



Nuevas herramientas de información en campañas de concienciación social. El *Big Data* aplicado a la Sequía y Salud Pública

Ubaldo Cuesta Cambra¹; José Ignacio Niño González²; Luz Martínez Martínez³; Carlos Díez Bielsa⁴

Recibido: 14 de mayo de 2018 / Aceptado: 30 de julio de 2018

Resumen. Este trabajo tiene por objeto proponer a los Medios de Comunicación una solución para crear, de forma sencilla y bajo nivel de inversión, una aplicación de *Big Data* que permita disponer de un instrumento de análisis histórico de las Sequías asociadas al impacto en la Salud Pública, con el fin de generar contenidos rigurosos que incidan sobre el comportamiento de los públicos mediante campañas informativas de concienciación y prevención. Implementada la aplicación, esta se puede reutilizar sin costes adicionales puesto que la estructura ya está operativa, para gestionar cualquier otro contenido informativo que sea relevante para crear una Sociedad sostenible.

Palabras clave: Aplicación; *Big Data*; medios de comunicación; sequía; salud pública.

[en] New information tools in social awareness campaigns. Big Data applied to Drought and Public Health

Abstract. The purpose of this work is to propose to the Media a solution to create, in a simple and low investment way, a Big Data application that allows to have an instrument of historical analysis of the Droughts associated with the impact on Public Health, with the purpose of generating rigorous content capable of influencing the behavior of the public through awareness prevention campaign. Once the application is implemented, it can be reused without additional costs since the structure is already working, to manage any other information content that is relevant to create a sustainable Society.

Keywords: Application; Big Data; Media; Drought; Public Health.

Sumario. 1. Introducción. 2. Descripción de la aplicación. 3. Perfiles requeridos. 4. Implementación de la aplicación; 4.1. Definición de requisitos de datos; 4.2. Adquirir, recopilar y almacenar datos; 4.2.1. AEMET (Agencia Estatal de Meteorología); 4.2.2. MAPAMA (Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente); 4.2.3. MSSSI (Ministerio de Sanidad Servicios Sociales e Igualdad); 4.2.4. CALIOPE (Sistema Pronóstico de Calidad de Aire); 4.2.5. Sistemas de datos

¹ Universidad Complutense de Madrid (España)
E-mail: ucuestac@ucm.es

² Universidad Complutense de Madrid (España)
E-mail: josenino@ucm.es

³ Universidad Rey Juan Carlos (España)
E-mail: luz.martinez@urjc.es

⁴ Profesional experto en Transformación Digital (España)
E-mail: cdiezbie@gmail.com

abiertos del gobierno de España; 4.2.6. Redes sociales; 4.3. Procesar y analizar datos; 4.4. Consumir / visualizar. 5. Herramientas de implementación; 5.1. Adquirir y recopilar datos; 5.2. Almacenar datos; 5.3. Procesar y analizar; 5.4. Visualizar. 6. Conclusiones. 7. Referencias bibliográficas.

Cómo citar: Cuesta Cambra, Ubaldo; Niño González, José Ignacio; Martínez Martínez, Luz; y Díaz Bielsa, Carlos (2019): "Nuevas herramientas de información en campañas de concienciación social. El *Big Data* aplicado a la Sequía y Salud Pública". *Estudios sobre el Mensaje Periodístico* 25 (2), 767-783.

1. Introducción

El cada vez más intenso uso de tecnologías en todos los ámbitos de la sociedad está dando lugar a un proceso de cambio bautizado como "Transformación Digital" que se caracteriza por su capacidad para facilitar nuevos recursos que permiten obtener mayores niveles de innovación y creatividad a múltiples áreas de actividad: el mundo empresarial, organismos públicos, educación, industria, ciencia, medios de comunicación, medio ambiente, salud, etc. Esta transformación conlleva asociada una necesidad de cambio de mentalidad por parte de todos los segmentos que intervienen en la misma, que están obligados a empatizar con el cambio y contar con capacidad para asumirlo.

Dentro de este nuevo marco tecnológico, una de las áreas que tiene mayor significación es la recolección de datos inteligentes que puedan ser utilizados de forma eficaz y de cuyo análisis se puedan extraer líneas de actuación que permitan mejorar la calidad de vida de los seres humanos. La capacidad para generar cada vez más volumen de datos y darles a estos una interpretación ha llevado a diferentes autores a hablar de un hito en la historia de la humanidad que (Hellerstein, 2008) definida como "la revolución industrial de los datos". Pero el valor de los datos no está en sí mismos sino en lo que se hace con ellos, por lo tanto, es básico delimitar su intencionalidad y utilidad (Murphy y Barton, 2014). Hoy en día los procesos de *Big Data* permiten el tratamiento conjunto de grandes volúmenes de datos, y esta capacidad de gestión es la puerta para identificar patrones de comportamiento que faciliten la predicción de situaciones complejas y así poder implementar acciones que permitan modificar las conductas. El análisis de datos debe servir para proveer a los ciudadanos de información que no solo sea capaz de fundamentar sus elecciones, opiniones o decisiones, sino debe ser un instrumento educativo que genere conductas responsables y permita reducir los efectos negativos de situaciones coyunturales que ponen en peligro la convivencia y el bienestar social. Una Sociedad bien informada es una Sociedad preparada para comportarse de un modo racional ante situaciones de riesgo como puede ser la falta de agua, y desde el conocimiento anticipado de lo que sucede, poner en práctica un modelo de consumo responsable y sostenible.

El objeto de este trabajo es proponer a los Medios de Comunicación una solución para crear de forma sencilla, basada en software de bajo coste, una aplicación de *Big Data* que les permita disponer de un instrumento de análisis histórico de las sequías y su predicción, asociadas al impacto en la Salud Pública. Las empresas informativas como emisores de contenidos capaces de incidir sobre los comportamientos de los públicos deben disponer de recursos que estén enfocados a crear una Sociedad sostenible. Identificar con antelación periodos de

sequía y promover usos controlados de los recursos hidrológicos puede ser el primer paso para evitar problemas futuros, máxime cuando en España las actuaciones preventivas activadas desde los organismos públicos se producen en el corto plazo y cuando el problema ya es de una dimensión muy compleja. Los modelos climáticos existentes en España indican un serio empeoramiento de las condiciones térmicas y sobre todo pluviométricas para las próximas décadas por lo que hay que promover de forma urgente sistemas de prevención que puedan paliar las graves consecuencias de una falta regular de agua. Además, la incidencia de la sequía sobre la salud es una realidad que hay que tener muy presente, como puede ser el efecto de la pésima calidad del aire en las grandes ciudades.

En esta misión de concienciación social juegan un papel determinante los Medios de Comunicación, ya que sobre ellos recaería la fundamental labor de transmitir a la ciudadanía las informaciones que se pueden derivar del proceso de *Big Data* y así promover comportamientos preventivos. Las tecnologías de la comunicación y la información, y los cambios en los usos y consumos informativos de los ciudadanos (Becker y Schönbach, 1989; Napoli, 2003) sitúan al periodismo y a las empresas informativas en un momento decisivo de su historia (Picard, 2010). Es importante resaltar que aunque la aplicación propuesta está orientada a la generación de contenidos informativos relacionados con la Sequía y la Salud Pública, es una solución que se puede reutilizar tantas veces como se considere oportuno para desarrollar cualquier otro tipo de contenido relevante para los ciudadanos.

Las empresas de comunicación del siglo XXI no pueden dejar pasar la oportunidad que el *Big Data* pone en sus manos para convertirse en abanderados del cambio social y liderar la emisión de campañas informativas que promuevan conductas sociales responsables.

2. Descripción de la aplicación

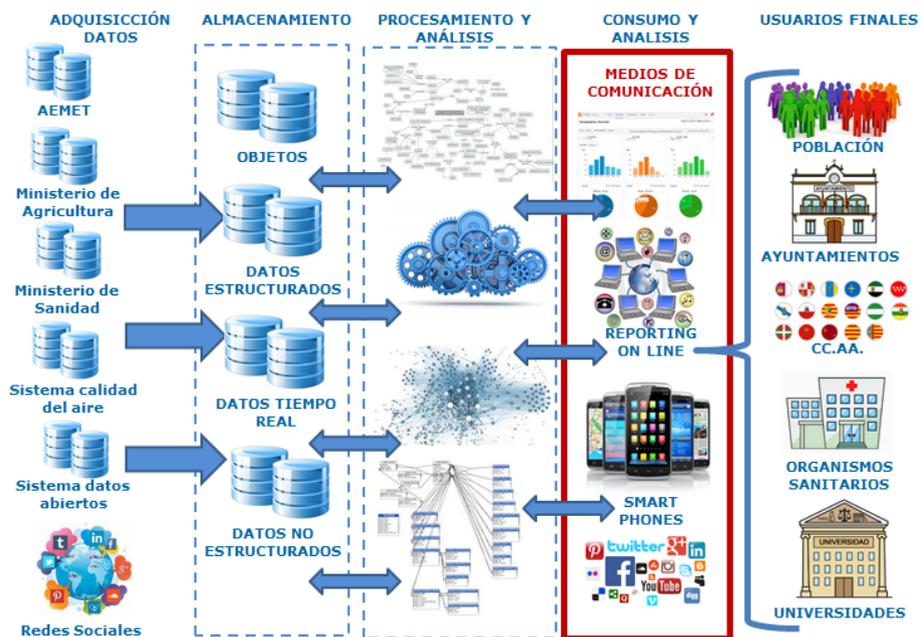
La solución *Big Data* que proponemos permitirá identificar patrones de comportamiento de la Sequía a partir de datos obtenidos de diferentes fuentes de información y correlacionarlos con la incidencia en la Salud Pública, de tal modo que facilitará la predicción de situaciones de riesgo futuras a partir del análisis de los datos históricos y conseguir variables que permitan paliar las situaciones en el largo plazo. Con esta aplicación los Medios de Comunicación podrán generar contenidos informativos de alto valor y dirigir con antelación campañas educativas preventivas a ciudadanos, comunidades autónomas, ayuntamientos, centros de salud, hospitales, universidades, etc... Además, la publicación podrán realizarla no solo a través de sus estructuras comunicativas clásicas, sino también empleando cualquier otro canal que se considere adecuado (aplicaciones móviles, *RRSS*, mensajería instantánea, etc.)

La aplicación *Big Data* propuesta, contará con la capacidad óptima para “recolectar” grandes volúmenes de información de diversas fuentes, con una gran variedad de datos a los que se les aplicará la inteligencia precisa mediante el uso de algoritmos, de tal manera que los datos tratados puedan ser explotados de una forma eficaz, permitiendo analizar el comportamiento pasado, presente y futuro

mediante el análisis de patrones de comportamiento. En nuestro caso, la solución nos facilitará la identificación de los parámetros precisos para determinar futuras sequías y su incidencia en la Salud Pública. La información resultante, se presentará de forma sencilla y de fácil comprensión para los Medios de Comunicación, que elaborarán desde su perspectiva informativa los resultados y los trasladarán a los destinatarios finales con el objetivo de promover la toma de decisiones responsables.

En la arquitectura de referencia de la aplicación, se han aplicado las recomendaciones de *Amazon Web Services*⁵. La representación funcional de la fase de implementación es la que se muestra en la siguiente imagen. (Figura 1)

Figura 1. Funcionalidad de la aplicación. Imagen de elaboración propia.



Las áreas de implementación son las siguientes:

1) Adquisición y Recopilación de Datos: Corresponde a la fase de identificación de los sistemas orígenes de datos, de dónde se identificarán y procesarán en las fases posteriores.

2) Almacenamiento: Fase en la que se registran los datos extraídos y se almacenan en las bases de datos más adecuadas a partir del tipo de dato a tratar: documentos, gráficos, datos transaccionales, datos de redes sociales, datos generados por dispositivos, etc.

⁵ <https://es.slideshare.net/AmazonWebServicesLATAM/patrones-de-arquitectura-para-big-data-en-aws>

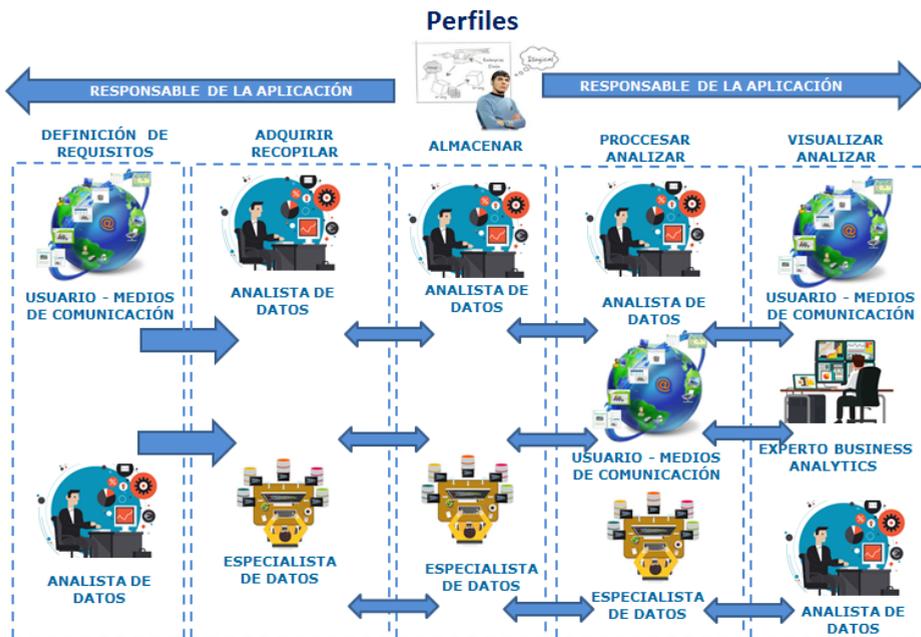
3) Proceso y Análisis: Es la fase en la que se aporta la inteligencia a los datos mediante la aplicación de diferentes metodologías de análisis: Descriptivo, Predictivo y Prescriptivo.

4) Visualización y Análisis: Corresponde a la fase de explotación, visualización y análisis de los datos, una vez aplicada la inteligencia precisa a los mismos en la fase anterior.

3. Perfiles requeridos

A partir de la publicación *Big Data COE* (<https://www.bigdatabcn.com/es/los-7-perfiles-clave-los-profesionales-big-data/>) dónde se definen perfiles idóneos para la implementación de un *Big Data*, se han seleccionado los recursos humanos que necesita un Medio de Comunicación para desarrollar la aplicación objeto de este trabajo. Es importante señalar que con objeto de reducir los costes de puesta en marcha, ciertos roles los puede asumir un mismo perfil, tal como se indica a continuación: (Figura 2)

Figura 2. Perfiles. Imagen de elaboración propia.



- Responsable de la aplicación: Máximo responsable de asegurar y garantizar que las decisiones están respaldadas por datos verificables. Lidera la gestión de datos y analítica asociada a los objetivos a conseguir y, por tanto, es responsable del equipo. En el modelo que estamos planteando, al tratarse de un proyecto inicial con una envergadura limitada, este rol puede ser asumido por el Analista de Datos.

- **Usuario de datos:** En nuestro caso este rol lo desempeñarán un periodista ya que corresponde a un perfil de alguien que no está formado en la ciencia del dato pero que puede extraer valor del mismo a través de su experiencia en los Medios de Comunicación. Es el responsable de generar los contenidos informativos que se transmitirán a la Sociedad.

- **Analista de datos:** Corresponde al rol clave de la ciencia de datos. A partir de todas las necesidades definidas por el Usuario de Datos, su labor principal consiste en extraer conocimiento e información valiosa de los datos. Es conveniente que cuente con habilidades técnicas, pero también debe ser un experto en comunicación que pueda presentar los resultados obtenidos de un modo fácilmente convertible en contenido informativo.

- **Especialista en Datos:** Se encarga de extraer los datos de una manera accesible y apropiada para que el Analista de Datos los interprete. Es un perfil que debe conocer y estar familiarizado con las herramientas de *Big Data*. En una primera fase este perfil puede ser asumido por el Analista de Datos.

- **Experto en *Business Analytics*:** Corresponde al perfil que genera los gráficos, infografías y otros contenidos visuales que ayuden al Usuario de Datos a confeccionar contenidos que expliquen de un modo sencillo datos complejos.

En definitiva, la puesta en marcha inicial de este proyecto requiere la participación de 3 personas: un Analista de Datos que asume 3 roles diferenciados (responsable de la aplicación, extraer datos y analizar datos), un Experto en *Business Analytics* que genera los elementos visuales que ilustran las informaciones y un Usuario de Datos que corresponde al encargado de confeccionar el contenido informativo que finalmente llegará a la opinión pública.

4. Implementación de la aplicación

Para implementar una aplicación de *Big Data* se debe emplear una metodología cíclica, puesto que según va evolucionando la solución se va adquiriendo mayor conocimiento, lo que permite identificar nuevas necesidades de depuración y optimizar los datos a procesar y analizar. (Figura 3)

Figura 3. Ciclo de vida de construcción. Imagen de elaboración propia.



A lo largo del ciclo de vida de implementación debe existir un rol responsable que asegure y garantice que las decisiones están respaldadas por datos verificables, liderando la gestión de datos y analítica asociada a los objetivos a conseguir. Como hemos comentado en la descripción de los perfiles, este rol puede ser asumido por el Analista de Datos. El desarrollo de las tareas en cada fase, son las que se exponen a continuación:

4.1. Definición de requisitos de datos

Esta primera fase consiste en la definición de los requerimientos conceptuales de los datos que se requieren para transformarlos en datos inteligentes que permitan analizarlos y correlacionarlos con el objeto de conseguir la capacidad de análisis deseada. Como metodología de trabajo, se propone una serie de reuniones, dónde el Analista de Datos plasme todas las necesidades conceptuales requeridas por el Usuario que explotará la solución. Los perfiles que deben intervenir en esta fase son los siguientes:

- Usuario de la aplicación: Este rol debe contar con las necesidades conceptuales y funcionales objeto de la solución. Debe tener una idea formada de los conceptos de información a tratar, modo de presentación y explotación de los datos.

- Analista de datos: Debe recopilar y modelar los conceptos y definiciones que le traslada el Usuario.

4.2. Adquirir, recopilar y almacenar datos

A partir de la información generada en la fase anterior, es en esta fase dónde se analiza y define la información a recopilar, normalizar y procesar. Se debe realizar un estudio riguroso de los requerimientos para materializarlos de forma efectiva, identificando la información que se requiere y localizando las diferentes fuentes de dónde extraer los datos en cuestión.

Para nuestra propuesta de plataforma, a modo de referencia, se han identificado los siguientes posibles orígenes de datos, aunque evidentemente se debe realizar un estudio detallado de la información que reside en cada origen de datos:

4.2.1. AEMET (Agencia Estatal de Meteorología)

AEMET es una entidad que almacena de forma continua los datos que recibe y genera para preservarlos de una forma estructurada y ordenada que facilite su acceso y utilización, preserva así la memoria de "el tiempo" aportando información meteorológica y climatológica. Cuenta con una funcionalidad denominada *OpenData API* que permite interactuar con los datos: esta interacción se caracteriza por la posibilidad de ser periódica e incluso programada, desde cualquier lenguaje de programación, con interfaces amigables y estándar.

Los datos residentes en este sistema son del siguiente tipo:

- Climatologías diarias
- Climatologías mensuales/anuales
- Valores Normales

- Valores Extremos
- Balance hídrico nacional
- Avance mensual climatológico
- Predicciones

4.2.2. MAPAMA (Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente)

MAPAMA cuenta con un servicio específico que pone a disposición pública información relativa a las *Sequías en España*. Una serie de datos de gran valor es la información relativa al seguimiento histórico de los indicadores de estado de sequía hidrológica de todas las Demarcaciones Hidrográficas intercomunitarias

Los datos almacenados en este sistema son del tipo:

Reservas hidráulicas.

Indicadores de Estado de la Sequía.

4.2.3. MSSSI (Ministerio de Sanidad Servicios Sociales e Igualdad)

MSSSI provee de información referenciada a enfermedades asociadas a los fenómenos meteorológicos y más concretamente información relativa a la sequía, como las enfermedades de origen hídrico, enfermedades infecciosas, respiratorias, lesiones por quemaduras...etc.

La información que se puede extraer de este sistema es de la siguiente tipología:

Evolución de las enfermedades asociadas a la sequía.

Población afectada por enfermedades asociadas a la sequía.

Seguimiento de las campañas de vacunación y medidas mitigadoras.

4.2.4. CALIOPE (Sistema Pronóstico de Calidad de Aire)

El Sistema CALIOPE ofrece información del pronóstico horario de la calidad del aire. La calidad del aire está directamente relacionada con las enfermedades respiratorias, por lo que una correlación de la información que aporta este sistema con la información extraída de los otros sistemas puede aportar un gran valor a los análisis y resultados a conseguir, como por ejemplo:

Datos Meteorológicos: temperatura, precipitación, humedad, presión, nubosidad, dirección y velocidad del viento, pronóstico de emisiones.

Emisiones: óxidos de nitrógeno, monóxido de carbono, dióxido de azufre.

Calidad del aire: parámetros que permiten determinar la calidad del aire.

4.2.5. Sistemas de datos abiertos del gobierno de España

Se trata de un catálogo con datos de toda índole para la reutilización de los mismos, entre los que cabe resaltar para nuestra propuesta datos de Medio Ambiente, Salud, Demografía...etc.

Este repositorio proporciona un sistema de acceso mediante una API que facilita la consulta de la base de datos semántica como:

- Intensidad de la Sequía

- Humedad del Cultivo
- Suministro de Aguas Superficiales
- Niveles de Precipitaciones
- Masa de nieve
- Caudal de los ríos
- Agua almacenada en los embalses.

4.2.6. Redes sociales

Las principales redes sociales (Facebook, Instagram, Twitter, YouTube), ofrecen APIs (*Interfaces* de conexión) para acceder a la información de las publicaciones que se realizan en las mismas. Estas APIs permiten buscar en las plataformas de redes sociales: posts, usuarios, canales y data demográfica.

Dentro de esta fase se deben transformar los datos extraídos de las fuentes en información útil. El proceso de inspeccionar, transformar y modelar los datos es lo que se denomina como análisis de datos. En esta fase y dependiendo los diferentes tipos de datos, estos se van almacenando en las BBDD más idóneas para cada tipología de dato:

- Base de Datos de Objetos: Datos de tipo objeto que se genera desde cualquier ubicación: sitios web y aplicaciones móviles, aplicaciones y datos de sensores o dispositivos IoT (Internet de las Cosas).

- Base de Datos de Tiempo Real: Para registrar y almacenar de manera continua altos volúmenes de información de datos por hora procedentes de diversos orígenes, como secuencias de clics de sitios web, fuentes de redes sociales, logs de TI y eventos.

- Base de Datos Estructurados: Almacena datos en modelos relacionales, es decir, se trata de la típica base de datos.

- Base de Datos de “datos No Sql”: Para almacenar datos no relacionados que permitan un acceso rápido y flexible.

Perfiles que participan en esta fase:

- Analista de Datos: Como especialista en análisis de datos y conocedor de las necesidades que debe cubrir la aplicación, deberá certificar la correcta modelización de los datos para su posterior procesamiento y análisis.

- Especialista en Datos: Es el especialista técnico en la implementación, empleando para ello herramientas para la extracción, recopilación, limpieza, normalización y almacenamiento de datos.

4.3. Procesar y analizar datos

Una vez almacenados los datos en su correspondiente Base de Datos según su tipología y formato, la siguiente fase consiste en el procesamiento y análisis de los datos almacenados.

El análisis de datos es un proceso que se basa en inspeccionar, limpiar, transformar y modelar los datos con el objetivo de descubrir la información de utilidad, sugerir conclusiones y respaldar la toma de decisiones. Para realizar esta tarea, se emplean las siguientes metodologías:

1) **Análisis Descriptivo:** Consiste en almacenar y realizar agregaciones de datos históricos, visualizándolos de forma que puedan ayudar a la comprensión del estado actual y pasado, en nuestro caso corresponde al análisis de la evolución de los parámetros que permiten determinar el índice de sequías y su incidencia en la Salud Pública desde tiempos anteriores a la actualidad. Esta tarea, requiere la experiencia del experto en *datos* ya que los datos a analizar corresponden de fuentes muy variadas y normalmente inconexas.

En este análisis se pueden realizar las siguientes funcionalidades:

- Detectar los índices de Sequía que se han producido por zonas geográficas.
- Visualizar la distribución geográfica de la población que publican/consumen información en las redes sociales relativa a la Sequía y Salud Pública.
- Observar la evolución histórica de la incidencia de la Sequía y la correlación con el uso de fármacos asociados a las enfermedades relacionadas con la Sequía, para diferentes periodos temporales.
- Identificar campañas contra las enfermedades asociadas a la sequía.

2) **Análisis Predictivo:** Proporciona herramientas para estimar aquellos datos que son desconocidos o inciertos, o que requieren de un proceso manual o costoso para su obtención. Más allá del puro análisis de la información histórica que realiza la analítica descriptiva, las predicciones de datos que realiza la analítica predictiva fortalecen las decisiones tales como:

- Anticipar periodos críticos de sequía y su asociación a la Salud Pública, teniendo en cuenta factores controlables, como efectos meteorológicos, calidad del aire, agua embalsada, índice de enfermedades, consumo de medicamentos asociados a las patologías asociadas a la sequía.
- Detectar si los parámetros analizados corresponden a un futuro periodo de sequía o bien a un periodo puntual.

3) **Análisis Prescriptivo:** Permite analizar los datos para encontrar cuál es la solución entre una gama de variantes. En nuestro caso, se pueden realizar técnicas de simulación y optimización, logrando señalar cuál es el camino que conviene realmente elegir ante las diversas variantes:

Identificando acciones a realizar para evitar periodos críticos de sequía y acciones mitigadoras en la sociedad para evitar la alta incidencia de las enfermedades en los periodos de sequía.

Actores que participan en esta fase:

- **Usuario de Datos:** Aportando su valor a través de su experiencia en los Medios de Comunicación y la Salud Pública, será el soporte tanto del Analista de Datos como del Especialista en Datos de modo que definan e implementen los procesos de análisis de los datos.
- **Analista de Datos:** Desarrollando tareas de supervisión, ya que cuenta con el detalle de los requerimientos que debe cubrir el análisis de datos y definiendo los criterios que son necesarios aplicar para aportar a los datos la inteligencia requerida.
- **Especialista en Datos:** Implementando los mecanismos de software definidos por el Analista de Datos.

4.4. Consumir / visualizar

Es en esta fase, a partir de herramientas específicas de visualización, dónde los Medios de Comunicación como usuarios de la aplicación, realizarán el análisis sobre la información procesada y almacenada en la fase anterior de construcción.

El consumo y visualización de datos, es una técnica que facilita al análisis de los mismos de manera gráfica, pero no debe ser un sustituto de la técnica de analítica de datos y procesamiento, actividades desarrolladas en la etapa de Procesamiento y Análisis en la construcción del *Big Data*. Esta técnica facilita notablemente la identificación de diferencias y correlaciones entre los datos objetos del análisis. En el caso de uso objeto de este artículo, es en esta etapa en la que los Medios de Comunicación podrán analizar gráficamente la evolución de la Sequía a partir de parámetros del tipo climatologías, evolución hídrica, seguimiento de predicciones, etc... que junto con la correlación de datos asociados a la incidencia en la Salud Pública como evolución de las enfermedades asociadas a la sequía, población afectada por enfermedades asociadas a la sequía, seguimiento de las campañas de vacunación, medidas mitigadoras, etc... permitirán obtener el comportamiento conjunto de los datos, que facilitarán a su vez determinar el comportamiento futuro del conjunto de datos analizados.

Las formas de presentar los datos y resultados son muy diversas, por lo que el Experto en *Business Analytics* será el responsable de diseñar y construir los gráficos precisos para aportar un mayor valor, proponiéndose los siguientes:

- Tabla: Permiten mostrar la información en números, haciéndolos más fáciles de descodificar a primera vista, ya que se pueden emplear colores y explicaciones previas.

- Gráfico de Barra: Facilita gráficamente la comparación entre elementos en un período de tiempo específico.

- Gráficos de Línea: Mostrando las relaciones de los cambios en los datos en un período de tiempo.

- Gráfico Circular: También conocido como pie en inglés. Se utiliza para mostrar cómo diferentes partes representan un total.

- Gráficos de Dispersión: Este tipo de gráfico son útiles para mostrar la relación entre diferentes conceptos de datos. Utilizando valores numéricos para ambos ejes en lugar de utilizar categorías en alguno de los ejes como en los gráficos anteriores.

- Gráfico de Burbujas: Se trata de una variación de un gráfico de dispersión en el que los puntos de datos se reemplazan por burbujas y el tamaño de las burbujas representa una dimensión adicional de los datos.

- *Treemap*: Consiste en un tipo de representación gráfica de datos jerárquicos en forma de rectángulos que ocupan el total del espacio de forma proporcional al valor de una variable.

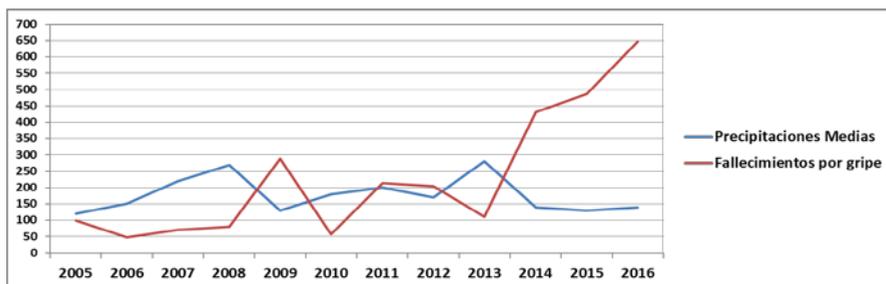
- Gráficos Sociales: Se trata de un mapa global que muestra con quién se relacionan las personas. Estos gráficos constan de nodos (personas) y flechas (relaciones) que conectan los nodos.

- Palabras: Se puede utilizar nubes de palabras o tags para analizar y descubrir tendencias en las Redes Sociales.

- Infografías: las infografías son más elaboradas y óptimas cuando se utiliza los datos para compartir información, difundirla y generar discusión, sobre todo con el objetivo de generar tráfico y enlaces para un sitio web.

A modo de ejemplo, a partir de la información almacenada en la aplicación de Big Data (información ya analizada, depurada y procesada previamente), los Medios de Comunicación como usuarios podrían generar un gráfico (mediante la herramienta de visualización) que correlacionase el nivel de precipitaciones medias anuales, con el índice de fallecimientos anuales por gripe en España. (Figura 4)

Figura 4. Ejemplo correlación precipitaciones anuales y fallecimientos por gripe. Imagen de elaboración propia.



Los actores que intervienen en esta fase son:

- Usuario de Datos: Aportando su valor a través de su experiencia en los Medios de Comunicación y la Salud Pública, será el validador final de los gráficos y de los resultados obtenidos.

- Experto en *Business Analytics*: Generará los gráficos, infografías y herramientas visuales que permita su explotación y análisis.

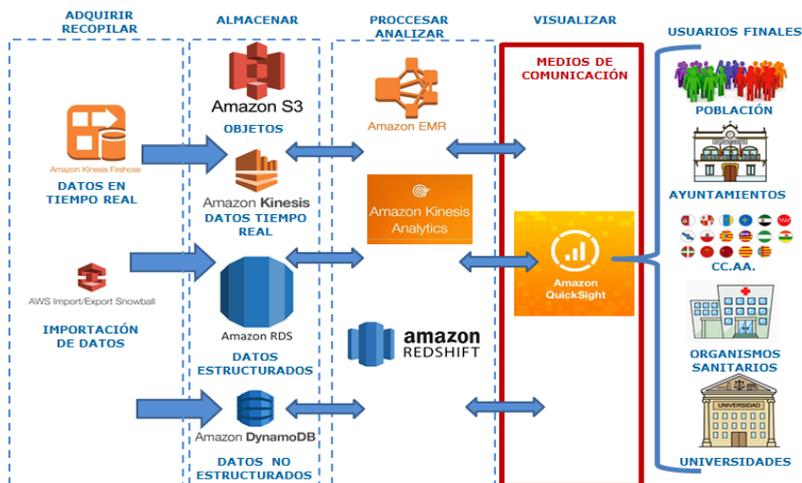
- Analista de datos: Aportará su conocimiento funcional y conceptual de la aplicación de tal manera que la visualización y análisis de los datos resultantes cubran las expectativas definidas inicialmente.

5. Herramientas de implementación

De cara a la implementación de nuestra plataforma de *Big Data*, una opción interesante es la que ofrece *Amazon Web Services (AWS)* (<https://aws.amazon.com/es/big-data/>), ya que proporciona una serie de herramientas que facilitan notablemente la construcción sencilla de este tipo de aplicaciones, siendo el coste de las mismas por uso y necesidades requeridas, lo que permite ajustar mucho el presupuesto. Además, ofrece un acceso rápido a recursos tecnológicos que permiten escalar la aplicación con rapidez, incluidos todos los componentes precisos para su construcción.

A continuación, se describen las herramientas recomendadas para cada una de las fases de construcción (Figura 5):

Figura 5. Herramientas para la construcción. Imagen de elaboración propia



5.1. Adquirir y recopilar datos

Datos en Tiempo Real (*Amazon Kinesis Firehose*): Esta aplicación, permite tratar datos en streaming. Los datos de streaming son datos que se generan constantemente a partir de miles de fuentes que normalmente envían los registros de datos simultáneamente en conjuntos de tamaño pequeño. Incluyen diversos tipos de datos, como archivos de registros generados por aplicaciones móviles o web, redes sociales, servicios geospaciales, así como telemetría de dispositivos conectados o instrumentación en centros de datos.

Importación de Datos (*AWS Import/Export Disk*): Solución que acelera la transferencia de grandes cantidades de datos desde y hacia la nube de AWS con dispositivos de almacenamiento portátiles para el transporte. *AWS Import/Export Disk* extrae datos de dispositivos de almacenamiento y transfiere datos a estos utilizando la red interna de alta velocidad de Amazon sin tener que pasar por Internet. En conjuntos de datos significativos, utilizar *AWS Import/Export Disk* suele ser más rápido que cuando se realizan transferencias a través de Internet y más económico que cuando hay que invertir en mejorar la conectividad.

5.2. Almacenar datos

- Almacenamiento de Objetos (*Amazon S3*): Es un servicio de almacenamiento de objetos creado para almacenar y recuperar cualquier volumen de datos desde cualquier ubicación: sitios web y aplicaciones móviles, aplicaciones y datos de sensores o dispositivos IoT (Internet de las Cosas)

- Almacenamiento en Tiempo Real (*Amazon Kinesis Streams*): Permite crear aplicaciones personalizadas que procesan o analizan datos de streaming para satisfacer necesidades especiales. *Kinesis Data Streams* puede registrar y almacenar de manera continua terabytes de datos por hora procedentes de cientos

de miles de orígenes, como secuencias de clics de sitios web, fuentes de redes sociales, logs de TI y eventos de seguimiento de ubicación.

- Almacenamiento Datos Estructurados (Amazon RDS): Corresponde a un servicio de base de datos relacionales en la nube. Permite obtener acceso a las funciones de las conocidas bases de datos *MySQL*, *Oracle*, *SQL Server* o *PostgreSQL*, entre otras.

- Datos No Sql (Amazon DynamoDB): Corresponde a un servicio de base de datos NoSQL rápido y flexible para todas las aplicaciones que requieren latencias constantes y de meros milisegundos a cualquier escala. Su modelo de datos flexible, desempeño fiable y escalado automático de la capacidad de desempeño lo hacen ideal para móviles, web, juegos, tecnología publicitaria, IoT y mucho más.

- Almacenamiento de Datos (*Amazon Redshift*): Simplifica la ejecución de consultas analíticas complejas en altos volúmenes de datos estructurados, utilizando una sofisticada optimización de consultas, almacenamiento en columnas en discos locales de alto desempeño y ejecución masiva de consultas paralelas. La mayoría de los resultados se producen en segundos.

5.3. Procesar y analizar

- Amazon EMR: Permite procesar de forma eficaz conjuntos de datos de gran tamaño. En lugar de utilizar un equipo grande para procesar y almacenar los datos, facilita analizar conjuntos de datos masivos en paralelo.

- En Tiempo Real (*Amazon Kinesis Analytics*): Ofrece la manera más sencilla de procesar datos de streaming en tiempo real con SQL estándar sin tener que aprender a usar lenguajes de programación ni marcos de procesamiento nuevos. Permite consultar datos de streaming o crear aplicaciones íntegramente de streaming con SQL, para que pueda obtener información procesable y satisfacer las necesidades de su empresa y sus clientes rápidamente.

5.4. Visualizar

- *Amazon QuickSight*: es un rápido servicio de análisis que se puede utilizar para crear visualizaciones, realizar análisis *ad hoc* y obtener rápidamente información útil a partir de los datos almacenados. *Amazon QuickSight* descubre los orígenes de datos de AWS, permite escalar a cientos de miles de usuarios y proporciona un desempeño fiable gracias a su sólido motor en memoria (*SPICE*).

6. Conclusiones

En la actualidad la mayoría de los Organismos Públicos y privados producen millones de datos, generando altos volúmenes de registros que ofrecen la posibilidad de extraer patrones de comportamiento. La aplicación propuesta en este trabajo es una sencilla herramienta de *Big Data* que estudia el comportamiento de la Sequía y su incidencia en la Salud Pública, con objeto de identificar comportamientos regulares e irregulares a partir de los datos históricos.

Vivimos la era de la inteligencia colectiva social, por tanto la utilización inteligente de este tipo de aplicaciones por parte de los Medios de Comunicación puede ser un camino para construir una sociedad más sostenible. De hecho, esta solución se propone como instrumento para generar contenidos informativos basados en evidencias fundamentadas con el objetivo de mejorar la calidad de vida de la sociedad a través de campañas de concienciación. Las empresas informativas juegan un papel esencial a la hora de concienciar a la ciudadanía en la práctica de conductas responsables a través de campañas preventivas. Contar con los Medios de Comunicación como usuarios de este tipo de aplicaciones, posibilita que los resultados obtenidos y recomendaciones analizadas lleguen de forma inmediata y amplia a la sociedad. La herramienta propuesta no tiene grandes barreras de entrada puesto que se construye a partir de software 100% accesible en el mercado, es de fácil uso y de baja inversión, pudiéndose dosificar los costes en función del uso que se realice del mismo.

Para concluir, es importante resaltar que la aplicación objeto de este trabajo se basa en el análisis de datos referidos a la Sequía y a la Salud Pública, pero una vez implementada esta solución de *Big Data* se puede reutilizar, sin costes adicionales puesto que la estructura operativa ya está funcionando, para gestionar cualquier otro contenido informativo que sea relevante para la población.

7. Referencias bibliográficas

- Agencia Estatal de Meteorología (AEMET): <http://www.aemet.es>.
- Agencia Europea de Medio Ambiente: “Contaminación atmosférica nociva”. En: <https://www.eea.europa.eu/es/pressroom/newsreleases/muchos-europeos-siguen-expuestos-a>
- Amazon Web Services: “Arquitecturas de Referencia para Big Data” En: <https://es.slideshare.net/AmazonWebServicesLATAM/patrones-de-arquitectura-para-big-data-en-aws>
- Amazon Web Services: “Big Data AWS”. En: <https://aws.amazon.com/es/big-data/>
- BBVA: “Ideas con 'big data' para luchar contra el cambio climático” En: <https://www.bbva.com/es/ideas-big-data-luchar-cambio-climatico/>
- Big Data COE: “Los 7 perfiles clave de los profesionales del Big Data” En: <https://www.bigdataabcn.com/es/los-7-perfiles-clave-los-profesionales-big-data/>
- Bonechi, Michele (2012): “Interview with Simon Rogers”. *Editor Guardian Data Blog, London*. En: www.youtube.com/watch?V=bU8DiD2J-tw
- CALIOPE [Sistema de pronóstico de calidad del aire]: <http://www.bsc.es/caliope/es>
- Centro de Investigaciones Sociológicas [CIS]: <http://www.cis.es/cis/opencms/ES/NoticiasNovedades/InfoCIS/2014/PlataformaOnLine/BancodeDatos.html>
- Deloitte: “Roles de un Big Data”. En: <https://www2.deloitte.com/es/es/pages/operations/articles/big-data-rol-empleado.html>
- El País* (2017): “Retrato de una España atrapada en la sequía”. En: https://politica.elpais.com/politica/2017/10/05/actualidad/1507198653_437781.htm
- Federación internacional de Sociedades de la Cruz Roja y de la Media Luna: “Definiciones de Sequías”. En:

- <http://www.ifrc.org/es/introduccion/disaster-management/sobre-desastres/definicion--de-peligro/sequias/>
- Flores-Vivar, Jesús M. (2015): "Contextualización, usos empíricos y etnografía de las redes sociales en el ciberperiodismo". *Estudios sobre el Mensaje Periodístico*, 21, nº especial Diciembre, 81-95.
- Fundación Ramón Areces: "Modelos econométricos de cambio climático" En: <http://www.fundacionareces.tv/watch/bigdataycambioclimatico?as=56f81c20a8c7fb81338b45e8>
- Greenpeace: "Sequía algo más que falta de lluvia". En: https://es.greenpeace.org/es/wp-content/uploads/sites/3/2017/11/Sequia-Falta-de-Agua_WEB-1.pdf
- Gray, Jonathan; Chambers, Lucy; and Bounegru, Lucy (2012): "The Data Journalism Handbook". *O'Reilly, Sebastopol*. ISBN: 978-1-449-33006-4
- Guerrero, Manuel A. (2008): "Medios de comunicación y la función de transparencia". *Instituto Federal de Acceso a la Información Pública (IFAI)*, México, D.F. ISBN: 968-5954-29-1
- Hellerstein, Joseph (2008): "The Commoditization of Massive Data Analysis". *Blog on O'Reilly.com*, 19. En: <http://strata.oreilly.com/2008/11/the-commoditization-of-massive.html>.
- Instituto de Ingeniería del Conocimiento: "Análisis de Datos". En: <http://www.iic.uam.es>
- IPCC (2007): "Climate Change 2007. Impacts, adaptation and vulnerability. (Working Group II Report). *United Nations Environment Programme*. En: <http://www.ipcc.ch/ipccreports/assessments-reports.htm>.
- IPCC (2007): "Climate Change 2007. The Physical Science Basis". (Working Group I Report). *United Nations Environment Programme*. En: <http://www.ipcc.ch/ipccreports/assessmentsreports.htm>.
- Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente (MAPAMA): <http://www.mapama.gob.es/es/>
- Meteorología en Red: "Predicciones". En: <https://www.meteorologiaenred.com/category/predicciones>
- Ministerio de Sanidad, Servicios Sociales e Igualdad: "Impactos del Cambio Climático". En: <https://www.msssi.gob.es/ciudadanos/saludAmbLaboral/docs/CCCompleto.pdf>
- Ministerio de Sanidad, Servicios Sociales e Igualdad: "Cambio Climático y Salud". En: <http://www.msssi.gob.es/ciudadanos/saludAmbLaboral/medioAmbiente/cambioClimatico.htm>
- Murphy, Michael & Barton, John. (2014): "From a sea of data to actionable insights: Big data and what it means for lawyers". *Intellectual property & technology law journal*, v. 26, n. 3, 8-17.
- Ministerio de Sanidad Servicios Sociales e Igualdad (MSSSI): <https://sede.msssi.gob.es/home.htm>
- Observatorio de Sostenibilidad: "Big Data". En: <http://www.observatoriosostenibilidad.com/big-data/>
- Organización Meteorológica Mundial: "Análisis de las olas de Calor". En: https://www.wmo.int/pages/index_es.html
- Picard, Robert (2010): "Value Creation and the Future of News Organizations: Why and How Journalism Must Change to Remain Relevant in the Twenty-First Century". *Lisboa, Media XXI*. ISBN: 9898143274

Salesforce: “Transformación digital”. En:

<https://www.salesforce.com/mx/blog/2017/12/Que-es-la-transformacion-digital.htm>

Sistemas de Datos Abiertos del Gobierno de España: <http://datos.gob.es/es>

Synergic Partners: “Extraer valor del dato”. En: <http://www.synergicpartners.com/como-se-extrae-valor-del-dato/>

Synergic Partners: “Las Vs en el Big Data”. En: <http://www.synergicpartners.com/donde-esta-la-v-en-big-data-entendiendo-que-es-big-data-de-una-vez-por-todas/>

Ubaldo Cuesta Cambra es Catedrático. Facultad CC. de la Información (Universidad Complutense de Madrid). Y Director Departamento Teoría y Análisis de la Comunicación. Director del Máster de Gestión Publicitaria (UCM). Director Revista Comunicación y Salud y Co-Director de la Revista Pensar la Publicidad (PLP). Vocal Consejo Español de Drogodependencias y otras Adicciones. Director laboratorio de Neuromarketing “NeuroLabCenter” (CAP-UCM) y del Grupo de Investigación “Creación y Efectos psicosociales y culturales del discurso audiovisual”(UCM). Autor de varios libros y artículos especializados en el ámbito de la psicología y la comunicación.

José Ignacio Niño González es Profesor Asociado Planificación Publicitaria. Facultad CC. de la Información (Universidad Complutense de Madrid). Doctor en Publicidad y Relaciones Públicas, Universidad Complutense de Madrid. Secretario Comisión Académica del Programa de Doctorado en Comunicación Audiovisual, Publicidad y Relaciones Públicas, Facultad de Ciencias de la Información (UCM). Estancias de investigación en la Universidad de Pekín (China), Universidad de Zacatecas (México) y Universidad de Oporto (Portugal). Subdirector laboratorio de Neuromarketing “NeuroLabCenter”, del Departamento CAP la Facultad de Ciencias de la Información (UCM). Autor de varios artículos en revistas científicas, capítulos de libros y libros. Trabaja en proyectos de innovación tecnológica y consumo mediático.

Luz Martínez Martínez es Profesora Asociada Departamento Comunicación y Sociología. Facultad CC. de la Información (Universidad Rey Juan Carlos). Doctora en Publicidad y Relaciones Públicas, Universidad Complutense de Madrid. Autor múltiples artículos en revistas científicas, capítulos de libros y libros. Trabaja en proyectos de innovación tecnológica y consumo mediático.

Carlos Díez Bielsa es Head of Digital Practices. Profesional experto en Transformación Digital, con más de 20 años de experiencia en Tecnologías de la Información, así con alta cualificación en el Desarrollo de Negocio e Innovación en Prácticas Digitales: Big Data, Smart Mobility, CRMs, Contact Center, Integration Platform, desarrollando dichas responsabilidades en las principales compañías globales IT e implementado soluciones de negocio para clientes pertenecientes a todos los sectores del Mercado.