



# Fundamentos de la Ciencia de Datos

---

## PEC I

*Julio M. Fernández Jiménez*  
21/03/2021

## Ejercicio I

Podemos definir *Open Data* como todo aquel conjunto de datos que se encuentra disponible para cualquier individuo sin restricciones de derechos de autor, patentes o limitaciones en su uso y explotación <sup>(1)</sup>. Se entiende, por ende, que son de dominio público.

La existencia y distribución pública de esta información está claramente motivada por la evolución tecnológica de los últimos años y la existencia de nuevos perfiles técnicos que, junto con la existencia de nuevas herramientas de minería y análisis de datos, han permitido que sea posible extraer conocimiento de los datos existentes.

Una de las ramas del conocimiento que más se está beneficiando de esta democratización de la información es la ciencia de la salud o biomédica <sup>(1)</sup>. Podría decirse, sin duda alguna, que el sector de la salud es el que más datos genera hoy en día. Las razones son evidentes, y es que cada uno de nosotros es una fuente de información de datos en crudo con un amplio abanico de posibilidades, y tan diverso como lo son las circunstancias individuales de cada uno de nosotros <sup>(2)</sup>.

Otra de las motivaciones de la recopilación, distribución y uso de la información en este campo es que el conocimiento extraído tiene un impacto directo, personalizado y práctico en cada persona mejorando la calidad de vida en términos de salud y bienestar a corto plazo; cosa que no ocurre precisamente con el conocimiento que se podría extraer de otro tipo de datos, como los relacionados con las *fintech* o la astrofísica por mencionar algunos.

Los beneficios del uso de datos sanitarios bajo la filosofía *open data* son diversos e implican tanto al sistema como al usuario. Entre los más destacados se encuentra el de mejorar la eficiencia del sistema de salud mediante el conocimiento adquirido del estudio de los datos, dando paso a una reducción en los costes en sanidad y un mejor servicio <sup>(3)</sup>. La disposición al público de dichos datos, promueve la transparencia del sistema, ya que el conocimiento extraído estaría al alcance de todos a través de los datos en crudo y su posterior interpretación. Este conocimiento globalizado, que a medio plazo se ve reflejado en una mejor educación sanitaria por parte del paciente, repercutiría directamente en una mejora de los tratamientos suministrados donde se hace al usuario participe del proceso de tratamiento y sanación <sup>(4)</sup>.

En adición, la apertura de datos sanitarios daría paso un nuevo mercado orientado a la salud y la prevención que transformaría por completo el sistema sanitario actual al convertirlo en un modelo preventivo de salud en decremento del actual sistema reactivo.

Uno de los requisitos fundamentales de cualquier información catalogada bajo la premisa de *Open Data* es que la información que se comparta sea anónima, no rastreable e irreversible <sup>(4)</sup>, es decir, que no sea posible identificar la identidad o fuente de cada una de las entradas que componen el conjunto de datos abiertos al cual nos referimos, incluso después de haber garantizado su anonimato. Este requisito esencial se ha convertido en uno de los principales retos de la política de datos abiertos en el ámbito de la salud.

Paradójicamente, lo que hace del campo de la biomedicina uno rico en diversidad y volumen de datos es lo que lo hace también altamente vulnerable a la privacidad y a los principios éticos de nuestra sociedad moderna. El hecho de que estemos tratando con información personal, que podría ser utilizada en nuestro perjuicio, hace del tratamiento de datos médicos un laberinto legal difícilmente salvable.

Entre los principales retos jurídicos que encontramos a la hora de divulgar datos médicos, está la disparidad legal entre las normativas de protección de datos de europeas, nacional y regionales. Aunque la normativa europea contempla la liberación de datos médicos anónimos con fines no lucrativos (y bajo estrictas normas de seguridad y garantías adicionales), es España, la legislación sólo contempla el uso de estos dentro del ámbito clásico médico-paciente, aunque recientemente se han añadido extensiones legislativas que permiten el uso de dicha información por terceros. Gracias a esta extensión sobre el uso de datos médicos con fines no clásicos, y como consecuencia de la emergente necesidad del uso de dicha información por terceros, algunas comunidades autónomas han legislado a favor de un uso controlado y monitorizado de dichos datos en el mejor interés del público <sup>(4)</sup>. Esta apertura gradual de la legislación al uso de datos médicos está permitiendo que los beneficios y objetivos previamente mencionados estén paulatinamente permitiendo el desarrollo de un entorno *Open Data* en el campo de la salud nacional.

La situación de alerta sanitaria por la que el planeta está pasando desde el último año ha puesto en el punto de mira nuestro modelo actual de privacidad de la información. En una pandemia global a tiempo real, se ha demostrado que el intercambio y uso de la información es primordial para promover una lucha efectiva contra el patógeno en cuestión. Acceder a datos médicos de pacientes infectados a escala global es esencial para extraer el conocimiento que nos permitiría conocer cómo actúa el virus y qué estrategias son las más eficientes para su control.

Muchos de los errores que como sociedad hemos cometido durante la lucha contra el COVID-19 han sido resultado de una evidente falta de información potencialmente derivada de datos reales. Personalmente, estoy convencido de que la burocracia, las normativas de protección de datos y privacidad y las ambiciones de terceros, en lo que concierne al control de la información, han obstaculizado en gran medida la apertura de la información.

La existencia de un sistema utópico de distribución de datos abiertos podría, en gran medida, haber ayudado a entender el COVID-19 (y otros patógenos) y, en consecuencia, a diseñar estrategias más eficientes y reales para su total exterminación. Sin embargo, una vez más, hemos sido víctimas de un sistema legal que en principio fue diseñado para nuestra protección y beneficio. La intrínseca falta de ética de nuestra sociedad moderna, y que demanda de un sistema legal basado en el control, ha jugado en nuestra contra cuando más falta nos ha hecho; y es que, hay circunstancias, en las que el sentido común debería estar por encima de cualquier sistema legal y nunca fue diseñado para circunstancias excepcionales como la que, al día de hoy, aún experimentamos.

(1) **Marçal Mora Cantallops**, “Los roles, ámbitos y nombres de la ciencia de datos”. Editorial UOC (págs. 42, 28)

(2) Data Saves Lives. *COVID-19: The future of healthcare: is it all about data?* [en línea]. Actualizada: 15 de octubre de 2020. [Fecha de consulta: 20 de febrero de 2021]. Disponible en:

<https://datasaveslives.eu/blog/isitalaboutdata15102020>

(3) Finanzas.com. HM Hospitales arranca su plan de transformación digital invirtiendo 20 millones de euros entre 2019 y 2023 [en línea]. Actualizada: 2019. [Fecha de consulta: 20 de febrero de 2021]. Disponible en: <https://opinno.com/es/insights/no-queremos-curar-queremos-llegar-antes-que-la-enfermedad>

(4) Andreu, M<sup>a</sup> Belén. Open Data en el ámbito sanitario y su compatibilidad con la privacidad del paciente [en línea]. Actualizada: 2017. [Fecha de consulta: 20 de febrero de 2021]. Disponible en: <https://ojs.imodev.org/index.php/RIGO/article/view/200/330>

## Ejercicio 2

La meta a largo plazo del grupo HM Hospitales es la de convertirse en una organización orientada a datos dentro del ámbito de su actividad, como es evidente. La divulgación de los datos COVID-19 ha sido una muestra de su capacidad e intención de acometer su labor <sup>(1)</sup>.

Para llevar a cabo esta labor, es necesario contar con una serie de perfiles profesionales que permitan la transformación de forma fluida. Los dos tipos de perfiles profesionales básicos identificados son los *quants* y los *non-quants* <sup>(2)</sup>.

Un perfil *quant* es aquel que está compuesto por analistas expertos en el manejo de los datos a nivel técnico así como en las herramientas, métodos y soluciones de análisis de datos. Por el contrario, el perfil *non-quant* cubre a los usuarios de negocio y gestión que hacen uso de las soluciones analíticas implementadas en la organización a través de los procesos de los profesionales *quant* <sup>(2)</sup>. Es indudable que existe una simbiosis profesional entre ambos perfiles; para que los perfiles *non-quants* pueda realizar su función, es necesario que los integrantes del perfil *quant* haya completado su trabajo. Por otro lado, el trabajo de los *quants* viene definido por las especificaciones de los perfiles *non-quants*.

Los perfiles *no-quants*, pueden ser catalogados como los *stakeholders* del proyecto. Son perfiles administrativos y nada o poco técnicos. En el ámbito del sector *quant* o técnico, tenemos dos tipos de subperfiles: los **perfiles tradicionales** y los **nuevos perfiles**. El perfil tradicional es más común y envuelve conocimientos en inteligencia y analítica de negocios; por otro lado, los nuevos perfiles incluyen destrezas orientadas a las nuevas tecnología orientadas a datos <sup>(2)</sup>.

En el caso concreto de HM Hospitales, es necesario cubrir una serie de perfiles que puedan, en conjunto, llevar a cabo la meta de transformación digital propuesta. Los perfiles más relevantes y necesarios son <sup>(2)(3)(4)</sup>:

- **Perfiles de negocio:** los siguientes roles son necesarios para la correcta puesta en marcha, gestión y organización del proceso de transformación digital de HM.
  - **Patrocinador del proyecto:** generalmente el CEO de la organización. Es quien apoya e impulsa el desarrollo del proyecto. Debe ser un perfil con conocimientos en gestión de empresa y negocios, buen comunicador y visionario.
  - **Gerente de proyecto:** responsable de ejecutar el proyecto de transformación. Este perfil requiere de conocimientos de gestión de proyectos desde un punto de vista no técnico.
  - **Analista de negocio:** proporciona los requisitos del proyecto. Es imprescindible que el analista de negocio conozca la dinámica del campo de la sanidad pública y privada, así como las necesidades del usuario final.
  - **Usuario de negocio:** usuario final que hará uso de las herramientas desarrolladas durante el desarrollo del proyecto. Este perfil es el más cercano al paciente y es quien aplicará las técnicas adoptadas sobre los usuarios. Es un perfil práctico con conocimientos del campo de la salud.

- **Perfiles técnicos tradicionales:** estos perfiles tradicionales son los encargados de aplicar el conocimiento extraído de los datos en la organización de una forma práctica y adaptada a las necesidades de HM.
  - **Director BI/DW:** está a cargo de las iniciativas de explotación de datos. Detecta las necesidades de la empresa y localiza un filón de datos de donde extraer inteligencia y conocimientos para llevar a cabo la función deseada. Este perfil es probablemente el más cercano a los perfiles de negocio y debe conocer el proyecto desde un punto de vista empresarial. Sin embargo, tiene que ser capaz de identificar aquellas áreas de minería de datos que permitan extraer el conocimiento necesario para las necesidades del negocio.
  - **Gerente de proyectos:** Permite la implementación de proyecto. Al igual que el perfil de gerente de proyecto descrito en el área de negocio. Este perfil técnico tradicional debe tener amplios conocimientos en gestión de proyectos pero desde una perspectiva técnica y orientada a sistema BI/DW.
  - **Analista de calidad de datos:** Es un perfil extremadamente técnico y muy cercano a la estadística. Su responsabilidad es la de proveer datos limpios y utilizables desde una conjunto de datos en crudo.
  - **Arquitecto de datos/DBA:** trabaja directamente con sistemas de gestión de bases de datos. Su función es el mantenimiento y arquitectura de sistemas de almacenamientos de datos masivos
  - **Gerente de seguridad:** experto en seguridad e imprescindible para cumplir con las normativas de protección de datos y evitar que la información almacenada pueda ser interceptada o manipulada por personal no autorizado.
  - **Especialista en minería de datos/estadística:** encargado de generar modelos de datos utilizados para la futura predicción y prevención. Es un perfil técnico que necesita conocimientos en estadísticas y matemáticas.
  - **Formador:** necesario para la formación de los usuarios de negocios. Necesita de conocimientos en educación con un perfil técnico en ingeniería de sistemas o gestión.
- **Nuevos Perfiles:** Orientado a las nuevas tecnologías de manejo y explotación de datos.
  - **Científico de datos:** procesa datos en bruto e identifica en tipo de análisis se deben aplicar. Este perfil debe tener conocimientos en matemáticas, estadísticas e inteligencia artificial entre otras disciplinas técnicas relacionadas. Entre sus habilidades, se espera que disponga del manejo de diferentes lenguajes de programación y herramientas de manipulación de datos.
  - **Especialista de visualización:** permite contextualizar los datos con la finalidad de obtener el conocimiento. Este perfil es el enlace entre los *quants* y los usuarios de negocios o *no-quants*.
  - **Arquitectos cloud computing:** imprescindibles para la creación de infraestructuras en las que implementar las soluciones generadas. Es un perfil basado en ingeniería de sistemas orientada a Dev-ops.
  - **Chief data officer:** Responsable de las iniciativas de explotación de datos a nivel técnico y contraparte técnica no tradicional del director BI/DW. Debe tener condimentos en gestión de proyectos e ingeniería de datos.

De forma general, los perfiles *non-quants* deben poder contar con tiempo y atención para comprender el problema a solucionar así como una noción real del tiempo y los gastos que la ejecución del proyecto pueda consumir.

Por otro lado, de los perfiles *quants* se espera que desarrollen un buen conocimiento del negocio, el lenguaje que conlleva y el proceso en el que trabajan así como saber ponerse en el lugar del usuario para el que realizan la labor. Saber gestionar y estimar el proyecto en términos de tiempo y coste, tener destrezas comunicativas para con el personal no técnico y saber extraer la información necesaria del cliente o personal de negocios.

(1) La Razón. HM Hospitales aúna investigación y Big Data para cercar a la Covid-19 [en línea]. Actualizada: 31 de diciembre de 2020. [Fecha de consulta: 20 de febrero de 2021]. Disponible en: <https://www.larazon.es/salud/20201231/yzm7vnrlzgrbpvc4f76fbak7q.html>

(2) **Josep Curto Díaz**, *“Organizaciones orientadas al dato”*. Editorial UOC (págs. 24-26, 34-37, 39-40)

(3) Finanzas.com. HM Hospitales arranca su plan de transformación digital invirtiendo 20 millones de euros entre 2019 y 2023 [en línea]. Actualizada: 2019. [Fecha de consulta: 20 de febrero de 2021]. Disponible en: <https://opinno.com/es/insights/no-queremos-curar-queremos-llegar-antes-que-la-enfermedad>

(4) El Español. Telefónica y HM Hospitales digitalizan el sector salud [en línea]. Actualizada: 30 de abril de 2019. [Fecha de consulta: 20 de febrero de 2021]. Disponible en: [https://www.elespanol.com/invertia/disruptores-innovadores/innovadores/20190430/telefonica-hm-hospitales-digitalizan-sector-salud/393962093\\_0.html](https://www.elespanol.com/invertia/disruptores-innovadores/innovadores/20190430/telefonica-hm-hospitales-digitalizan-sector-salud/393962093_0.html)

## Ejercicio 3

El proyecto **Covid Save Lives** (CSL) es una iniciativa del grupo sanitario HM Hospitales que pretende poner en disposición de centros de investigación, públicos y privados, datos anónimos de los historiales clínicos de pacientes que han sido contagiados con COVID-19 y tratados en los centros de Galicia, Madrid y Barcelona durante el último año <sup>(1)(2)</sup>.

El proyecto CSL es una iniciativa *Open Data* que está limitada por la legislación nacional vigente en lo que a su divulgación concierne, razón por la que su divulgación es limitada y restringida a un sector reducido de usuarios y tras previa evaluación y aprobación de una solicitud de acceso y explotación <sup>(3)</sup>.

Todo conjunto de datos hace uso de un ciclo de vida programático que permite que puedan ser reutilizados sucesivamente en otros procesos. Este ciclo genera valor a partir de los datos en cada uno de los pasos que lo compone. Los pasos a seguir en el proceso de ciclo de vida o de transformación de los datos son <sup>(4)</sup>:

1. **Captura:** En este paso se obtienen los datos que deseamos procesar. Los datos pueden ser generados, a través de procesos concretos como la medición de parámetros ambientales o acciones realizadas por individuos o máquinas, o capturados desde sistemas existentes como servicios API o bases de datos *Open Data* entre otros.
2. **Almacenamiento:** Para poder disponer de los datos, o distribuirlos de forma adecuada, es necesario almacenarlos en sistemas físicos destinados al almacenamiento de información. Los formatos de almacenamiento y el medio pueden diferir según las necesidades y los diferentes estándares a seguir.
3. **Preprocesado:** Este es primero de los pasos en vistas a la extracción de la información y que conlleva una transformación de los datos en crudo orientado a la hipótesis que deseamos probar. En este paso podemos descartar datos irrelevantes (no relacionados con nuestro trabajo), fusionar datos existentes para generar nuevos datos, transformarlos o limpiarlos.
4. **Análisis:** Este es quizás el paso más crítico. Conlleva la creación de modelos que expliquen el comportamiento de los datos en base a patrones y relaciones matemáticas encontradas. Este proceso conlleva la aplicación de estadísticas, procesos de aprendizaje de máquina o inteligencia artificial. También incluye la explicación del modelo en términos de la morfología y la estructura de los datos.
5. **Visualización:** El proceso de visualización permite explorar los datos obtenidos mediante herramientas de visualización tales como gráficos, tablas, filtros y cruces de información con otros modelos entre otros.
6. **Publicación:** Finalmente, este paso conlleva la publicación de los datos y modelos obtenidos en los tres pasos anteriores para que otros puedan hacer uso de ellos.

Los datos actuales publicados para el CSL están disponibles desde la [web oficial de HM hospitales](#) <sup>(2)</sup> y han sido desarrollados hasta el segundo paso del ciclo de vida de los datos. Según podemos inferir de la descripción del juego de datos, estos han sido previamente **capturados** directamente de los pacientes o usuarios de los centros de salud ya mencionados. De ellos se ha recopilado cierta información en crudo tales como los datos demográficos del paciente, su estancia en UCI, constantes vitales en urgencias, medicación suministrada, resultados de pruebas

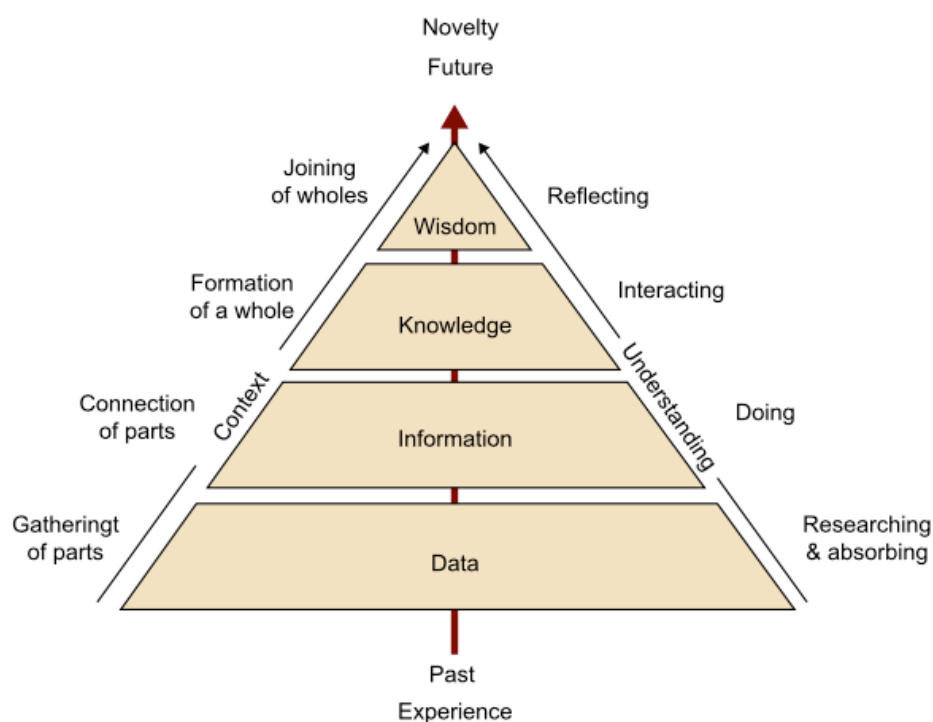


de laboratorio e información de diagnósticos. Esta información ha sido **almacenada** en archivos *datasets* (probablemente preparados para su procesamiento en R o CSV). No queda, claro sin embargo, si los datos han pasado por algún preprocesado previo a su **publicación**.

Como bien puede preciarse, los datos CSL no han sido llevados a través de ningún proceso de análisis o visualización, ya que éstas fases son delegadas a los equipos de investigación que solicitan el acceso a los mismos.

A pesar de que los datos publicados han pasado por ciertos procesos incluidos en el ciclo de vida de datos, desde el punto de vista del centro de investigación que haga uso de ellos, éstos se encuentra en el paso uno, es decir, en el de captura. Desde la perspectiva de estas instituciones, se estaría accediendo a los datos desde un sistema de *Open Data*, invitándoles a pasar de nuevo por los pasos subsiguientes según su línea de investigación. Cada organización que haga uso de los datos debe pues cumplimentar con un proceso de almacenamiento apropiado a su sistema de análisis, la realización de un preprocesado adaptado a su hipótesis, análisis, visualización y finalmente publicación de los resultados para su futuro uso y adaptado a su modelo de investigación.

La pirámide DIKW <sup>(4)</sup>, que incluye las siglas del inglés para los términos *data*, *information*, *knowledge* y *wisdom* (datos, información, conocimiento y sabiduría), determina el ciclo de vida de los datos. Este ciclo de vida permite escalar la información desde una colección de datos en crudo hasta la extracción de un conocimiento consciente de la información que da lugar a la extrapolación del conocimiento con fines prácticos. Como cabe de esperar, esta pirámide conlleva el ciclo de vida de lo datos discutido con anterioridad, aunque descrito desde un punto de vista práctico y orientado a la extracción de la información. Los estados por los que un conjunto de datos debe pasar son:

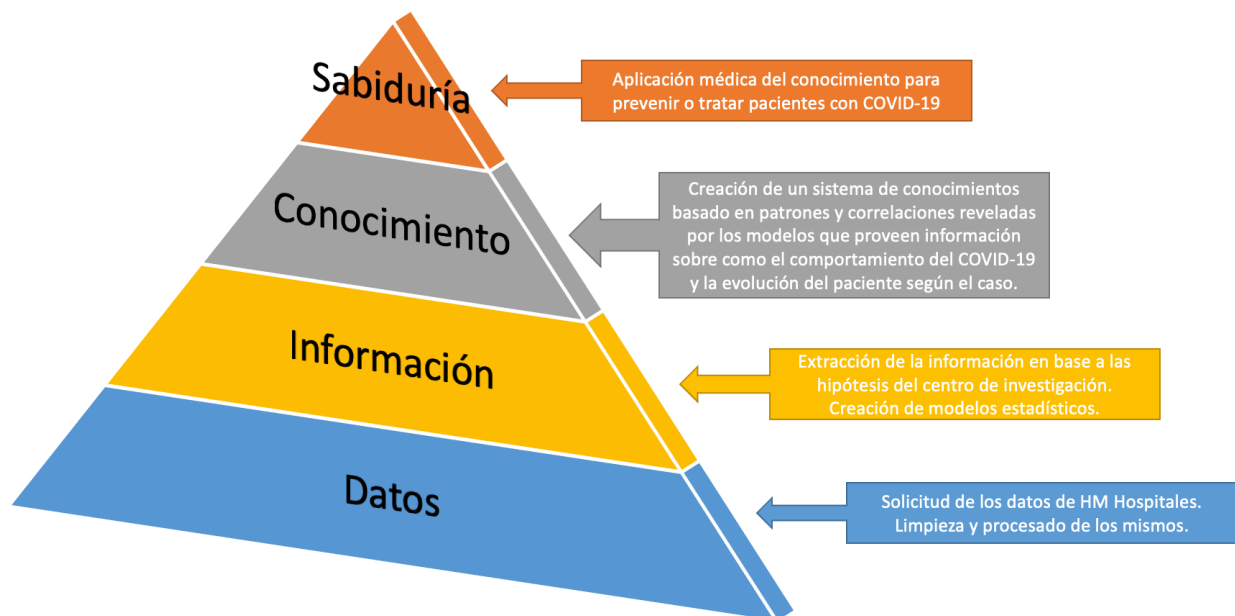


1. **Dato:** Este bloque se encuentra en la parte inferior de la pirámide y representa una colección de datos recolectados en crudo y sin ninguna interpretación lógica o práctica.
2. **Información:** representa el primer paso del procesamiento de los datos en crudo donde se le provee de contexto.
3. **Conocimiento:** Denota un uso práctico sobre cómo hacer uso de la información adquirida de los datos.
4. **Sabiduría:** Implica la capacidad de hacer uso del conocimiento adquirido con fines concretos.

Como todo conjunto de datos, los datos CSL también siguen esta evolución piramidal. A continuación aplicaremos la pirámide DIKW a los datos de ***Covid Data Save Lives:***

1. **Datos:** En la base inferior de la pirámide tenemos los datos en crudo provistos por HM Hospitales. Estos datos contienen información sobre el cuadro clínico de los pacientes, constantes vitales, ingresos en UCI, tratamiento, estado social y posibles antecedentes médicos. Este conjunto de datos es un sistema vivo que está en constante crecimiento y actualización. Si el proceso del ciclo de datos esta bien diseñado, éste debe fácilmente admitir la adición de nuevos datos para su incorporación automatizada, resultando en un modelo más preciso, que a su vez, repercute en un conocimiento más certero. En este escalón piramidal podemos ver reflejados los pasos de captura y almacenamiento de los datos descritos anteriormente.
2. **Información:** En el siguiente escalón de la pirámide, procedemos a la extracción de la información de los datos. Es aquí donde damos contexto y sentido a los datos orientándolos hacia el conocimiento que deseamos extraer de los mismos. En este paso realizamos el procesamiento de los datos y su análisis. La entidad que lleva a cabo la investigación a través de los datos debe decidir qué datos desea conservar y cuales descartar, así como cuales deben asociarse o derivarse. Un ejemplo de este proceso seria el de relacionar los datos demográficos de los pacientes, la medicación suministrada y la evolución de la enfermedad, siendo ésta última el resultado de la fusión de los análisis de laboratorio y el diagnostico médico. Por último, el *dataset* resultante es sometido a un análisis que debe dar como resultado el modelo a utilizarse para la extracción del conocimiento. La información extraída en este proceso es considerada como hecho derivado de los datos y será utilizada en los siguientes escalafones de la pirámide.
3. **Conocimiento:** En esta fase visualizamos la información para extraer conclusiones de los datos haciendo uso de los modelos generados en el paso anterior. Es aquí donde debemos interpretar los tests realizados sobre los modelos generados llegando a conclusiones practicas sobre los hechos ocurridos y reflejados en los datos estudiados. Por ejemplo, podríamos deducir qué pacientes denotan un mayor riesgo de ingreso en UCI tras una infección con COVID-19. Este paso conlleva el proceso de visualización y publicación del ciclo de vida de los datos.
4. **Sabiduría:** En este nivel ponemos en práctica los conocimientos adquiridos en el paso anterior mediante su implementación en la medicina práctica. Por ejemplo, si en el paso de la recopilación de conocimiento se ha detectado una correlación entre el el sexo del paciente y su respuesta concreta a un medicamento, lo ideal sería comenzar a suministra dicho medicamento pacientes que encajen con el patrón demográfico identificado.

A continuación mostramos una pirámide DIKW adaptada al modelo CSL estudiado basado en la explicación anterior.



(1) Razón. HM Hospitales aún investiga y Big Data para cercar a la Covid-19 [en línea]. Actualizada: 31 de diciembre de 2020. [Fecha de consulta: 20 de febrero de 2021]. Disponible en: <https://www.larazon.es/salud/20201231/yzm7vnrlzgrbpvc4f76fbak7q.html>

(2) HM Hospitales. Covid data save lives [en línea]. Actualizada: Abril de 2020. [Fecha de consulta: 20 de febrero de 2021]. Disponible en: <https://www.hmhospitales.com/coronavirus/covid-data-save-lives>

(3) Andreu, M<sup>a</sup> Belén. Open Data en el ámbito sanitario y su compatibilidad con la privacidad del paciente [en línea]. Actualizada: 2017. [Fecha de consulta: 20 de febrero de 2021]. Disponible en: <https://ojs.imodev.org/index.php/RIGO/article/view/200/330>

(3) Julià Minguillón, “Fundamentos de Data Science”. Editorial UOC (págs. 5-6, 17-32)

## Ejercicio 4

El grupo HM Hospitales está actualmente pasando por un proceso de transformación digital enfocada en sus centros de la Comunidad de Madrid, Castilla y León, Galicia, Castilla La Mancha y Cataluña. El proyecto aspira a mejorar la experiencia del cliente al aumentar la eficiencia y transparencia de los servicios. Con la finalidad de alcanzar dicha meta, HM Hospitales se ha afiliado con Telefónica como proveedor de servicios digitales entre los que destacan aquellos que involucran al almacenamiento de datos en la nube y la seguridad de los mismos <sup>(1)</sup>.

Gracias a las nuevas infraestructuras digitales incorporadas, el grupo HM aboga por el denominado *hospital líquido*. Este nuevo concepto de salud aspira a la descentralización de la asistencia hospitalaria mediante el uso de video consultas, digitalización de servicios generales y la incorporación de comparecencias telemáticas <sup>(2)</sup>.

HM quiere reducir la estancia de los usuarios en sus facilidades a lo mínimamente imprescindible mediante la digitalización de los procesos que se puedan llevar a cabo antes y después de una intervención física; esta nueva dinámica de atención semi-virtual pretende agilizar los procesos de diagnóstico, tratamiento y seguimiento del paciente en vistas a mejorar la experiencia del usuario y a liberalizar el uso de las facilidades a lo estrictamente necesario <sup>(2)(3)</sup>.

Esta transformación conlleva el reto de un cambio cultural y digital en el usuario, especialmente en aquellos de edad avanzada, que con la amenaza del COVID-19 ya han visto sus hábitos médicos modificados. De igual modo, la aparición de la reciente pandemia, ha forzado a una rápida implementación de dicha transformación permitiendo realizar en meses lo que en principio se había estimado como una labor de años de implementación <sup>(2)</sup>.

Otro objetivo de esta transformación digital es el de orientar HM hospitales a una organización orientada a datos con el objetivo de ofrecer tratamientos personalizados basados en el perfil genético del paciente y de encaminar las prácticas sanitaria de HM a un enfoque preventivo más que reactivo. El uso de big data, cuando el contexto lo requiera, permitirá que HM pueda prever escenarios médicos en los usuarios antes de que éstos ocurran <sup>(3)</sup>.

Actualmente, HM no puede ser considerada como una organización completamente orientada a datos, que se gestiona fundamentalmente en hechos y datos, aunque sí lo es de forma parcial. El proyecto *Covid Data Save Lives* es, sin duda alguna, un primer gran paso para esta transformación de HM Hospitales a una organización orientada a datos. A pesar de ello, y como ya se han comentado, el grupo HM tiene una ruta claramente trazada para alcanzar este objetivo.

Los requisitos mínimos para que una organización sea considerada como una orientada a datos son <sup>(4)</sup>:

- Identificar, combinar y gestionar múltiples fuentes de datos. Actualmente, sólo CSL ha sido puesta a disposición pública como fuente oficial de datos, aún quedaría por ver qué otros recursos están por ser liberados como resultado de la actividad de HM.
- La organización debe ser capaz de construir modelos avanzados para predecir y optimizar sus resultados. Este paso conlleva una de las acciones propuestas por HM en vistas a actualizar

sus servicios. Los datos de su recientemente publicado *dataset* están siendo modelados por terceros, aunque sí es cierto que no existe ninguna declaración oficial que demuestre lo contrario. Estaría en manos de HM el crear modelos productivos de estos datos e incorporarlos en el sistema de diagnóstico y tratamiento de sus paciente de COVID para cumplimentar con este requisito. Para ello, sin embargo, HM necesita de un equipo técnico de perfiles *quant* operando internamente, o al menos mantener cierta coordinación entre grupos dispersos a través de la red.

- Debe poseer la capacidad de moldear la organización para que los datos y modelos transformen su forma de operar. Este paso es básicamente una extensión avanzada de lo anterior, que como hemos dejado claro, aún no es una realidad en los centro del grupo HM.

Sin embargo, la transformación de una organización hacia el dato no depende solamente de su capacidad de generar datos y aplicar el condimento extraído de éstos. Existe una amplia gama de grises que permite indicar en qué estado se encuentra una organización en su proceso de metamorfosis. El modelo DELTTA (*data, enterprise, leadership, targets, technology and analyst*) permite identificar el punto exacto de madurez de una organización a través de su transformación.

Este modelo se basa en la evaluación del estado de desarrollo de seis áreas críticas y necesarias para la transformación digital orientada a datos. Estos factores son datos, organization, liderazgo, objetivos, tecnología y personas

Cada uno de estos factores puede escalar en cinco diferentes fases según sea su desarrollo. Para determinar el nivel de evolución en cada una de estas áreas, debemos evaluar qué criterios cumple dentro de cada una de las fases de las que están compuesta.

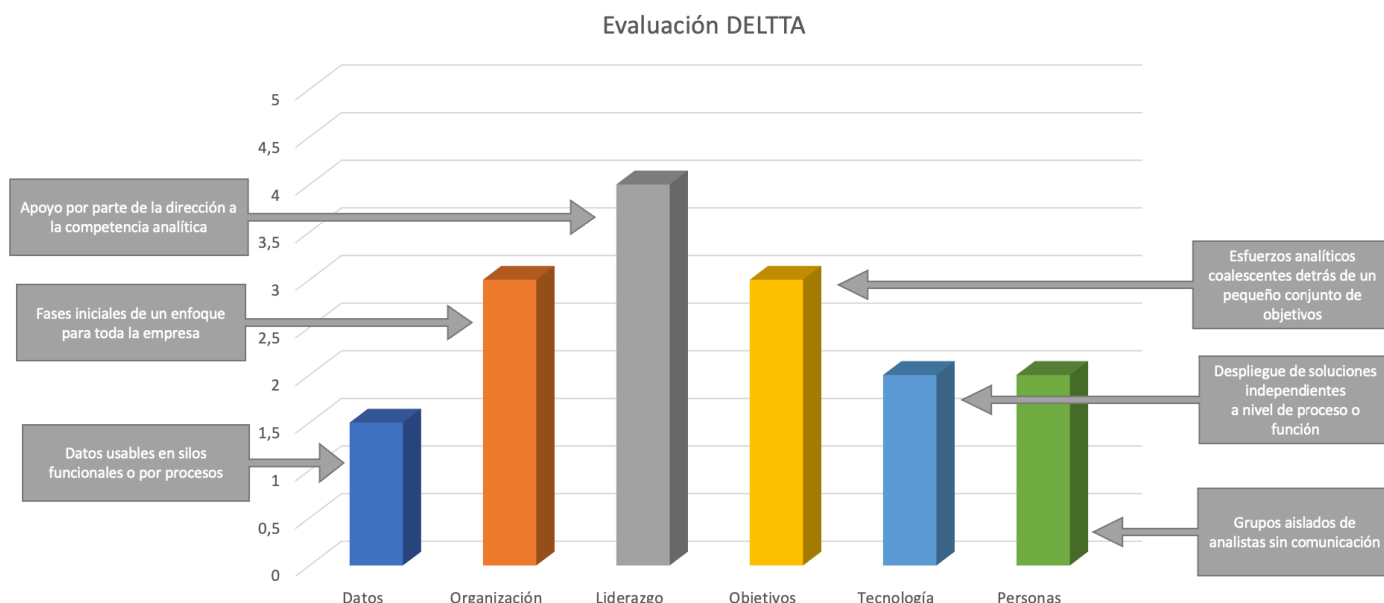
La siguiente tabla resume los requisitos a cumplimentar por área para determinar la fase de desarrollo relacionada a dicha competencia.

Factores de éxito	Fase 1		Fase 2	Fase 3	Fase 4	Fase 5
Datos	Inconsistentes, baja calidad, organización de datos pobre	Datos usables en silos funcionales o por procesos	La organización empieza a crear un repositorio central de datos	Datos integrados y precisos, que son comunes en un <i>data warehouse</i> central	Incesante búsqueda de nuevos datos y métricas	
Organización	No existe/no aplica	Silos de datos, tecnología y conocimiento	Fases iniciales de un enfoque para toda la empresa	Datos, tecnología y analistas clave están centralizados o en red	Todos los recursos son gestionados de modo central o en red	
Liderazgo	No hay conciencia, no interesa	Solo a nivel de función o proceso	Los líderes empiezan a reconocer la importancia de los análisis	Apoyo por parte de la dirección a la competencia analítica	Fuerte liderazgo para la competencia analítica	
Objetivos	No existe/no aplica	Múltiples objetivos inconexos que pueden no ser de importancia estratégica	Esfuerzos analíticos coalescentes detrás de un pequeño conjunto de objetivos	Actividad analítica centrada en torno a ciertas áreas clave	La estrategia y las ventajas competitivas se fundamentan en la analítica	
Tecnología	Uso mayoritario de Excel como herramienta de análisis	Despliegue de soluciones independientes a nivel de proceso o función	Inicio del despliegue de iniciativas a nivel corporativo	Uso coordinado de tecnologías comunes en ciertas áreas clave	Búsqueda continua de nuevas tecnologías que desplazar a nivel de organización	
Personas	Capacidades analíticas limitadas y ligadas a una función específica	Grupos aislados de analistas sin comunicación	Afluencia de analistas en las áreas clave	Analistas altamente capacitados organizados centralmente o en red	Equipo de profesionales de alto calibre y en ampliación continua	

A continuación evaluaremos en que fase se encuentra HM Hospitales para cada una de las áreas mencionadas con el objetivo de determinar cuál es el estado actual de evolución de HM como organización orientada a datos. Es importante mencionar que para llevar a cabo la evaluación se ha asumido que HM sólo cuenta con el juego de datos CLS previamente mencionado y que aún depende de equipos analíticos externos que modelen los datos en su camino hacia el conocimiento.

- **Datos:** Gracias al precedente de CSL, podemos estipular que HM se encuentra **entre la fase 1 y 2** del factor datos. La razón por la que no consideramos que HM cumple con todos los requisitos para satisfacer el nivel 2 es que el grupo sólo cuenta con un *dataset* publicado, a pesar de que éste esté usable en un silo de datos y esté bien estructurado. Sería necesario que HM contase con más de un dataset disponible y organizados por áreas o procesos.
- **Organización:** Aquí sí podemos afirmar que HM se encuentra en una fase avanzada de desarrollo, más concretamente en la **fase 3**. Para llegar a este nivel, HM ha tenido que concretar su iniciativa de enfocar la organización hacia un completo desarrollo del proyecto de digitalización en lo que a datos se refiere. El siguiente paso necesario para que HM pueda moverse a la fase 4 sería la incorporación de tecnologías claves para que sus analistas puedan operar sobre ellos en red.
- **Liderazgo:** Este factor es probablemente el más avanzado de todos, **fase 4**, en el que se reconoce que existe una conciencia global en la organización que desea promover e implementar una actividad orientada a datos, quedando sólo la intervención de un liderazgo fuerte que lleve a cabo un proyecto sólido de análisis y modelaje a partir de los datos.
- **Objetivos:** Aquí podemos aventurarnos en definir el nivel de evolución de HM como en **fase 3**, es decir, que se ha definido una estrategia con una meta clara y se han podido despejar algunos objetivos, a baja escala, sobre qué se desea hacer en un futuro.
- **Tecnología:** Aunque arriesgado, podríamos decantarnos por una **fase 2** dentro de este ámbito ya que hasta el momento sólo disponemos de un conjunto aislado de analistas, sin comunicación centralizada y dispersos alrededor del planeta. Sería necesaria la intervención directa de HM en estas tareas de investigación como coordinador central para operar en fase 3.
- **Personas:** Podríamos definir una **fase 2** bajo este ámbito dado que HM depende de grupos aislados de analistas que trabajan en una única área de concurrido.

El siguiente diagrama resume la relación de HM con su plan de transformación digital orientado a datos bajo el modelo DELTTA.



(1) El Español. *Telefónica y HM Hospitales digitalizan el sector salud* [en línea]. Actualizada: 30 de abril de 2019. [Fecha de consulta: 20 de febrero de 2021]. Disponible en:

[https://www.elespanol.com/invertia/disruptores-innovadores/innovadores/20190430/telefonica-hm-hospitales-digitalizan-sector-salud/393962093\\_0.html](https://www.elespanol.com/invertia/disruptores-innovadores/innovadores/20190430/telefonica-hm-hospitales-digitalizan-sector-salud/393962093_0.html)

(2) Computerworld. *En HM Hospitales hemos vivido una revolución digital a todos los niveles* [en línea]. Actualizada: 9 de diciembre de 2020. [Fecha de consulta: 20 de febrero de 2021]. Disponible en:

<https://www.computerworld.es/entrevistas/en-hm-hospitales-hemos-vivido-una-revolucion-digital-a-todos-los-niveles>

(3) Finanzas.com. *HM Hospitales arranca su plan de transformación digital invirtiendo 20 millones de euros entre 2019 y 2023* [en línea]. Actualizada: 2019. [Fecha de consulta: 20 de febrero de 2021]. Disponible en: <https://opinno.com/es/insights/no-queremos-curar-queremos-llegar-antes-que-la-enfermedad>

(4) **Josep Curto Díaz**, *“Organizaciones orientadas al dato”*. Editorial UOC (págs. 24-26, 34-37, 39-40)

(5) Computerworld. *Juan Abarca, de HM Hospitales, Premio ComputerWorld 2020 al presidente del año* [en línea]. Actualizada: 4 de febrero de 2021. [Fecha de consulta: 20 de febrero de 2021]. Disponible en: <https://www.computerworld.es/entrevistas/juan-abarca-de-hm-hospitales-premio-computerworld-2020-al-presidente-del-ano>