso electrónico: Academic Search Complete (Ebsco) | BJM Clinical Evidence | ClinicaKey en Español | Dentistry & Oral Sciences Source (Ebsco) | Medline Complete (Ebsco) | Ovid | Pediatrics | Springer | Vlex |

Recurso electrónico	Sesiones	Descargas	Búsquedas	Promedio Descargas por sesión	Promedio Búsquedas por sesión	Cantidad de usuarios
Academic Search Complete (Ebsco)	3930	6445	19	2	0	1731
BJM Clinical Evidence	468	28	355	0	1	279
ClinicaKey en Español	5546	5823	12097	1	2	1222
Dentistry & Oral Sciences Source (Ebsco)	1938	5588	8	3	0	522
Medline Complete (Ebsco)	5760	6988	9398	1	2	1600
Ovid	6995	8756	5294	1	1	1349
Pediatrics	1062	652	1720	1	2	344
Springer	4146	4752	5902	1	1	1115
<b>Sesiones: 29845   Descargas: 39032   Búsquedas: 34793   Promedio descargas: 10   Promedio búsquedas: 9</b>						

Figura 4. Resultados de la consulta de ELOGIM

Tener claridad en cómo se aplica el Big Data en las unidades de información a través de herramientas como ELOGIM permite:

- **Actualizar los procesos:** Agiliza los procesos de gestión y da respuesta a solicitudes de información que sobre la biblioteca requieren los organismos nacionales o directivas universitarias respecto a los trámites de obtención de permisos para el funcionamiento de programas académicos o su acreditación.
- **Gestionar el conocimiento:** Permite conocer y administrar las actividades relacionadas con la divulgación y uso de la información científica en las instituciones facilitando la identificación de las necesidades informacionales y de conocimiento de los usuarios, con miras a suplirlas mediante el análisis de los datos sobre el uso los recursos electrónicos que estos prefieren en sus búsquedas de información y si estas son efectivas.
- **Business Intelligence (BI):** El análisis comportamental de los usuarios de cualquier producto o servicio es una tarea hacia la cual, las organizaciones están dirigiendo su atención, pues el conocimiento de los clientes – usuarios, sus hábitos, patrones de compra o uso de servicios permiten diseñar y desarrollar programas y productos según las necesidades de los clientes. Las bibliotecas no son ajenas a este fenómeno de mercadeo y a través de los datos almacenados en ELOGIM es posible evidenciar tendencias y patrones comportamentales que ayuden a identificar y predecir las necesidades informacionales de los usuarios, a generar perfiles o a crear o modificar productos y servicios.
- **Definir indicadores de gestión (KPI's):** Los indicadores de gestión y la medición del desarrollo de procesos favorecen la gestión del conocimiento en las unidades de información en la medida en que sean un insumo útil de información que fundamente la toma acertada de decisiones, facilite los procesos, mejore los resultados y aporte a las actividades bibliotecarias administrativas. Los datos por si solos carecen de significado, se hace necesario procesarlos y darles valor lógico para convertirlos en indicadores que permitan una interpretación profunda de los procesos y evidencien la calidad de la gestión. En este sentido ELOGIM posee un alto porcentaje de certeza en la calidad de los datos debido a su interconectividad con los demás

sistemas universitarios, a la unificación del acceso de los usuarios a los recursos electrónicos a través de una sola contraseña, y la administración autónoma la cuenta de acceso de cada usuario. (ISO, 2014)

## Referencias bibliográficas

Gartner. (2016). Big Data. Recuperado 11 de febrero de 2016, a partir de

<http://www.gartner.com/it-glossary/big-data/>

IBM Global Business Services. (2013). Analytics: el uso de big data en el mundo real Cómo las empresas más innovadoras extraen valor de datos inciertos. Recuperado a partir de

[ibm.co/1QYqewy](http://ibm.co/1QYqewy)

ISO. (2014). ISO 11620:2014. Recuperado 11 de febrero de 2016, a partir de

<https://www.iso.org/obp/ui/#iso:std:iso:11620:ed-3:v1:en>

Marr, B. (2015). *Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance* (1 edition). Chichester, West Sussex, United Kingdom ; Hoboken, New Jersey: Wiley.

Mitchell, I., Locke, M., Wilson, M., & Fuller, A. (2013). *The White Book of... Big Data: The Definitive Guide to Big Data*. Fujitsu Services Ltd. Recuperado a partir de

<http://www.fujitsu.com/global/Images/WhiteBookofBigData.pdf>

Quiroz, N. L. (2012). Aplicación del proceso de kdd en el contexto de bibliomining: El caso

Elogim. *Revista Interamericana de Bibliotecología*, 35(1), 97-108.