

## **La subutilización de los datos académicos en las Universidades**

José Gerardo Moreno Salinas  
[gerardo\_moreno@cuaed.unam.mx]  
Coordinación de Universidad Abierta y Educación a Distancia – UNAM

### **Resumen**

En las anteriores intervenciones en Virtual Educa se ha defendido la tesis de la importancia de contar con las bases de datos de los alumnos (antecedentes académicos, sociodemográficos y económicos), así como el conocimiento de la técnica para realizar los procesos de análisis y hacer llegar más información y mejor conocimiento a los responsables de los programas académicos, y que sean éstos los encargados de definir las estrategias a seguir en beneficios de todos los alumnos de la Universidad. Ahora el discurso ha cambiado, y se defiende la idea que los principales tomadores de decisión deben ser los mismos alumnos, ya que éstos son los que están en condiciones en hacer los cambios o modificaciones en el corto plazo. Además de ser lo principales interesados de su trayectoria académica en sus respectivas carreras.

Es por lo anterior, que en el presente documento se resalta que cada vez vivimos en un mundo más conectado, donde los datos que generamos, transmitimos y almacenamos son cada vez mayores, sino sólo hay que ver el actual problema que tenemos con nuestros dispositivos móviles, particularmente los celulares que cada vez debemos estar depurando nuestras tarjetas de almacenamiento. También describo algunos de los principales conceptos relacionados con la ciencia de datos y su importancia en el análisis de los datos.

Además se presentan algunos ejemplos de cómo algunas organizaciones se han beneficiado de considerar al dato como un activo estratégico y por otro lado, tenemos el caso de las universidades que particularmente subutilizan el valor de dato académico de los alumnos, siendo que éstos pueden beneficiarse, y se invita a la reflexión con el planteamiento de una serie de preguntas.

Por último, se comenta la necesidad de una estructura en el uso y manejo del dato académico en la Universidad, particularmente para el caso de la Universidad Nacional Autónoma de México (UNAM).

*Palabras clave: Subutilización del dato, ciencia de datos, datos académicos, universidades.*

## Introducción

El mundo que ahora habitamos se ha transformado de manera radical, vivimos en una aldea cada vez más global, donde pareciera que se desvanecen las fronteras físicas y privilegiamos los puentes digitales al momento de comunicarnos, relacionarnos, divertirnos (consumir entretenimiento), estudiamos, compramos, opinamos ante los medios y la sociedad. Es decir, cada vez nuestro mundo está más conectado, tenemos acceso a más redes y servicios. Ahora más que nunca consumimos y producimos mayores cantidades de datos e información. En el mundo digital que ahora vivimos y con la tecnología que contamos hace más fácil crear, transmitir, almacenar, acceder y usar los datos (Bhansali, 2013).

De acuerdo con un estudio realizado por la empresa “International Data Corporation” (IDC, por sus siglas en inglés), publicó el artículo: El universo digital de las oportunidades, donde Turner (2014) señala que al igual que el universo físico, el universo digital es grande. Y estiman que para el 2020 habrá tantos bits digitales como estrellas hay en el universo. El universo digital está duplicando su tamaño cada dos años y en 2020 los datos que creamos y transmitimos alcanzarán los 44 ZB ( $10^{21}$ ).

Ahora bien, no todos los datos están sujetos a hacer analizados. Podemos distinguir los que son del tipo estructurado, semiestructurado y los no estructurados. Los primeros son los referentes a los datos relacionales, es decir, filas y columnas claramente identificables. Los semiestructurados son los que tienen un tipo de estructura implícita, por ejemplo: datos espaciales, temporales y textuales. Y los no estructurados son los que provienen principalmente de sitios web y/o son del tipo multimedia, como son los audios, imágenes y vídeos.

Para el caso de los datos académicos a los que nos referimos (antecedentes académicos, sociodemográficos y económicos) son del tipo estructurado, ya que residen en bases de datos relacionales. Y sí pueden ser sujetos a procesos automáticos sin mayor problema, a diferencia de los semiestructurados y no estructurados. De lo anterior, mucho se ha hablado sobre qué técnicas utilizar para el análisis de los datos, pero también resulta importante mencionar e impulsar el valor que tienen los datos académicos.

Las organizaciones deben saber aprovechar al máximo la información y explorar de manera inteligente cómo pueden beneficiarse del análisis de los datos que generan sus usuarios, operaciones, productos o servicios. No hay que olvidar que ahora, más que nunca, el recurso intangible más valioso en nuestros tiempos es el poder de la información y del conocimiento que obtengamos de éste.

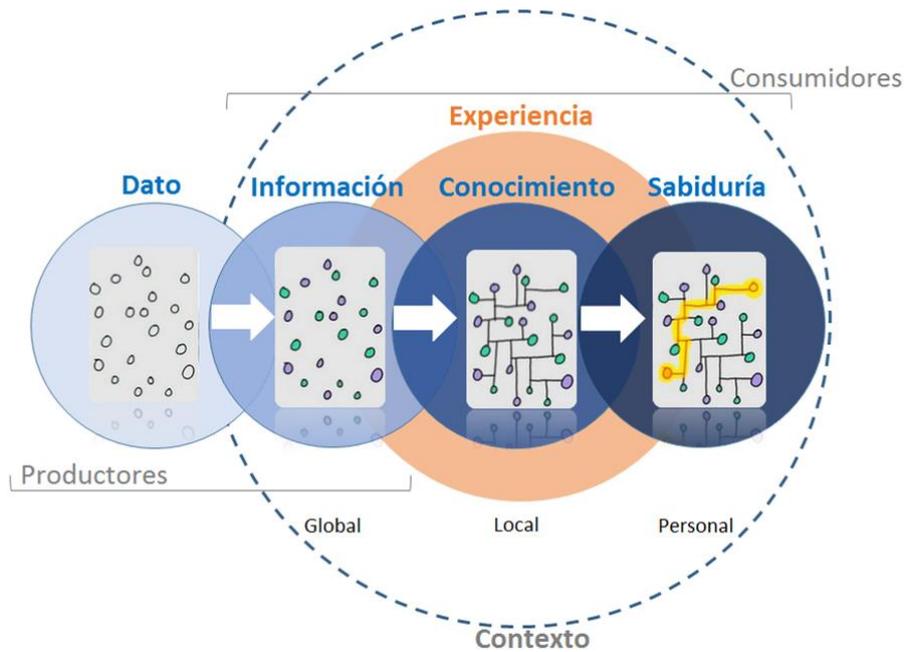
## Los conceptos emergentes

Recientemente hemos visto el surgimiento de nuevos conceptos que incluyen las áreas de la estadística, matemáticas y cómputo, principalmente. Es el caso de la ciencia de datos, “Big Data”, minería de datos y visualización de datos, por mencionar algunos. Los cuales vienen a enmarcar las técnicas, los análisis, el software especializado, las formas y los métodos en los que se almacenan, procesan y presentan los datos para el mejor entendimiento en términos de su variedad de formatos y cantidad.

Comenzaremos por definir que los científicos de datos son los trabajadores del conocimiento, los cuales tiene la responsabilidad de entender en su máxima expresión los datos y sus relaciones, con el objetivo de tomar decisiones más informadas a la vez que mejoran los productos y servicios de las organizaciones (Moreno, 2017).

En términos generales, el científico de datos combina estadística, matemáticas, programación y solución de problemas, con la captura datos de forma ingeniosa y la capacidad de mirar las cosas de manera diferente (encontrar patrones), además de hacer las actividades propias de limpieza, preparación e integración de datos (Monnapa, 2017). Cabe mencionar que los autores Davenport y Patil en su artículo “Data Scientist: The sexiest job of the 21st century” (2012), definieron por primera vez el concepto de científico de datos e hicieron notar su gran pertinencia en la valoración de los datos y su importancia en la toma de decisiones de cualquier organización.

Es pertinente señalar que el *dato* es la representación simbólica básica de un atributo o variable, que al procesarlos en su conjunto obtenemos información de acuerdo a su contexto, es decir, datos con significado. Al apropiarse y entender la *información* generada por los datos, más las experiencias y una comprensión teórica o práctica dan lugar al *conocimiento*, mismo que da la certidumbre cognitiva mensurable a las posibles repuestas a ¿por qué?, ¿cómo?, ¿cuándo? y ¿dónde? Por último, llegar al nivel más alto que puede aspirar cualquier conjunto de datos, es decir, *sabiduría* la cual tiene la característica de la aplicación de la inteligencia en la experiencia propia, obteniendo conclusiones que nos dan un mayor entendimiento y a su vez nos capacitan para reflexionar. Con lo anterior hacemos alusión al modelo jerárquico de datos, información, conocimiento y sabiduría, mejor conocido por sus siglas en inglés (DIKW). Que va desde un nivel básico a lo más complejo, la siguiente imagen representa su relación estructural y funcional.



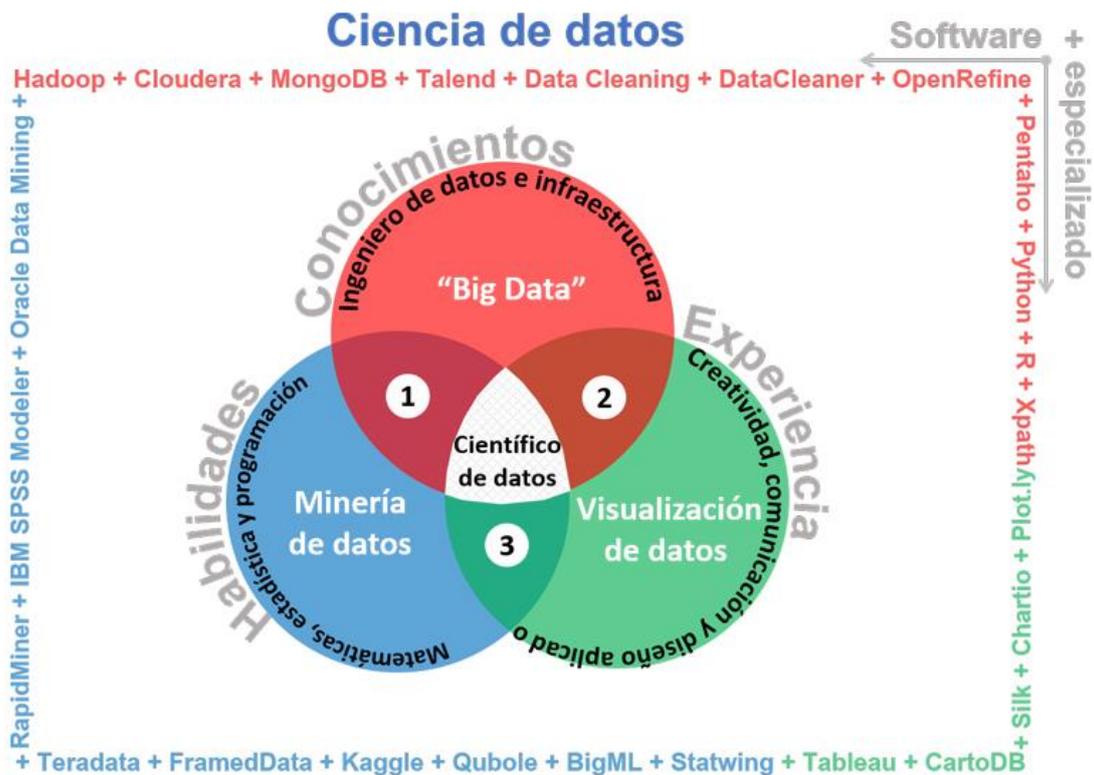
Jerarquía DIKW.

Fuente: Diseño elaborado con base a lo expuesto por Rowley, 2007.

Principalmente son tres las áreas en las que se desarrollan los científicos de datos:

- Big Data para procesar datos.
- Minería de datos para analizar e identificar relaciones ocultas, patrones y tendencias.
- Visualización de datos para explicar y socializar mejor la información obtenida.

Por cada una de estas áreas existen una amplia gama de técnicas y software especializado que el científico de datos utiliza para desarrollarse, de las cuáles se pueden clasificar por herramientas de extracción, almacenamiento, limpieza, minería, visualización, programación, análisis e integración de datos. En la siguiente figura se presentan las habilidades, los conocimientos y la experiencia que debe poseer el científico de datos, así como una muestra del software especializado y técnicas que existe por sus áreas de desarrollo.



**Técnicas:** Regresión lineal, Reconocimiento de patrones, Series de tiempo, Árboles de decisión, Estadística Bayesiana, Redes neuronales, Aprendizaje supervisado y no supervisado, K-NN, Sistemas de recomendación, Modelos predictivos, Teoría de juegos, Aprendizaje profundo...

El científico de datos y su entorno.  
Fuente: elaboración propia.

1. Big Data. Los investigadores Cox y Ellsworth (1997) de la Administración Nacional de la Aeronáutica y del Espacio (NASA por sus siglas en inglés), fueron los primeros en utilizar el término Big Data en un artículo científico, en el que señalaron el problema al que se enfrentaban al visualizar y el procesar grandes cantidades de datos, así como las limitantes técnicas de las computadoras (gráficos, memorias y almacenamiento) que tenían en esos tiempos. Ha sido un término que, al igual que su nombre, ha tenido una gran aceptación en todas las industrias y son muchas las definiciones<sup>1</sup> que existen al respecto en particular me gusta definir el concepto de Big Data como: “el gran cúmulo de datos compuesto por diferentes tipos, estructuras y relaciones de datos, que a su vez tienen veloces tasas de generación y dispersión, y el procesarlos con tecnologías convencionales para su posterior análisis es parte del problema (Big Problem)”.
2. Minería de datos. En términos sumamente prácticos la minería de datos la podemos definir como a la extracción de conocimientos de grandes cantidades

<sup>1</sup> Para conocer la recapitulación de las diferentes definiciones hechas al concepto Big Data, se sugiere revisar el sitio de la Escuela de Información de Berkeley, E.U., disponible en: <https://datascience.berkeley.edu/what-is-big-data/#List>

de datos. Han y Kamber (2006), hacen una interesante crítica al concepto de minería de datos: “la extracción de oro de las rocas o la arena se conoce como minería de oro en lugar de minería de roca o arena. Por lo tanto, la minería de datos debería haber sido más apropiadamente llamada minería del conocimiento a partir de datos. Sin embargo, la minería es un término vívido que caracteriza al proceso de encontrar un pequeño conjunto de preciosas pepitas en una gran cantidad de materia prima”. La minería de datos en principio trabaja sobre todo tipo de datos. Los estructurados se refieren a las bases de datos relacionales (filas y columnas claramente identificadas); los semiestructurados son los que tienen un tipo de estructura implícita, pero no como para ser automatizada como la estructurada (datos espaciales, temporales y textuales); los no estructurados son los que principalmente provienen de sitios en Internet y son del tipo multimedia (imágenes, audio y videos). Los dos últimos se identifican con la minería de textos y la minería web, respectivamente.

3. Visualización de datos. Es considerada por algunos como una ciencia y hay quienes la clasifican como un arte, cuando en realidad es una combinación de ambas. Sus principales precursores justo provienen de las ciencias exactas, que han tenido la necesidad de recurrir al campo de la creatividad y del arte, con el propósito de representar con fines estéticos algún aspecto de la realidad (Moreno, 2017). La visualización de datos sólo tendrá éxito en la medida que nuestros ojos codifiquen la información para poder discernirla y nuestros cerebros la pueden entender. El objetivo es traducir de maneras fáciles, eficientes, precisas y decodificadas la información abstracta en representaciones visuales significativas (Few, 2013).

### **La utilización del dato por las organizaciones**

Cada vez son más las organizaciones que están considerando al dato como un activo estratégico, como bien señala Bhansali (2013), *los datos se han convertido en la base de la competencia, productividad, crecimiento e innovación de las organizaciones*. Por mencionar algunos ejemplos, las compañías como Google, Facebook y Twitter gastan increíbles cantidades de dinero para mantener sus sistemas, sin embargo, los usuarios finales no son quienes pagan directamente esos gastos, en lugar de ello proveen contenido a la vez que son objeto de ambiciosas campañas publicitarias, lo que significa que otras compañías están pagando los costos de infraestructura a cambio de obtener datos de los usuarios (Van der Aalst, 2014).

Para Twitter existen aplicaciones web donde se calcula el valor que tiene una cuenta, lo cual es un estimado con base al número de seguidores que tengas, la cantidad de personas que te siguen, los tweets que escribes y la velocidad con la que ganas seguidores. Por ejemplo, al hacer la prueba en los sitios tvalue.com y tweetvalue.com reportaron que mi cuenta en Twitter (@jgmorenos) está valuada en \$18.47 y \$44 dólares, respectivamente. Recientemente el analista Cakmak (2017), analizó el valor que tiene para Twitter la cuenta de Donald Trump (37.4 millones de seguidores con más de 35 mil Tweets) y la calculó en 2 mil millones de dólares. Hay que considerar que estos valores son estimaciones y habrá que tomarlos con reserva, pero al menos son una invitación para reflexionar y no subestimar el valor que tienen los datos.

Son varios los casos de éxito en donde las compañías se han beneficiado por codificar el valor oculto que tienen sus datos, para así mejorar sus productos y servicios, principalmente (Moreno, 2017). Por ejemplo, Netflix ha sabido utilizar bien sus datos, pues tiene como objetivo principal: “ayudar a sus suscriptores a encontrar el contenido que realmente disfrutan, maximizando así su satisfacción y retención” (Elahi, 2015). Desde sus inicios en 1997, con el servicio de renta y envío de DVD por correo postal, le dio una gran importancia a los datos de sus usuarios y en 2000 comenzó a desarrollar lo que sería su primer algoritmo (Cinematch) para crear un sistema que permitiera recomendar contenido de alto interés para cada uno de sus suscriptores. En el 2006, Netflix abrió su algoritmo a la comunidad científica y ofreció una recompensa de 1 millón de dólares para quién(es) lograran mejorar en un 10% su capacidad predictora, tuvieron que pasar tres años para que el grupo BellKor's Pragmatic Chaos lograra resolverlo. En 2007 comenzó con su servicio de descarga y reproducción (streaming) de películas y series.

El sitio statista.com reportó que en el segundo cuatrimestre de 2017 Netflix tiene 103.9 millones de suscriptores a nivel mundial, de los cuales procesa en promedio 695 mil millones de eventos por día, es decir, una base de datos de 1.8 Petabytes diarios.4 Algunos de los eventos registrados por Netflix son:

- ¿Desde dónde se conectan?
- ¿A través de qué dispositivo?
- ¿En qué horarios se conectan?
- ¿El tipo de contenido (película, serie) varía con el dispositivo?
- ¿Ven los créditos?
- ¿Cuánto tardan en ver el contenido?
- ¿Cuáles son sus actores y directores favoritos?
- ¿Qué y cómo califican?
- ¿Qué buscan?
- Etcétera.

### **La subutilización del dato en la Universidad**

El considerado padre del “management”, Peter Drucker (2004), reconoció que la sociedad postcapitalista es una sociedad basada en el conocimiento, donde el centro de la producción de la riqueza es el saber y no el capital. Los protagonistas claves en esta economía del conocimiento serán los “trabajadores del conocimiento”, es decir, los que posean las capacidades, las habilidades, el pensamiento creativo y la tecnología para procesar, analizar y visualizar las grandes bases de datos.

Así también, la Universidad Nacional Autónoma de México (UNAM) tiene presente la importancia del dato y lo estratégico que éste puede ser para los alumnos, profesores y administradores de los servicios educativos. La Universidad para el caso particular de los alumnos, registra:

- Antecedentes académicos
- Perfiles sociodemográficos y económicos al momento del ingreso,
- Cuántas y cuáles asignaturas inscriben a cada semestre,
- Calificaciones que éstos promedian
- Frecuencia con la que usa los servicios bibliotecarios y los títulos que más consulta,
- Periodicidad con la que se conecta a la red inalámbrica universitaria (RIU) y desde qué tipo de dispositivos

- Asignaturas con mayor índice de reprobación y aprobación
- Tiempos que le toma para titularse de la carrera
- Entre otros tantos.

Y para el caso de los alumnos en la modalidad a distancia, además de los anteriores se puede saber:

- Frecuencia con la que ingresan a sus respectivas plataformas
- Tiempo de permanencia
- Cómo se expresan en los foros y chat
- La forma en cómo avanzan (paralelo o serial)
- Red de contactos o comunidades de estudio
- Entre otros.

El obtener información de todos estos datos y que se pueda traducir en conocimiento, sin duda empoderará a cada uno de los involucrados (alumnos/profesores/administradores de los servicios educativos), lo que posibilitará una mejor toma de decisiones. Mucho se ha hablado sobre la personalización de la educación y de la importancia de considerar a cada alumno como un individuo diferente a todos los demás, el cual tiene necesidades diferentes y no considerarlo como otro elemento más del grupo, *“nunca debemos estar interactuando con los estudiantes como si fueran todos iguales”* Sclater (2017).

Como podemos notar para el caso particular de los alumnos, son muchos los datos que la UNAM registra y procesa para generar información de manera concentrada, por ejemplo está la agenda estadística que cada año la Dirección General de Planeación publica y se pueden consultar la estadística descriptiva por nivel y modalidad en cuanto a la población escolar, egreso, exámenes de grado, entre otros más, ver en: <http://www.planeacion.unam.mx/estadistica.php>.

Es claro que la Universidad tiene un alto nivel de granularidad de la información, lo que posibilitaría cruzar datos entre diferentes bases, pero por razones de seguridad y cuidar la identidad de los usuarios/alumnos, es claro que no todos tenemos acceso a los mismos datos y en caso de tenerlos, no son claros los lineamientos para usarlos, incluso ni para los mismos dueños de los datos.

Cómo hacer que las diferentes entidades de la Universidad ofrezcan el valor del dato procesado y sean los mismos usuarios/alumnos los empoderados de su propia información y de los procesos de enseñanza- aprendizaje, ya que éstos son los que pueden aplicar los posibles cambios en el corto plazo inmediato. Por ejemplo, podrán hacer una mejor selección de sus asignaturas, definir sus cargas académicas, compararse con el grueso de la población escolar, por mencionar algunos.

Por lo anterior, surgen las siguientes preguntas.

- ¿Cómo utilizan los datos las diferentes entidades de la Universidad que sí cuentan con los privilegios de acceso?
- ¿Cuál es la información y conocimientos que obtienen de los datos?
- ¿Qué marcos regulatorios manejan para el procesamiento de datos?
- ¿Qué técnicas/modelos/software utilizan para analizar y visualizar los datos?

- ¿Qué beneficios o ventajas obtienen los usuarios/alumnos de sus propios datos? y ¿cómo pueden utilizarlos para una mejor toma de decisiones?
- ¿Qué beneficios traerá a la Universidad en contar con una estructura en gobierno de datos?

Todo lo anterior con el firme propósito de tener informado al tomador de decisiones de primer orden en su tránsito académico por la Universidad, es decir, el alumno.

### **La necesidad de una estructura en el uso y manejo del dato académico en la Universidad**

Las altas tasas de abandono es uno de los principales problemas que aquejan a las universidades, y para el caso de los programa a distancia el problema es aún más grave. Yolanda Gayol (2016) presentó un comparativo entre las tasas de deserción de las principales universidades con programas en la modalidad a distancia en América Latina, donde hace evidente el gran problema que enfrentan las instituciones educativas, además de una interesante crítica sobre la utilización de la palabra deserción y sus implicaciones. En particular, para la modalidad a distancia Yukselturk (2014), investigó lo que varios autores dijeron estar de acuerdo *“las tasas de abandono para la educación a distancia son generalmente más altas que para la educación convencional. Muchos estudiantes fácilmente están abandonando sus cursos y programas en modalidades a distancia o en su defecto, los terminan insatisfechos”*.

El costo de la pérdida de un alumno es muy alto en términos de pérdida de tiempo, esfuerzo y dinero por parte de todos los involucrados (alumnos, maestros y universidades).

La Universidad no se ha visto ajena a los grandes cambios en cuanto al procesamiento y análisis de datos. Desde un plano académico ofrece programas de posgrado donde incentiva la investigación en el procesamiento y análisis de datos, por ejemplo están los Posgrados en Ciencias e Ingeniería de la Computación y en Ciencias Matemáticas y de la Especialización en Estadística Aplicada, respectivamente.

En cuanto a la protección de los datos que procesa la Universidad, en agosto del 2011 publicó en la gaceta universitaria el Reglamento de Transparencia y Acceso a la Información Pública de la Universidad Nacional Autónoma de México<sup>2</sup>, con la finalidad de establecer los órganos, criterios y procedimientos institucionales que garanticen, a toda persona, la transparencia, el acceso a la información pública y la protección de datos personales en posesión de la Universidad. Entre sus disposiciones establece en el capítulo IV (de la protección de datos personales) artículo 12, las Unidades Universitarias desarrollarán o tendrán sistemas de datos personales sólo cuando éstos se relacionen directamente con sus facultades o atribuciones reglamentarias; artículo 13, no se requerirá el consentimiento de los involucrados para proporcionar los datos personales en los necesarios con fines estadísticos, científicos, históricos o de interés general, previo procedimiento por el cual no puedan asociarse los datos personales con el individuo a quien se refieran. Lo anterior sí posibilita que el propio

---

<sup>2</sup> Reglamento recuperado de <https://goo.gl/2efiMA>.

individuo/alumno pueda relacionar sus datos con los demás, y haga un uso responsable de sus propios datos en su beneficio, con la reserva que siempre se deberá cuidar la confidencialidad de los demás alumnos, es decir, nula posibilidad a ser identificados ni identificables.

De acuerdo con Porter (1979), hay cinco fuerzas que afectan de manera directa el nivel de competencia de una organización, es decir: la amenaza de nuevos competidores, la amenaza de productos sustitutos, la rivalidad entre competidores, el poder de decisión de los compradores y poder de negociación de los proveedores. Evidentemente para el caso que aquí tratamos las primera tres fuerza no aplican en la Universidad, pero si consideramos a la cuarta como el poder de decisión de nuestros alumnos en lugar de los compradores. Al igual si consideramos el poder de negociación de los proveedores con el que tienen los profesores, diseñadores y desarrolladores de los planes y programas de estudio, cobra un mayor sentido para nuestro caso particular.

Es necesario contar con otros enfoques o marcos de referencia, que consideren la complejidad y riqueza del dato y de cómo éste llega a definir la huella digital<sup>3</sup> de cada alumno -para el caso de la Universidad-, así como el valor e importancia que ha venido adquiriendo en los últimos años. Así como tener clara una estructura en el uso y manejo de datos académicos entre entidades universitarias.

Por un lado la Universidad cuenta con los posgrados para el análisis y procesamiento de datos, así con un reglamento de transparencia y acceso a la información que sea de carácter público, pero también es cierto que la Universidad tiene la necesidad de contar con un marco de referencia que direcciona un mejor uso y manejo del dato entre las diferentes entidades, en términos de gobernanza de datos, en beneficio de los alumnos, no olvidemos que son los principales tomadores de decisiones en su proyecto formativo académico.

En tanto más y mejor conozcan los alumnos su situación en la Universidad y sean sus propios datos (historias académicas, perfiles de ingreso, trayectorias académicas) los que ayuden a marcar la pauta del avance que deberán tener en sus estudios, directamente se estará abonando hacia una mejor retención de estudiantes. Los datos pueden contener información esencial para mejorar la calidad de la educación, con el gran potencial de personalizar las experiencias de aprendizaje de los alumnos (Zhao y Luan, 2006).

De acuerdo con Bhansali (2013), ha probado que las organizaciones se han visto beneficiadas por diseñar e implementar una estructura en gobierno de datos, especialmente en las que registran y procesan grandes cantidades de datos. Además asegura que al tener un efectivo gobierno de datos reduce la incertidumbre y ayuda a mejorar el desempeño de las organizaciones. Por lo que resulta interesante demostrar los beneficios que tendrán tanto la Universidad y sus involucrados al establecer los lineamientos del uso responsable del dato, así como diseñar una estructura que soporte el gobierno de datos. Donde seguramente la Universidad se verá favorecida

---

<sup>3</sup> Es el conjunto único de actividades digitales, acciones y comunicaciones que dejan una traza de datos en Internet o en una computadora u otro dispositivo digital y pueden identificar al usuario o dispositivo en particular. Concepto recuperado de dictionary.com

en mejorar las tasas de retención, así como una clara identificación de los alumnos que están en vías de recuperación y motivarlos a retomar sus estudios.

## **Conclusiones**

Es vital profundizar sobre la problemática en el uso responsable de los datos en la Universidad y cómo es que generan información y conocimiento en beneficios de los alumnos, es decir, cómo dotarlos de elementos que propicien una más y mejor toma de decisiones. A la vez que la Universidad tiene una mejor estructura en el procesamiento de los datos.

En la Universidad no hay claridad en cuanto las líneas que habrá que seguir referente al manejo y el uso del dato, para su posterior análisis y obtención de información y conocimiento. Estoy seguro que todas las entidades de la Universidad hacen su mejor esfuerzo con los académicos y personal administrativo con el que cuentan para hacer frente a los procesos cotidianos, como es: registro escolar, asignación de grupos, determinación de cursos remediales, etcétera. Pero también hay que decirlo, no hay entidad o dependencia en la Universidad que no se muestre celosa de sus datos, habrá que transformar la vieja idea que son los dueños de los datos, son los poseedores, los dueños de los datos son cada uno de los alumnos.

Con base a la experiencia propia, muchas de las veces no es suficiente contar con los conocimientos técnicos y metodológicos del almacenamiento, procesamiento, y visualización del dato, se requiere de una misma base sobre la que sustente el uso generalizado del dato, sin comprometer su integridad y asegurando la confiabilidad del mismo.

## **Bibliografía**

- Bhansali N. (2013), Data Governance: Creating Value from Information Assets, CRC Press, 268 pages, ISBN: 978-1-4398-7913-9.
- Cakmak, J. (2017). What Is Trump Worth to Twitter? One Analyst Estimates \$2 Billion. Fortune.com. Disponible en <<https://goo.gl/AdjDQz>>.
- Cox, M. y Ellswort, D. (1997). Managing Big Data for Scientific Visualization. ResearchGate.net. Disponible en <<https://goo.gl/DLj8sd>>.
- Davenport, T. H. y Patil, D. J. (2012). Data Scientist: The Sexiest Job of the 21st Century. Harvard Business Review, 90 (10). Disponible en <<https://goo.gl/65IMw1>>.
- Drucker, P. F. (2004). La sociedad postcapitalista. Medellín, Colombia: Norma
- Elahi, E. (2015). Spark and GraphX in the Netflix Recommender System. SlideShare. Disponible en <<https://goo.gl/LUQqx4>>.
- Few, S. (2013). Data Visualization for Human Perception. En M. Soegaard (2nd Ed.), The Encyclopedia of Human-Computer Interaction. Aarhus, Denmark: The Interaction Design Foundation. Disponible en <<https://goo.gl/7uYrrp>>.
- Gayol, Y. (2016). Deserción: Explorando resultados en la educación a distancia. Disponible en <<https://goo.gl/nYEMg1>>.

Han, J. y Kamber, M. (2006). *Data Mining Concepts and Techniques*. Disponible en <<https://goo.gl/jmXNua>>.

Monnappa, A. (2017). *Data Science vs. Big Data vs. Data Analytics*. Simplilearn.com. Disponible en <<https://goo.gl/EAYQRc>>.

Moreno, José G. (2017). "Científico de datos: codificando el valor oculto e intangible de los datos", en *Revista Digital Universitaria (RDU)*, vol. 18, núm. 7, septiembre-octubre. Disponible en <<https://goo.gl/LR58UC>>.

Porter M. (1979), *How Competitive Forces Shape Strategy*. Harvard Business Review. Disponible en <<https://goo.gl/Pnua9p>>.

Rowley, Jennifer (2007). "The wisdom hierarchy: representations of the DIKW hierarchy". *Journal of Information and Communication Science*. 33 (2): 163–180. Disponible en <<https://goo.gl/7ndWGs>>.

Slater N., Mckay T. (2017), *Learning Analytics Explained*, Routledge, 278 pages, ISBN: 978-1-13893-172-5.

Turner V. (2014), *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*. IDC Analyze the Future. Disponible en <<https://goo.gl/VEWutA>>.

Van der Aalst, W. M. P. (2014). *Data Scientist: The Engineer of the Future*. *Enterprise Interoperability*, volume 7, 13-28. Springer. Disponible en <<https://goo.gl/yiaE9F>>.

Yukselturk, E; Ozekes S. y Türel Y. (2014). Predicting Dropout Student: An Application of Data Mining Methods in an Online Education Program. *European Journal of Open, Distance and E-Learning*. Volume 17, Issue 1, Pages 118–133, ISSN (Online) 1027-5207.

Zhao, C. y Luan, J. (2006). *Data mining: Going beyond traditional statistics*. In *New Directions for Institutional Research*, 131(2), (pp. 716).