

¿Cuál es la metodología para consolidar datos en una cultura de paz y reconciliación en Colombia?*

What is the Methodology for Consolidating Data in a Culture of Peace and Reconciliation in Colombia?

John Velandia Vega**

Marco Emilio Sánchez Acevedo***

Fecha de recepción: 5 de mayo del 2020

Fecha de aprobación: 25 de mayo del 2020

RESUMEN

La cultura de paz y reconciliación en Colombia ha producido grandes volúmenes de datos en los últimos años a partir de diferentes fuentes de información. Ahora es primordial definir una metodología que garantice una forma ordenada y estructurada para abordar proyectos de investigación que requieran consolidar y limpiar datos para la toma de decisiones informadas. Esta investigación busca responder a la siguiente pregunta: ¿cuál es el proceso que permite consolidar datos de forma estructurada y gráfica para la toma de decisiones de manera informada en el marco de paz y reconciliación en Colombia? Como resultado de esta investigación, se presenta una metodología que nace a partir del análisis de diferentes investigaciones de todo el mundo.

-
- Artículo resultado del proyecto de investigación “Derecho, cambio climático y *big data*” gestionado en la Universidad Católica de Colombia.
 - Citar como: Velandia Vega, J. y Sánchez Acevedo, M. E. (2020). ¿Cuál es la metodología para consolidar datos en una cultura de paz y reconciliación en Colombia? *Via Inveniendi et Iudicandi*, 15(2), 131-163. doi: <https://doi.org/10.15332/19090528/6245>
 - Magíster en tecnologías de la información e investigador del Grupo de Investigación Software Inteligente y Convergencia Tecnológica (GISIC) de la Facultad de Ingeniería de la Universidad Católica de Colombia. Correo: javelandia@ucatolica.edu.co. Magíster en Tecnologías de la Información ORCID: <https://orcid.org/0000-0001-9217-4726>
 - Doctor en Derecho; investigador del grupo Derecho Público y TIC de la Facultad de Derecho de la Universidad Católica de Colombia. Correo: mesanchez@ucatolica.edu.co. ORCID: <https://orcid.org/0000-0002-7745-2182>

Palabras clave: metodología, analítica de datos, paz, reconciliación, decisiones informadas.

ABSTRACT

In recent years, the culture of peace and reconciliation in Colombia has produced large volumes of data coming from different sources of information. It is now a matter of paramount importance to define a methodology that ensures an orderly and structured way to address research projects that require consolidating and clearing data for informed decision-making. This paper seeks to answer the following question: what is the process that allows to consolidate data in a structured and graphic way for informed decision-making in the framework of peace and reconciliation in Colombia? As a result of this paper, a methodology that is born from the analysis of different research papers from around the world is presented.

Keywords: methodology, data analytics, peace, reconciliation, informed decisions.

INTRODUCCIÓN

En el marco del proyecto “Derecho, cambio climático y *big data*”, patrocinado por la Universidad Católica de Colombia y en el cual participan los grupos Investigación en Derecho Público y TIC, de la Facultad de Derecho, y el Grupo de Investigación de Software Inteligente y Convergencia Tecnológica (GISIC), de la Facultad de Ingeniería, con la participación además de investigadores externos de la Universidad de Texas (EE. UU.), se plantea un análisis del estado actual de analítica de datos y su representación gráfica, orientado a facilitar la toma de decisiones en los aspectos de cultura de paz y reconciliación a través de contenidos basados en *big data* y en datos abiertos.

La expresión *big data* se refiere a un ingente volumen de datos, de toda variedad, que se procesan a una velocidad impensable, de manera tal que se logre así un potencial de valor incalculable (Joshi, 2015; Zhou, Pan, Wang y Vasilakos, 2017). Esta gestión de datos masivos se proyecta en los sectores público y privado para generar y desarrollar negocios, fijar estrategias de marketing, consumidores y contratación para tomar decisiones en los ámbitos más variados de la salud, la educación y las políticas públicas sociales, culturales, de vivienda, de defensa, de seguridad y de criminalidad. Se trata de un auténtico nuevo mundo por explorar que no solo tiene decisivas consecuencias económicas, sino también sociales, políticas y jurídicas (Bertot, Gorham, Jaeger, Sarin y Choi, 2014; Khan, Liu, Shakil y Alam, 2017).

En ese sentido, y con motivo de la variedad de datos existentes, Colombia reconoce el derecho de acceso a documentos públicos (Sánchez Acevedo, 2015), que reviste por lo menos tres funciones esenciales en el ordenamiento interno, a saber: 1) garantizar la participación democrática y el ejercicio de los derechos políticos; 2) cumplir una función instrumental para el ejercicio de derechos constitucionales de las víctimas, ya que permite conocer las condiciones necesarias para su cumplimiento y 3) garantizar la transparencia de la gestión pública; por lo tanto, ello se constituye en mecanismo de control ciudadano de la actividad estatal (C. C., Sentencia C-274-2013, Colom.).

Asimismo, el derecho de acceso a documentos públicos impone al menos dos deberes correlativos a todas las autoridades estatales. En primer lugar, para garantizar el ejercicio de este derecho, las autoridades públicas tienen el deber de suministrar a quien lo solicite

información clara, completa, oportuna, cierta y actualizada sobre su actividad (Rengifo Lozano, 2011). En segundo lugar, también es necesario que las autoridades públicas conserven y mantengan la información sobre su actividad, ya que, de no hacerlo, vulnerarían el derecho de las personas al acceso a la información pública y, en consecuencia, el derecho a que ejerzan control sobre sus actuaciones.

Sin embargo, el mencionado derecho también cuenta con limitaciones en su ejercicio, fijadas por la ley, ya que se debe indicar la restricción de acceso y su negación ante una eventual solicitud, si es objeto de reserva. Esto se presenta con el fin de proteger derechos fundamentales o bienes constitucionalmente valiosos, como 1) la seguridad nacional, 2) el orden público, 3) la salud pública y 4) los derechos fundamentales, conforme al principio de proporcionalidad (Corte Constitucional, Sentencia C-274 de 2013).

Por otro lado, la promoción de una cultura de paz y de reconciliación es un eje fundamental para construir paz en Colombia, país que enfrenta un conflicto armado interno particularmente complejo. La larga duración y la complejidad de este conflicto armado interno hicieron que amplios sectores sociales aceptaran el uso de violencia como mecanismo normal para resolver disputas y conflictos. Rasgos culturales que justifican y aceptan el uso de violencia, como el machismo o la exclusión de jóvenes y minorías étnicas, se establecieron en la cotidianidad de las relaciones, obstaculizando la convivencia pacífica y la reconciliación (Transparencia por Colombia, n. d.).

Dentro de las estrategias que contribuyen a la construcción de una cultura de paz, se identifican cinco (Le Blanc, 2016):

- fomento de la unidad y de la acción colectiva;
- promoción de diálogo y de participación informal;
- promoción de la identificación con el territorio y de la permanencia en él;
- promoción del uso de herramientas culturales y artísticas y de actividades lúdicas;
- promoción de los jóvenes como constructores de paz.

Por todo esto, se observa la necesidad de definir indicadores que apoyen las estrategias para promover la cultura de paz y reconciliación.

Los indicadores de paz se encuentran clasificados en conflicto armado, seguridad social y militarización. A partir de estos indicadores se definen métricas, como la del número de muertes del conflicto armado o el número de conflictos armados durante cierto periodo de tiempo, entre otros (Transparencia por Colombia, n. d.).

PROBLEMA

Los indicadores de paz en todo el mundo son métricas que se muestran a través de investigaciones (Mac Ginty, 2013) y reportes sociopolíticos (“Índice de Paz Global”, 2017). Estos indicadores tienen asociadas cifras como las del número de muertos y heridos asociados a los conflictos internos y externos de un país, el porcentaje de regiones geográficas donde se presentan los conflictos armados contra el número de zonas libres de acciones terroristas provocadas por un conflicto, el porcentaje de corrupción en las zonas donde se desarrolle el conflicto armado, el número de convenios entre partidos políticos que identifiquen pactos por la paz, el número de acuerdos firmados por la paz entre universidades y el Estado y muchas otras métricas que se suscitan alrededor de la paz y el posconflicto (“Índice de Paz Global”, 2017; Le Blanc, 2016; Mac Ginty, 2013).

Entidades públicas, privadas, mixtas y no gubernamentales se encuentran comprometidas con el apoyo a la paz y el posconflicto; sin embargo, estas entidades encuentran obstáculos en situaciones en que deben planificarse y ejecutarse recursos financieros y físicos, debido a que desde el Estado no se suministran datos centralizados que permitan tomar decisiones informadas para la inversión en temas de paz (García y Fino, 2014).

La carencia de datos e información no brindan confianza a ningún tipo de empresa que quiera orientar recursos financieros destinados a responsabilidad social; por lo tanto, esta investigación responderá a la siguiente pregunta de investigación: ¿cuál es el proceso que permite consolidar datos de forma estructural y gráfica para respaldar la toma de decisiones de manera informada en el marco de paz y reconciliación en Colombia?

DISCUSIÓN Y RESULTADOS

En este apartado se presenta la metodología de la ciencia de datos con el fin de suministrar las bases para abordar cualquier necesidad relacionada con la manipulación de estos. La metodología abarca cinco fases: formulación del problema, recopilación de datos, exploración de datos, modelamiento de datos y visualización de datos (Kayser, Nehrke y Zubovic, 2018; Szymańska, 2018).

Formulación del problema

La formulación del problema inicia con la identificación de una necesidad particular para luego abordar la solución mediante los siguientes pasos:

Definir el objetivo

Al definir el objetivo se sugiere preguntar a quién o quiénes impactará; es decir, cuáles son las partes interesadas (*stakeholders*) que estarán involucradas en el proceso de diseño y análisis de datos; por ejemplo: las personas que toman decisiones basadas en indicadores. Es crucial identificar a quienes realmente intervendrán en el proceso, ya que los líderes de proyectos tienden a incluir a personas que no tienen relación con el proceso que se va a desarrollar.

Por otra parte, también es importante responder a la pregunta de “cuándo”. En este caso, es vital identificar la variable tiempo, con el fin de conocer el periodo de los datos a recopilar y hasta cuando tendrán vigencia para sustentar la toma de decisiones informadas.

Es también primordial establecer la variable espacio para identificar los posibles lugares o fuentes de información. Incluso, el lugar físico o virtual donde se gestionan los datos y se guardan. Como consecuencia, los costos de almacenamiento y procesamiento de datos se identifican desde el inicio del proyecto, a fin de evitar costos ocultos en el futuro.

Por último, y no menos importante, también se debe identificar qué tipo de datos e información se quiere obtener, garantizando objetividad en los resultados requeridos.

Las variables de costo y tiempo son dependientes de los datos que se van a acopiar, por lo tanto, se recomienda considerar la aplicación de esta actividad.

Determinar las tareas que se ejecutarán con los datos obtenidos

La identificación y selección de las tareas que se ejecutarán a partir de los datos recogidos representan un conjunto de actividades fundamentales para desarrollar patrones y tendencias de información. Dependiendo de la entidad pública o del interesado, se establecen las tareas que se desplegarán con la información obtenida a partir del procesamiento de los datos. Por ejemplo, definir estrategias de mercadeo digital con el fin de aumentar el número de interesados en participar en convocatorias privadas o públicas; asimismo, la creación de una nueva línea de trabajo que ayude a mejorar los indicadores públicos, entre otros ejemplos.

En conclusión, la definición y ejecución de las tareas basadas en la información producida depende de la estrategia, las tácticas y los grupos de interés —o *stakeholders*— del proyecto. Generalmente, los resultados se enfocan en el incremento de ventas y la optimización de recursos.

Definir qué se quiere estimar o predecir

Esta actividad consiste en el establecimiento de indicadores de gestión y resultados que se construyen a partir de los datos suministrados por un proceso particular. Al igual que en el punto anterior, las organizaciones buscan optimizar el recurso humano, el tiempo y los costos de ejecución de procesos. Los indicadores más utilizados en las empresas de todo el mundo son los de eficiencia, eficacia y efectividad.

Recopilación de datos

La recopilación de datos consiste en identificar las fuentes de datos, sus tipos y su representación, con el fin de establecer un proceso repetitivo de obtención de datos de cómo almacenarlos para analizarlos en la siguiente etapa de exploración de datos (Gröger, Hillmann, Hahn, Mitschang y Westkämper, 2013).

Identificación de fuentes de datos

Establecer un único conjunto de fuentes de datos es imposible, debido a que existe un amplio rango de estos, según su naturaleza. A continuación, se presenta un resumen de la heterogeneidad de fuentes que se podrían encontrar en las organizaciones:

Gestor de base de datos: se conoce generalmente como *motores de bases de datos*, capaces de almacenar datos de manera relacional, de objetos, jerárquica, XML, entre otros.

Sistemas de información: un sistema de información también se considera fuente de información solo cuando suministra sus datos, independientemente de la manera en que lo haga: datos compartidos, integraciones, entre otros. Los sistemas de planificación de recursos empresariales (ERPs, según su sigla en inglés de *enterprise resource planning systems*) también se encuentran en esta categoría.

Sistemas de información externos: esta categoría puntuiza los sistemas de información que comparten o exponen datos de una organización a otra. Por ejemplo, una aerolínea expone los datos de disponibilidad de viajes a agencias de viaje, con el fin de que estas tengan información en tiempo real y puedan ofrecer planes turísticos a sus clientes.

Archivos o aplicaciones estándar: esta categoría se refiere a aplicaciones de escritorio, como Excel, Word o Power Point, entre otras. A pesar de que estas aplicaciones normalmente las utiliza una persona o un conjunto reducido de ellas, también se pueden considerar como fuentes de datos. Esto se evidencia en empresas emergentes (*startups*) u organizaciones con un bajo número de empleados. Incluso, puntos de ventas, que registran toda su operación en hojas de cálculo que después se cargan en algún sistema de información o de base de datos.

Portales de internet: son datos publicados o albergados en sitios de internet (Ríos Sierra, 2014) que son obtenidos a través de API o interfaces de aplicaciones; estas exponen funciones preconstruidas con el fin de suministrar datos específicos, de manera gratuita o comercial; por ejemplo: Google Maps API, Facebook API y Twitter API. También existen otras formas de exponer información; por ejemplo, los rss (*rich site summary*),

es decir, contenido resumido que se ofrece normalmente en blogs o portales de noticias. Por su parte, *web scraping* consiste en consumir de manera automática contenido HTML mediante programas de *software*.

Las fuentes heterogéneas también pueden diferir entre versión del software, marca o tipo de vendedor y modelo, cuyas variables también deben considerarse durante el proceso de identificación de fuentes de datos.

También es crucial identificar el tipo de datos que la fuente de información suministra. Los datos pueden ser de naturaleza estratégica, de planeación y operativa. Los datos estratégicos están asociados a datos agregados, consolidados e indicadores de cumplimiento, que son utilizados por los empleados de mayor rango en las organizaciones para tomar decisiones por parte de presidentes, vicepresidentes y directores. Los datos de planeación aluden a porcentajes de ejecución de procesos, ingreso y gasto, que son utilizados por gerentes y cargos de nivel medio, mientras que los datos operativos incluyen los registros diarios de procesos; por ejemplo, ventas, que son producidos y utilizados por parte de los cargos operativos de una organización.

Dimensiones de las fuentes de datos

La dimensión de las fuentes de datos permite caracterizar cada una de ellas, a fin de establecer los parámetros de extracción, transformación y cargue de datos al repositorio definido para la siguiente fase, de exploración de datos. Las dimensiones se presentan a continuación:

Origen: consiste en definir si la fuente de datos se encuentra dentro de la organización o, si, por el contrario, es externa a ella. De esta manera, pueden identificarse atributos de calidad de manera anticipada, lo que ayuda a prevenir futuros problemas, como los relativos a seguridad, disponibilidad, integridad y confidencialidad, entre otros.

Tiempo: mediante esta actividad se establece el periodo en el que los datos deben haberse creado o actualizado. Dos categorías sirven para definir el periodo: datos actuales o datos históricos. Los actuales engloban los datos del día o semana en curso, mientras que los históricos se refieren a meses o años pasados. En términos de tiempo de procesamiento,

los datos actuales tardan menos en ser visualizados, en comparación con los datos históricos, debido al tamaño de estos.

Utilización: esta actividad consiste en identificar si los datos a procesar son datos o metadatos. Un metadato es la descripción de un dato, mientras un dato es el valor de una variable. Por ejemplo, del conjunto de datos que contiene las ventas efectuadas por mes en una organización, los metadatos son los meses del año y los datos, el total de ventas.

Tipo: esta actividad se refiere a la definición de los tipos de datos que podrían manipularse y guardarse en un repositorio determinado. Por ejemplo, los datos pueden ser de tipo numérico, alfanumérico, de tiempo, de carácter gráfico, de audio, de video, booleanos o binarios, entre otros. El tipo de datos depende del repositorio donde se vayan a almacenar de manera temporal o permanente.

De todos modos, existe una tipología de datos que el ordenamiento jurídico colombiano ha reconocido, así como una clasificación de la información para determinar si esta se encuentra sujeta a reserva o si, por el contrario, puede ser revelada (Becerra, Cotino Hueso, García Vargas, Sánchez Acevedo y Torres Ávila, 2015). Por ejemplo, 1) cuando se trate de un *dato personal sensible*, relacionado, por ejemplo, con la orientación sexual, los hábitos del individuo y su credo religioso o político, en principio, solo su titular podría tener acceso a él; 2) el dato *personal privado o reservado* contenido en documentos públicos solo puede ser revelado a través de una autoridad administrativa o judicial; 3) en el caso del *dato semiprivado*, que no tiene naturaleza íntima, reservada ni pública, su conocimiento o divulgación puede interesar no solo a su titular, sino a cierto sector o grupo de personas y 4) los *documentos públicos que contengan información personal pública* son objeto de libre acceso (Corte Constitucional, Sentencia C-274 de 2013).

Lo anterior obedece a la seguridad jurídica y el reconocimiento de los demás derechos fundamentales conexos: habeas data, intimidad, honra y buen nombre, información y libertad informática en el ámbito de las nuevas tecnologías de la información y las comunicaciones y las exigencias internacionales (Galvis, 2012).

Tipo de codificación: mediante esta actividad se busca definir el tipo de codificación de caracteres. Los más conocidos son ASCII, EBCDIC y Unicode.

Orientación: es la manera mediante la cual se organizan los datos en archivos, sistemas de información, portales o gestores de bases de datos. La orientación puede ser de izquierda a derecha, de derecha a izquierda, de arriba hacia abajo y viceversa. De esta manera, los datos pueden obtenerse de una misma forma.

Confidencialidad: esta actividad depende de las políticas definidas por parte de la organización, país o convenio internacional. Por ejemplo, la política de protección de datos de la Unión Europea establece que los datos confidenciales de las personas naturales son el número de teléfono, el correo electrónico, la dirección y cualquier otro dato que permita la identificación y el contacto con una persona específica.

En el caso colombiano, la política de protección de datos personales se encuentra consagrada en la Ley estatutaria 1581 del 2012, que se rige bajo el principio de confidencialidad previsto en el artículo 4 de la referida ley (Becerra *et al.*, 2015).

Calidad de las fuentes de datos

El proceso de recopilación de datos incluye la definición de un marco de calidad de datos tanto en la fuente de datos como durante su procesamiento y almacenamiento (Moyano *et al.*, 2017). Las métricas que integran el marco de calidad son: consistencia, correctitud, completitud, exactitud, confiabilidad y relevancia. A continuación, se presentan las preguntas asociadas para la definición del modo de evaluación de las métricas:

Consistencia: es la medida o cantidad de contradicciones que surgen en datos o metadatos. Por ejemplo, un conjunto de datos con los meses del año, y los valores contienen los días de la semana, esto conduce a inconsistencias en los valores de los datos.

Correctitud: es la evaluación de los datos y metadatos en cuanto a la correspondencia entre los valores almacenados y la realidad. Por ejemplo, el valor de ventas almacenado en el repositorio de datos es de \$10, pero en realidad son \$50, lo que significa que existe una carencia en la correctitud de los datos de ventas.

Completitud: este atributo se encarga de evaluar si existen valores faltantes en los datos. Es el caso, por ejemplo, de consultar los valores consolidados de ventas por mes en una

organización, pero no se encuentran los datos de un mes específico. Podría afirmarse entonces que los datos carecen de completitud, en alguna medida.

Exactitud: se encarga de medir si los datos recopilados son identificables y exactos; por ejemplo, el valor de ventas almacenado es de \$10, pero realmente fueron \$10.1. En este caso particular, los valores carecen de exactitud. Otra situación que ayuda a reflejar el grado de exactitud consiste en diferenciar valores homónimos.

Confiabilidad: se encarga de evaluar el proceso estándar para obtener los datos. Un caso puntual para la evaluación de este atributo es la manera mediante la cual los datos son protegidos contra su modificación durante el proceso de consulta, ya que podrían un agente externo podría alterarlos. Por ejemplo, un desafío nacional es el de ponerse a tono con los postulados de la Unión Europea, con el fin de actualizar la normatividad, garantizar mayor eficacia del derecho a la protección de datos y despertar mayores niveles de confiabilidad en el desarrollo del ecosistema digital, la revolución tecnológica y la globalización (Galvis, 2018).

Relevancia: este atributo se encarga de medir si el dato o conjunto de datos contribuyen al propósito de la necesidad que se requiere resolver. Una situación común en las organizaciones es la tendencia a recopilar todos los datos que forman parte de una transacción en una venta para la toma de decisiones; no obstante, datos como el número de identificación tributario y la dirección del establecimiento no ayudan a la toma de decisiones durante el proceso de consolidación y análisis de información.

Además de identificar los atributos de calidad, es primordial asociar una fórmula o métrica para la evaluación del grado de cada atributo.

Exploración de datos

La exploración de datos engloba el análisis de datos a través de tecnologías y herramientas de software con sólidos y avanzados componentes gráficos, evitando conocimientos previos en programación (SQL, NO-SQL, entre otros). De esta manera, los analistas de datos tienen como objetivo combinar, comparar, clasificar y segmentar datos que

permitan descubrir patrones, tendencias, anomalías y comportamientos en ellos. Incluso, durante esta etapa normalmente se definen hipótesis y formulan problemas que deben ser abordados durante la subsiguiente etapa, modelamiento de datos (Kayser *et al.*, 2018).

Al respecto, en el contexto nacional se cuenta con el documento Conpes 3920, Política Nacional de Explotación de Datos (Big Data), cuyo objeto es aumentar el aprovechamiento de datos en Colombia mediante el desarrollo de las condiciones para que sean gestionados como activos y, así, generar valor social y económico. Para lograrlo, busca masificar la disponibilidad de datos de las entidades públicas que sean digitales, accesibles, usables y de calidad, además de favorecer seguridad jurídica en la explotación de datos y disponer de capital humano para otorgar valor con los datos y crear cultura de datos en el país (Niño González, 2016).

Adicionalmente, para consolidar este potencial, el documento propone la definición de un marco jurídico e institucional que maximice la obtención de beneficios del aprovechamiento de datos, permita la cooperación entre los sectores público y privado y, al mismo tiempo, refuerce la protección de los derechos de los ciudadanos en el contexto de la creciente transformación de la vida diaria en datos digitales cuantificables y procesables (llamada datificación). Para ello, se resalta la necesidad de establecer distinciones claras entre las categorías de datos: personales, del sector privado y públicos. Asimismo, se precisan límites que orienten a los actores hacia fines que aumenten el bienestar social (Consejo Nacional de Política Económica y Social, 2018).

Modelamiento de datos

Consiste en el diseño, construcción, refinamiento y validación de un modelo de datos. Este consiste en la selección de un modelo estadístico, probabilístico o estocástico que permite procesar los datos con el fin de encontrar una solución a la hipótesis o problema formulado en la primera fase. En el siguiente apartado se presenta una introducción a los diferentes modelos de datos que pueden definirse en ciencias de datos, usando el aprendizaje de máquina y los algoritmos más reconocidos en el momento de definir e implementar modelos de datos (Cheng *et al.*, 2016; Larson y Chang, 2016).

Visualización de datos

La visualización de datos debe basarse en el concepto de *visualización de datos con excelencia*, que consiste en mostrar los datos con claridad y calidad, entendiendo calidad como un conjunto de variables entre las cuales están precisión, completitud, exactitud y relevancia, entre otras. Adicionalmente, durante este proceso, el público debe recibir el mayor número de ideas en el menor tiempo posible a través de pocos gráficos e información.

Según Edward Tufte, autor de varios libros e investigaciones relacionadas con la visualización de datos, deben considerarse los siguientes principios: enfocado en el contenido, comparación en vez de meras descripciones, integridad, alta resolución y la utilización de diseños y conceptos clásicos (Globus, 1994).

Enfocado en contenido

Los datos presentados deben ser lo más importante, más allá de la técnica de representación de datos, el diseño multimedia o herramienta de software. La presentación del contenido debe ser clara, simple y concisa, con alta solidez en los datos; es decir, debe mostrar información que se pueda convertir en conocimiento nuevo y sustancial para la toma de decisiones informadas en las organizaciones, manteniendo siempre la calidad implícita en los datos.

Comparación en vez de descripción

Los datos que se presentan deben facilitar la identificación de sus causas, efectos y comportamiento mediante la comparación entre ellos. La mayoría de las visualizaciones, tanto en la academia como en la industria, se concentran en la descripción de los datos, enfocándose únicamente en la lectura de los asociados a los ejes x y y , que conducen a la identificación de las causas de los comportamientos. En conclusión, deben compararse los datos con variables como las de períodos anteriores; por ejemplo, los ingresos obtenidos durante los últimos tres años, en comparación con los ingresos del año actual.

Integridad

La representación de los datos debe estar acompañada de las etiquetas de los gráficos y sus títulos; de esta manera, se proporciona al público la mayor cantidad de información con claridad y se evitan ambigüedades en medidas, pesos y tipo de moneda, entre otros. Se garantiza así una lectura autosuficiente, sin tener que esperar una explicación detallada de lo que se visualice.

Alta resolución

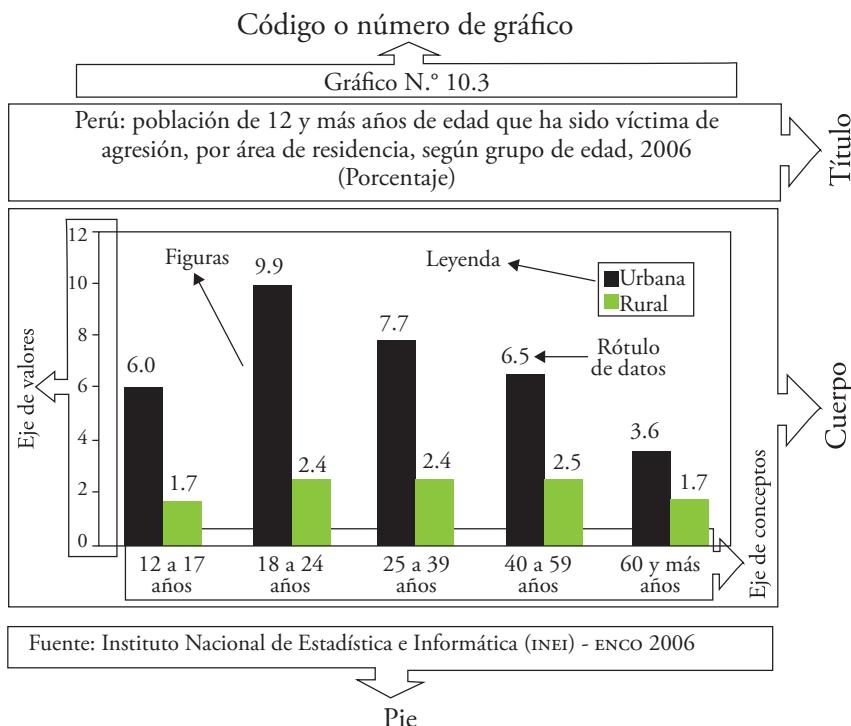
Los datos deben mostrarse por paquetes o secciones que presenten pequeñas porciones de datos, en vez de suministrar grandes cantidades de ellos en una sola pantalla. Esto facilita su lectura, pues se entienden de modo más fácil, debido a que se encuentran clasificados.

Diseños probados

Consiste en la selección y utilización de gráficos adecuados para visualizar los datos. A continuación, se presenta la estructura fundamental de un gráfico y sus variaciones:

Código o número de gráfico

Elemento numérico que permite identificar al gráfico estadístico (aplica para documentos como tesis, guías y otros).

Figura 1. Estructura fundamental del gráfico

Título

El título expresa el contenido del gráfico en forma ordenada, clara y breve. Debe evitarse la descripción excesiva o la brevedad extrema. Se debe indicar, además, la fecha o período de referencia de la información.

Otras características

La primera letra debe escribirse en mayúscula, así como poner las tildes correspondientes, incluso si todo el título está en mayúscula, por supuesto. La fecha de referencia debe ser lo

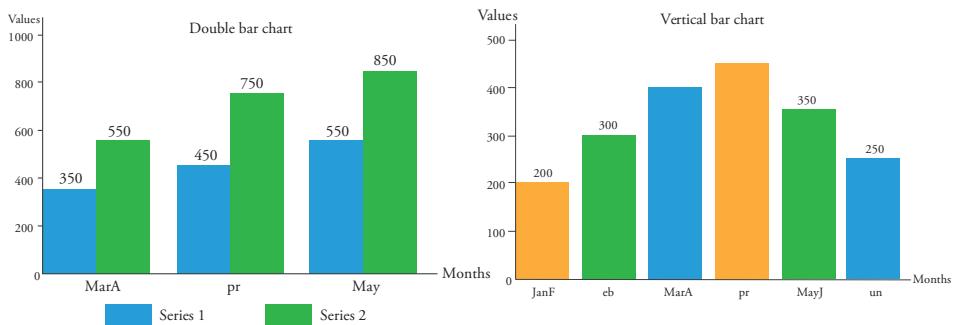
más exacta posible; por ejemplo: 2020-08-31 15:03. Debe especificarse la unidad de medida o el tipo de moneda; por ejemplo, COP (“peso de Colombia”). Las notas aclaratorias del gráfico deben ir debajo de este. Palabras sugeridas para series de tiempo: tendencia, comportamiento, crecimiento, cambio, transformación.

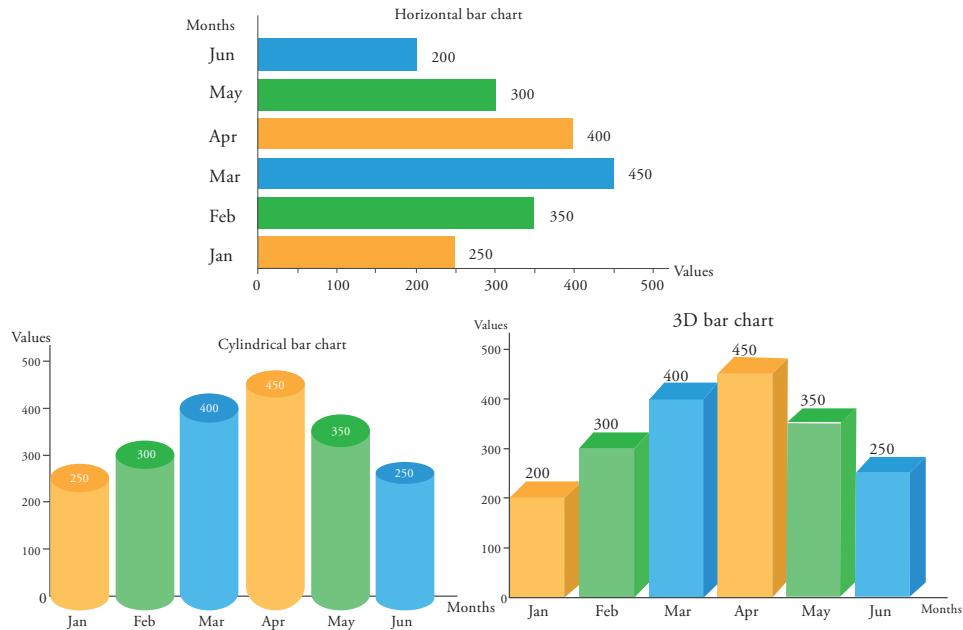
Cuerpo del gráfico

Lo conforma el conjunto de elementos visuales que se utilizan para representar los datos. Entre los más importantes están los de barras, lineal, circular, mapas, diagrama de dispersión y Gantt, entre otros.

Gráficos de barras: se utilizan para comparar datos entre categorías. Son el resultado de combinar dos variables a través de los ejes *x* y *y*. Estos gráficos se crean utilizando modelos en dos y tres dimensiones, según el objetivo del gráfico. Adicionalmente, este tipo de gráficos pueden presentarse de manera vertical u horizontal. Se utiliza la forma horizontal para situaciones en las que se tiene una gran cantidad de categorías que se requieren comparar, organizadas de mayor a menor con diferente color para facilitar la lectura del gráfico.

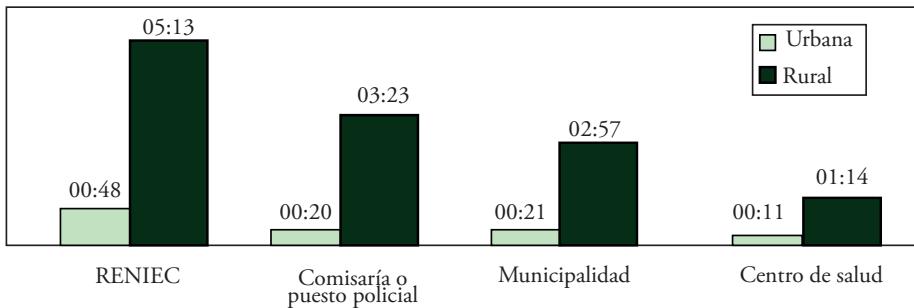
Figura 2. Tipos de gráficos de barras





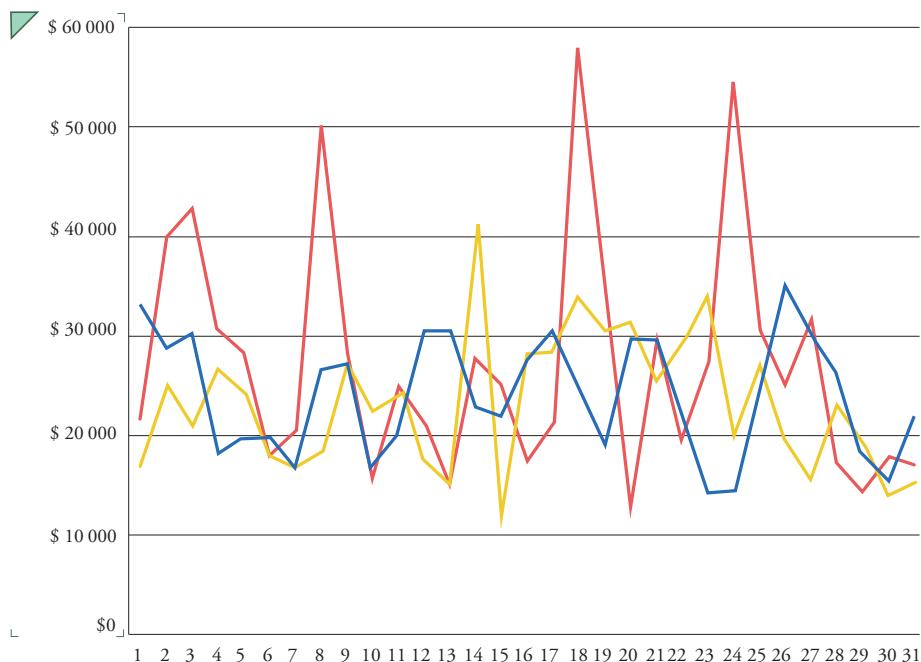
Barras comparativas verticales (dos o más series): se utilizan para comparar la magnitud de dos o más variables mediante barras que pueden ponerse juntas. Estas son una evolución de las gráficas de barras.

Figura 3. Ejemplo de barras comparativas



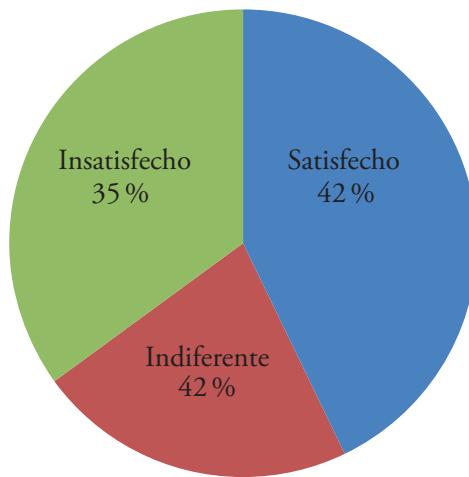
Gráficos de líneas: conectan puntos de datos individuales en una vista. Proporcionan una forma sencilla de visualizar una secuencia de valores. Son útiles para visualizar tendencias en el tiempo y permiten pronosticar valores futuros mediante la comparación de datos históricos.

Figura 4. Gráfico lineal que representa las ventas diarias de tres productos determinados



Gráficos circulares: se utilizan para mostrar proporciones. Comúnmente presenta distribuciones porcentuales y colores distintos para cada porción del círculo. Aunque este tipo de gráficos se pueden representar en dos y tres dimensiones, la mayor cantidad de reportes se basa en dos, debido a que es más fácil y rápido de leer.

Figura 5. Ejemplo de gráfico circular: representa los resultados de una encuesta



Gráficos de dispersión: se utilizan para visualizar las relaciones entre variables numéricas. La condición para utilizar este tipo de gráficos consiste en definir como máximo dos variables. En caso de que se necesite buscar la relación con una tercera variable, debe utilizarse un gráfico de burbujas.

Figura 6. Ejemplo de gráfico de dispersión en dos dimensiones

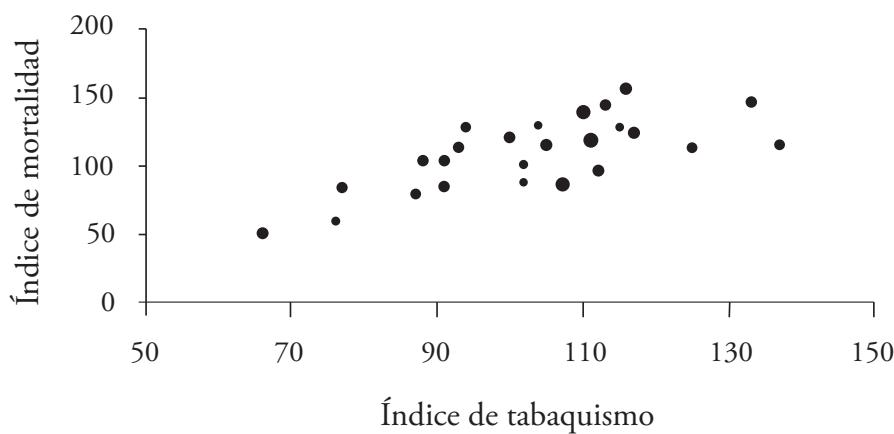
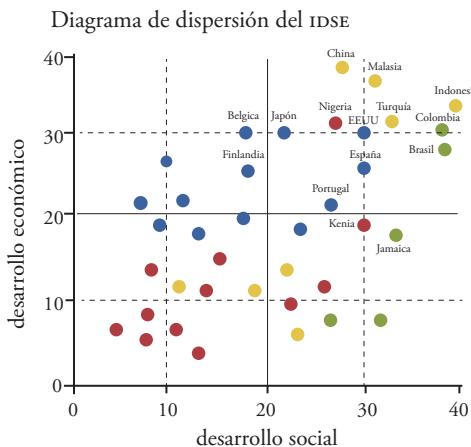


Figura 7. Ejemplo de un gráfico de dispersión en tres dimensiones



Mapa de calor: se emplean para comparar datos categóricos mediante colores. Los mapas de calor no están asociados únicamente a mapas geográficos, sino que también se pueden extraer a la representación del área de cualquier objeto, por ejemplo, o a una cancha de fútbol, en caso de que se necesite conocer el área con mayor presencia de juego.

Figura 8. Ejemplo de mapa de calor: mapa geográfico

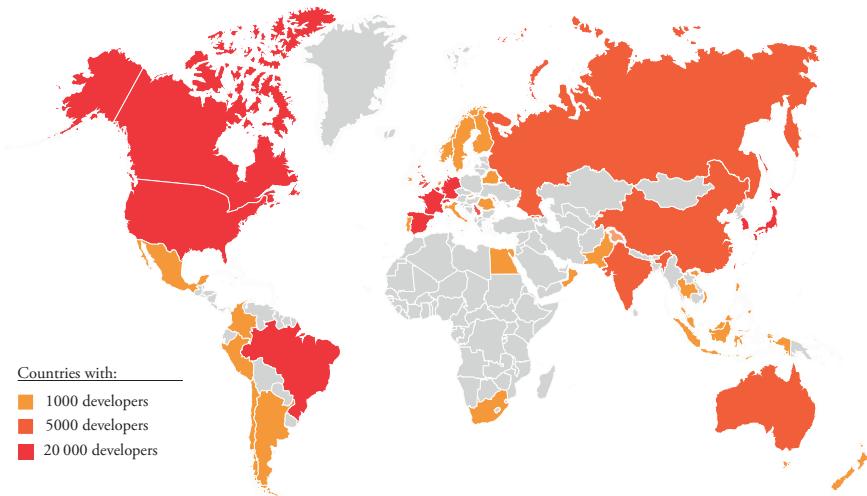
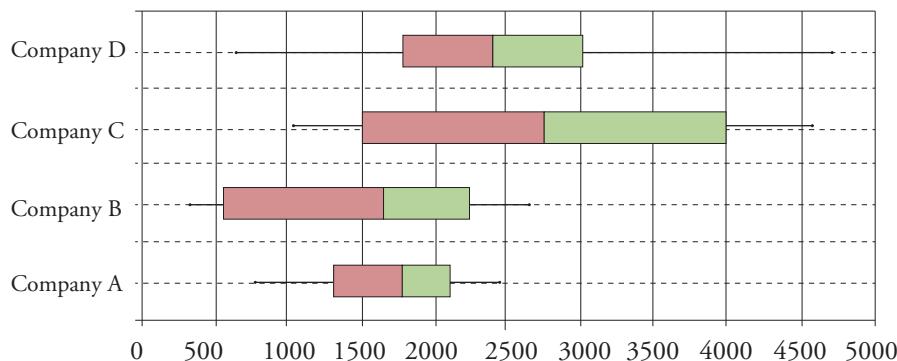


Diagrama de cajas (bigots): son muy útiles para establecer comparaciones. En 1977, John Tukey publicó un tipo de gráfico estadístico para resumir información utilizando cinco medidas estadísticas: el valor mínimo, el primer cuartil, la mediana, el tercer cuartil y el valor máximo. Un gráfico de este tipo consiste en un rectángulo (caja), donde los lados más largos muestran el recorrido intercuartílico (RIC). Este rectángulo está dividido por un segmento vertical que indica dónde se posiciona la mediana y, por lo tanto, su relación con los cuartiles primero y tercero (el segundo cuartil coincide con la mediana). Este rectángulo se ubica a escala sobre un segmento que tiene como extremos los valores mínimo y máximo de la variable. Estos segmentos que quedan a izquierda y a derecha de la caja se llaman bigotes.

Figura 9. Ejemplo de diagrama de cajas



Gráficos de burbujas: se utilizan para mostrar datos en un grupo de círculos. Las dimensiones se usan para definir las burbujas individuales, y las medidas, para definir el tamaño y el color de los círculos individuales. Este gráfico se utiliza para encontrar relaciones entre variables, normalmente más de dos.

Figura 10. Ejemplo de diagrama de burbujas

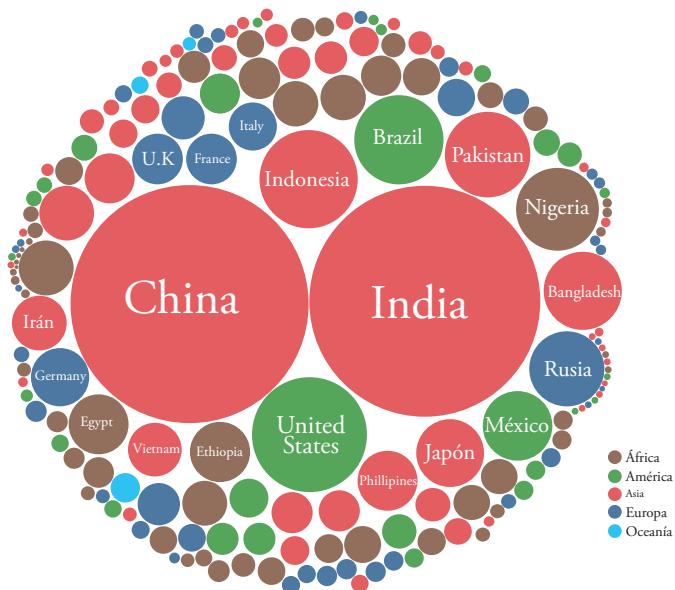
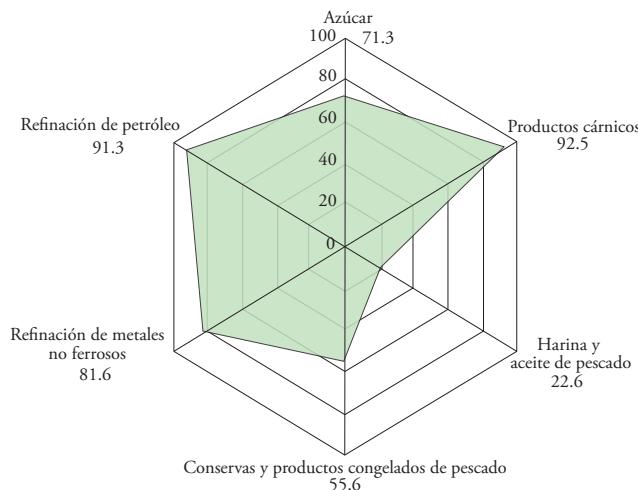
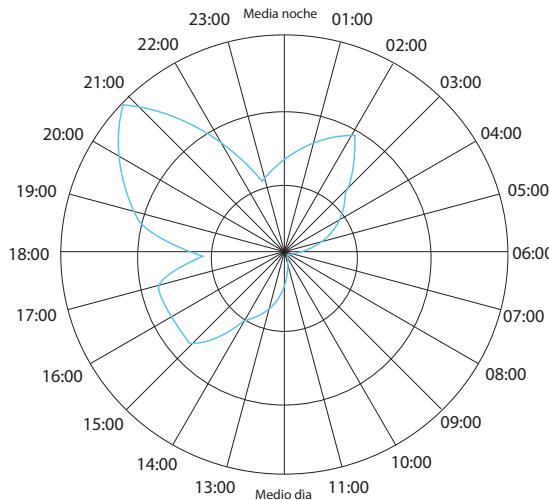


Gráfico radial: compara los valores agregados de cada serie de datos. Cada eje se utiliza para indicar el valor de un indicador específico. Regularmente se representan indicadores distintos con valores previamente estandarizados para fines de comparación.

Figura 11. Ejemplo de gráfico radical: comparación de diferentes categorías

Fuente: INEI Dirección Nacional de Cuentas Nacionales.

Figura 12. Ejemplo de gráfico radical: comparación de esfuerzo en una hora específica del día

Nube de palabras: se encargan de representar un conjunto de datos no estructurados que no tienen asociados valores numéricos. La construcción de este tipo de gráfico se basa en el conteo de palabras que se repiten; la palabra con mayor cantidad de coincidencias va aumentando de tamaño y cambiando de color. El siguiente ejemplo está basado en las palabras encontradas en un foro de discusión de un aula de aprendizaje virtual.

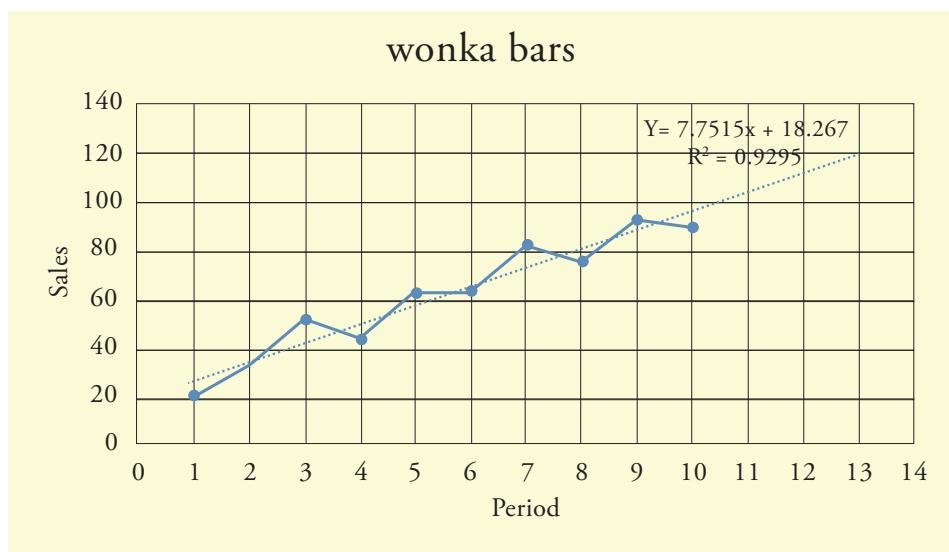
Figura 13. Ejemplo de nube de palabras



Líneas de tendencia: permite predecir los comportamientos de variables a través del tiempo, construidas a partir de un histórico de datos y la variable R^2 asociada, que representa el nivel de confianza de la línea de tendencia. El rango de valores del nivel de confianza está establecido entre 0 y 1. Para un nivel de confianza de 100 %, el valor de R debe ser igual a 1, lo que significa que los datos del gráfico se ajustan a la línea de tendencia propuesta. Adicionalmente, las líneas se pueden clasificar por su grado de libertad, que indica el número mínimo de parámetros que permite especificar el modelo por completo. El grado de libertad para las tendencias lineales, logarítmicas y exponenciales es igual a 2. Las tendencias polinómicas tienen grados de libertad de modelo de 1 más el grado del polinomio.

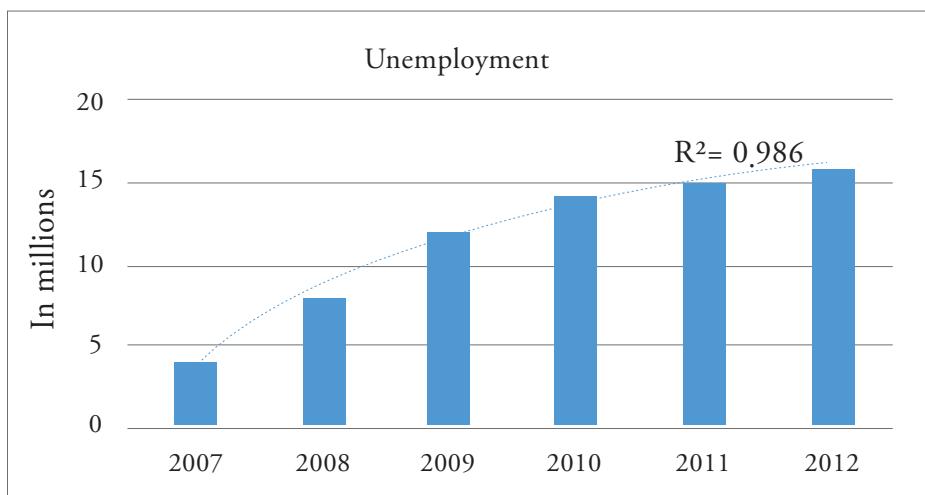
Línea de tendencia lineal: es una línea recta que se ajusta perfectamente y se utiliza con conjuntos de datos lineales simples. Una línea de tendencia lineal normalmente muestra que algo aumenta o disminuye a un ritmo constante. La condición para este tipo de tendencia consiste en que los datos son lineales si la trama de los puntos de datos se parece a una línea.

Figura 14. Ejemplo de línea de tendencia lineal

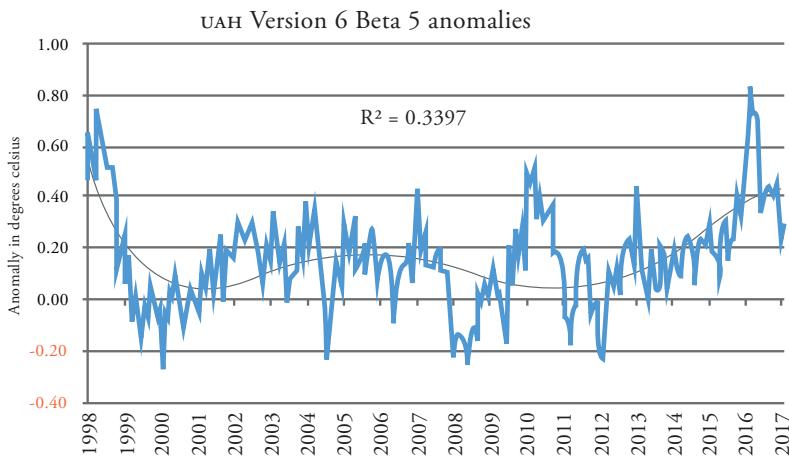


Línea de tendencia logarítmica: es una línea curva que se ajusta perfectamente y que puede utilizar valores positivos o negativos. La condición para obtener este tipo de comportamiento consiste en que el índice de cambios de los datos aumente o disminuya rápidamente y que después se estabilice.

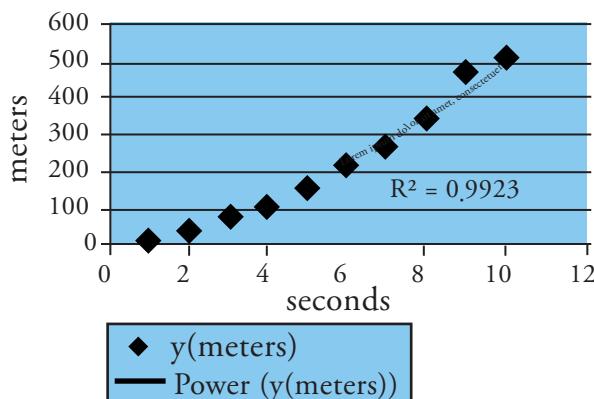
Figura 15. Línea de tendencia logarítmica



Línea de tendencia polinómica: línea curva que se utiliza cuando los datos fluctúan. Es útil para analizar las pérdidas y ganancias de un conjunto de datos grande. El orden del polinomio puede determinarse mediante el número de fluctuaciones en los datos o en función del número de máximos y mínimos que aparecen en la curva. Una línea de tendencia polinómica de orden 2 suele tener solo un máximo o un mínimo. Una de orden 3 normalmente tiene uno o dos máximos o mínimos. Una de orden 4 tiene más de tres. La condición para obtener este comportamiento consiste en que el conjunto de datos sea fluctuante.

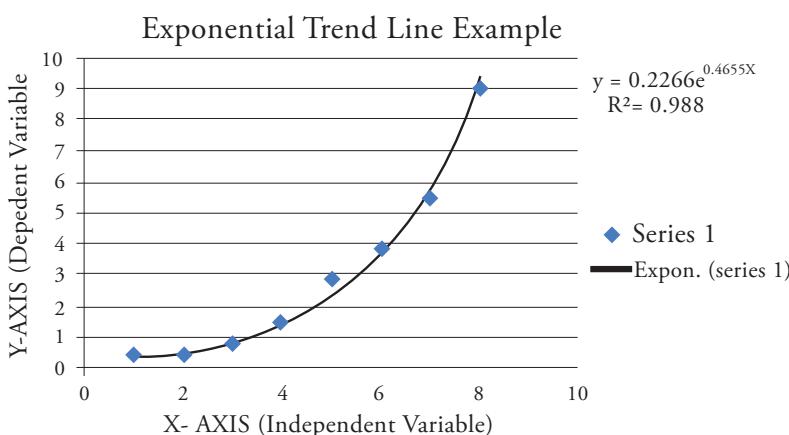
Figura 16. Línea de tendencia polinómica

Línea de tendencia potencial: línea curva utilizada con conjuntos de datos que comparan medidas que aumentan a un ritmo concreto; por ejemplo, la aceleración de un automóvil de carreras a intervalos de un segundo. No es posible crear una línea de tendencia de potencia si los datos contienen valores cero o negativos. La condición para obtener este tipo de tendencia consiste en que los valores aumenten a un ritmo concreto.

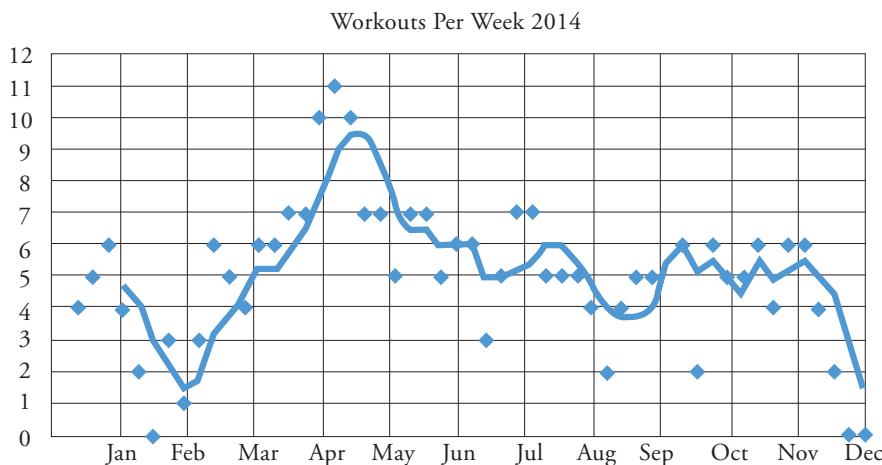
Figura 17. Línea de tendencia potencial

Línea de tendencia exponencial: línea curva que resulta muy útil cuando los valores de los datos aumentan o disminuyen a intervalos cada vez mayores. No es posible crear una línea de tendencia exponencial si los datos contienen valores cero o negativos. La condición para obtener este tipo de tendencia consiste en que los valores de los datos aumenten o disminuyan a intervalos cada vez mayores.

Figura 18. Línea de tendencia exponencial



Línea de tendencia media móvil: esta atenúa las fluctuaciones en los datos para mostrar con mayor claridad la trama o la tendencia. Esta línea de tendencia de media móvil utiliza un número concreto de puntos de datos (establecido por la opción período), hace un promedio de ellos y utiliza el valor del promedio como punto en la línea de tendencia. Por ejemplo, si el valor de período está establecido en 2, el promedio de los dos primeros puntos de datos se utiliza como el primer punto en la línea de tendencia de media móvil.

Figura 19. Línea de tendencia media móvil

Pie de página (*footer*)

Es la última parte del reporte, pero no la menos importante. Siguiendo la estructura fundamental del reporte, el pie de página normalmente se ubica en la parte inferior del reporte e indica la fuente de los datos; por ejemplo, una organización o un sistema de información, entre otros. Adicionalmente, debe indicarse la fecha de la última actualización de los datos; de esta manera, el público obtiene más información respecto de la frecuencia de actualización de los datos.

TRABAJOS FUTUROS

Considerando que la primera fase de esta investigación busca definir una metodología apropiada para abordar un problema de consolidación y representación de datos, de manera tal que se asegure la ejecución ordenada y estructurada de la solución, futuras investigaciones deberían poner en operación la metodología propuesta y su aplicación en problemas como los de paz y reconciliación.

CONCLUSIONES

La metodología propuesta en esta investigación permitirá a las siguientes fases del proyecto abordar de una manera ordenada y estructurada la consolidación y limpieza de datos, con el fin de presentar la información asociada al proceso de paz y posconflicto en Colombia. Como consecuencia, las personas que toman decisiones lo harán de una manera oportuna y eficiente.

Igualmente, el país cuenta con un marco normativo de protección de datos, como la Ley 1581 del 2012, de protección de datos personales, que orienta sobre el manejo efectivo y correcto de los datos personales y que debe acatarse en cuanto al uso, acceso, intercambio, aprovechamiento y recopilación de información con motivo del posconflicto. Se garantizan así los derechos fundamentales de las partes e intervenientes.

Finalmente, también debe reconocerse y garantizarse el derecho al acceso a la información pública conforme a la normatividad vigente, en atención a la libertad informática en el ámbito de las nuevas tecnologías de la información y comunicaciones, dentro del marco del proceso de paz y posconflicto que vive la nación.

REFERENCIAS

- Becerra, J., Cotino Hueso, L., García Vargas, C. B., Sánchez Acevedo, M. E. y Torres Ávila, J. (2015). Responsabilidad del Estado, derechos y TIC. En M. Sánchez (Ed.), *La responsabilidad del Estado por la utilización de las tecnologías de la información y la comunicación (TIC)*. Bogotá: Universidad Católica de Colombia.
- Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C. y Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information Polity*, 19, 5-16. <https://doi.org/10.3233/IP-140328>
- Cheng, S., Liu, B., Ting, T. O., Qin, Q., Shi, Y. y Huang, K. (2016). Survey on data science with population-based algorithms. *Big Data Analytics*, 1(1), 3. <https://doi.org/10.1186/s41044-016-0003-3>

Consejo Nacional de Política Económica y Social. (2018). Documento Conpes 3920. Política Nacional de Explotación de Datos (Big Data). Recuperado de <https://colaboracion.dnp.gov.co/CDT/Conpes/Económicos/3920.pdf>

Galvis, L. (2012). Protección de datos en Colombia, avances y retos. *Revista Lebret*, (4), 195-214. <https://doi.org/https://doi.org/10.15332/fl.v4i4.336>

Galvis, L. (2018). El panóptico digital de la protección de datos personales en Colombia. *Revista Temas*, (12), 125-140. <https://doi.org/https://doi.org/10.15332/rt.v0i12.2038>

García, M. y Fino, G. (2014). Los impuestos territoriales en Colombia y la inequidad social, ¿la voluntad de la clase dominante erigida en ley? *IUSTA*, 2(41), 61-75. <https://doi.org/https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.15332%2Fs1900-0448.2014.0041.04>

Globus, A. (1994). *Principles of Information Display for Visualization Practitioners*.

Gröger, C., Hillmann, M., Hahn, F., Mitschang, B. y Westkämper, E. (2013). The operational process dashboard for manufacturing. *Procedia CIRP*, 7, 205-210. <https://doi.org/10.1016/j.procir.2013.05.035>

Índice de Paz Global. (2017). Recuperado de Datosmacro.com website: <https://datosmacro.expansion.com/demografia/indice-paz-global>

Joshi, P. (2015). Analyzing big data tools and deployment platforms. *International Journal of Multidisciplinary Approach and Studies*, 2(2), 45-56. Recuperado de <http://ijmas.com/upcomingissue/05.02.2015.pdf>

Kayser, V., Nehrke, B. y Zubovic, D. (2018). Data science as an innovation challenge : from data to value proposition. *Technology Innovation Management Review*, 8(3), 16-26. <https://doi.org/http://doi.org/10.22215/timreview/1143>

Khan, S., Liu, X., Shakil, K. a. y Alam, M. (2017). A survey on scholarly data: From big data perspective. *Information Processing & Management*, 53(4), 923-944. <https://doi.org/10.1016/j.ipm.2017.03.006>

Larson, D. y Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management*, 36(5), 700-710. <https://doi.org/http://dx.doi.org/10.1016/j.ijinfomgt.2016.04.013>

- Le Blanc, J. (2016). Cultura de paz y reconciliación. *En Construcción de desarrollo y paz: aprendizajes y recomendaciones desde los territorios*. Bogotá: Redprodepaz. Recuperado de http://biblioteca.clacso.edu.ar/Colombia/cinep/20161102012014/20160707_Construccion.pdf
- Mac Ginty, R. (2013). Indicators+: a proposal for everyday peace indicators. *Evaluation and Program Planning*, 36(1), 56-63. <https://doi.org/https://doi.org/10.1016/j.evalprogplan.2012.07.001>
- Mishra, P. y Azeez, A. (2014). Analyzing big data tools and deployment platforms. *International Journal of Multidisciplinary Approach and Studies*, 01(5), 138-147.
- Niño González, C. (2016). Exportación de conocimiento: Colombia como proveedor de seguridad para la lucha contra el terrorismo y el crimen internacional. *Vía Inveniendi et Iudicandi*, 11(2), 87-104. <https://doi.org/https://doi.org/10.15332/s1909-0528.2016.0002.04>
- Rengifo Lozano, D. (2011). Las instituciones y sus presupuestos de información en el orden social: un enfoque económico. *Vía Inveniendi et Iudicandi*, 6(1), 1-44. Recuperado de <http://www.redalyc.org/articulo.oa?id=560258669009>
- Ríos Sierra, J. (2014). Política, poder e internet: nuevas posibilidades frente a viejos dilemas. *Vía Inveniendi et Iudicandi*, 9(1), 100-131. Recuperado de <https://www.redalyc.org/articulo.oa?id=560258675006>
- Sánchez Acevedo, M. E. (2015). El régimen de responsabilidad de la administración pública colombiana por la publicación de contenidos mediante el uso de las tecnologías de la información y comunicación (TIC). *En El derecho y las tecnologías de la información y la comunicación (TIC)*. Bogotá: Universidad Católica de Colombia.
- Szymańska, E. (2018). Modern data science for analytical chemical data – A comprehensive review. *Analytica Chimica Acta*, 1028. <https://doi.org/10.1016/j.aca.2018.05.038>
- Transparencia por Colombia. (n. d.). Transparencia por Colombia. Recuperado de <https://transparenciacolombia.org.co/>
- Zhou, L., Pan, S., Wang, J. y Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing*, 237, 350-361. <https://doi.org/https://doi.org/10.1016/j.neucom.2017.01.026>

