



Universitat
Oberta
de Catalunya

PEC 1

**M2.850 - Fundamentos de la
ciencia de datos**

Curso 2020-2021 Segundo semestre

Alumno: Gustavo Alejandro Álvarez Bea
Estudios: Máster Universitario en Ciencia de Datos

Índice de contenidos

Pregunta 1. Open Data en la salud	3
1.1 Obstáculos y retos	3
1.2 Limitaciones	4
1.3 Beneficios	4
1.4 Impacto durante la pandemia	6
Pregunta 2. Perfiles involucrados	7
2.1 Conocimientos previos y formación.....	7
2.2 Habilidades	8
2.3 Buenas prácticas.....	9
Pregunta 3. Proyecto <i>Covid data save lifes</i>	10
3.1 Ciclo de vida	10
3.1.1 Captura	10
3.1.2 Almacenamiento	11
3.1.3 Preprocesado	11
3.1.4 Análisis.....	12
3.1.5 Visualización	12
3.1.6 Publicación	12
3.2 Pirámide D-I-K-W.....	13
Pregunta 4. Transformación digital.....	14
4.1 Orientación al dato.....	14
4.2 Modelo de madurez	16
Referencias bibliográficas	17

Pregunta 1. Open Data en la salud

1.1 Obstáculos y retos

El primer obstáculo que emerge cuando se combinan salud y *open data*¹ viene dado por la propia **sensibilidad del dato a tratar** y por el desconocimiento ante las técnicas de anonimizado. No hay nada máspreciado que la salud, y todos tenemos un especial celo por nuestra intimidad y por la salvaguarda de nuestro historial clínico. De hecho, los datos personales sobre salud gozan del mayor nivel de protección en la LOPD, RGPD y demás normativa derivada. Evidentemente, cualquier dato clínico compartido, fuera de la exclusiva atención al propio paciente, debe estar desligado de toda identificación personal, con disociación irreversible. No necesariamente debería ser un obstáculo, si se explica lo anterior y se garantiza el debido tratamiento mediante protocolos, procedimientos, evaluación de impacto y auditoría adecuados, sin menoscabo de todas las medidas coercitivas y disuasorias de aplicación, excluyendo aquellos conjuntos de datos para los que no se pueda contar con dichas garantías.

Un segundo obstáculo es ya más general e independiente de la naturaleza del dato, y apunta a **cuestiones culturales**. No se nos educa debidamente para compartir, ni se nos inculca que justamente en eso reside la pirámide del conocimiento y el avance de la sociedad. La manida frase “la información es poder” hace que muchas personas, y las organizaciones que conforman, lo eleven a grado extremo y vean como una pérdida de ventaja competitiva abrir sus datos al público. A este respecto, es notorio que, en la PEC que nos ocupa, nos llame positivamente la atención esta iniciativa de apertura de datos de salud, y que quien la aborde tenga que ser justamente un grupo empresarial privado. Es difícil entender que no sea ya una obligación de la sanidad poner a disposición, cuando menos de la comunidad científica, todos nuestros datos de salud anonimizados. Juan Abarca Cidón, presidente de HM Hospitales y premiado por ComputerWorld, sentenciaba convencido que “el dato sanitario debería considerarse patrimonio universal de la humanidad” (ComputerWorld, 2021).

¹ “Los datos abiertos son aquellos que son libres y cualquier persona los puede extraer y utilizar como quiera, sin derechos de autor, patentes o mecanismos de control” (Mora Cantallops, 2019, p. 43).

Otro obstáculo, que ofrece un reto nada menor, es el **consenso y empleo de estándares**, con el fin de que los datos sean homologables, y puedan ser incorporados a los *datasets* existentes, reutilizados y redistribuidos por la comunidad. Cuando se trabaja en proyectos con equipos de distintas procedencias, uno se da cuenta la facilidad que tenemos los humanos por particionar la realidad de forma diferente, y por llamar de manera distinta a lo que debería ser un mismo concepto. En esta línea, los metadatos y diccionarios técnicos juegan un papel crucial.

1.2 Limitaciones

Los procesos de anonimizado, la combinación de datos que en origen se encuentran en diferentes repositorios, la exportación a formatos abiertos, etc. requieren un esfuerzo en tiempo y recursos con los que no siempre se cuenta, si no existe una política activa de *open data*, con liderazgo ejecutivo. Cuando se consultan repositorios abiertos, es habitual que la periodicidad de actualización no sea la esperada para un sector en constante cambio y evolución. De hecho, suele ser frustrante constatar la **falta de actualización** que tienen muchas veces los últimos datos disponibles, o las lagunas o cortes inexplicables de las series temporales.

El anonimizado puede limitar también la **trazabilidad** de los datos, especialmente cuando se quiere combinar un subconjunto de resultados con otros estudios con los que no existen identificadores de relación. Otra limitación es el excesivo **nivel de agregación** con el que se acostumbra ofrecer la información en abierto, empleando operaciones básicas para presentar variables numéricas agregadas, como son la simple suma o la media aritmética, donde se pierde la posibilidad de establecer rangos o especificidades concretas, en dimensiones de especial interés en el ámbito de la salud: edad, masa corporal, hábitos alimenticios, práctica de deporte, patologías previas, etc.

1.3 Beneficios

Uno de los logros más directos e interesantes del análisis de un gran volumen de datos es la **detección de patrones** que, sin las debidas técnicas analíticas, y dispuestos en conjuntos dispersos, pasarían desapercibidos. Acceder a datos abiertos de otros pacientes, con similares

síntomas o con cualquier otro indicador que suscite correlación, es obviamente de gran interés para la comunidad científica y un gran beneficio a la hora de realizar diagnósticos, solicitar pruebas específicas y prescribir tratamientos.

Al abordar proyectos con ambición global en ciencia de datos, rápidamente también se entiende la necesidad de la **multidimensionalidad**, que requiere disponer de *datasets* muy variados que ofrezcan una visión no sesgada de la realidad. Para esto el *open data* es fundamental, a la vez que se potencian **nuevos proyectos** y estudios, que no se podrían llevar a cabo internamente por limitación de presupuesto, tiempo y recursos, pero que sí pueden abordarse desde terceras partes, que cuentan con los datos que no tenían previamente, fomentando así la **participación**. Y se sustenta, sin duda, en la revolución que supone “la generalización de las wearables y las aplicaciones sobre salud y bienestar [...] la irrupción de herramientas de realidad virtual, inteligencia artificial y técnicas de big data sanitario [...] medicina de precisión o medicina personalizada” (Andreu Martínez, 2017).

Por otro lado, los datos abiertos **mejoran a los propios datos** en origen, por variadas razones. En primer lugar, porque están al alcance de muchas más personas, con visiones complementarias, e incluso críticas con la entidad que los ha producido. Esto sustenta sinergias interesantes, con visiones mucho más amplias de las que se podría contar internamente. Por otro lado, al saber que los datos se publicarán en abierto, se imprime una responsabilidad extra hacia el productor de la información, que sabe que sus datos serán inspeccionados desde el exterior.

Se favorece, a su vez, la **transparencia y toma de decisiones**, tan demandada en la sociedad. Y los empleados de una organización, que saben que sus datos serán públicos, se sienten más partícipes con los proyectos, ya que compartir datos se convierte en una señal de identidad y un escaparate hacia la sociedad. A este respecto, es muy apropiada la visión que se da en los materiales del dato como “bien común y [que] debe ser compartido” (Curto Díaz, 2017, p. 22), en línea con el demandado **empoderamiento del paciente**, consciente de la realidad sanitaria, responsable de su correcto uso y beneficiario de la buena *praxis*. Experiencias de personas con enfermedades incurables y diagnóstico tardío, como la que se traslada en el blog de *Data Save Lives*, ofreciéndose como *data donor* y solicitando

comportamientos análogos del resto de la sociedad, demuestran la necesidad de disponer de conjuntos de datos fiables en abierto. Es preciso contar con la seguridad que aporta el conocimiento y el debido contexto para alcanzar decisiones informadas, y mucho más si a enfermedades crónicas se les suma una pandemia como la actual (Data Saves Lives, 2020).

Y este tipo de decisiones –que necesariamente deben adoptarse desde las capas organizativas más altas– facilita incluso que los propios trabajadores puedan enseñar sus logros en foros sectoriales o de difusión pública, sin temor a infringir clausulados restrictivos de confidencialidad, con las salvaguardas debidas en cuanto a información de carácter exclusivamente interno. Lo anterior no es un detalle menor, ya que comienza a convertirse en un mantra la frase de que ‘no se consiguen científicos de datos’. Estoy seguro que a varios de nosotros nos ha cambiado la visión que teníamos de HM Hospitales, si es que acaso conocíamos al grupo empresarial. El *open data* puede ser una **fuerza de atracción** que consiga que buenos profesionales de otros sectores quieran ser parte de un proyecto, por compartir la filosofía de que los resultados finales de su trabajo puedan revertir, de una u otra manera, en el bien de la sociedad y en el avance del conocimiento. Y consolida, sin duda, la **marca de empresa**, favoreciendo un sentimiento positivo hacia la organización, siempre tan buscado en las campañas de marketing y en redes sociales.

Acceder a *datasets* de calidad, finalmente, favorece la **reducción de costes**, porque la comunidad se beneficia a coste cero de un conjunto de datos vital, en ocasiones para muchos totalmente impensable de generar. Y esta filosofía se retroalimenta, ya que esa disponibilidad de datos posibilita y justifica el acopio de otros datos complementarios, que revierten en el ecosistema abierto.

1.4 Impacto durante la pandemia

Considero que sería difícil encontrar a alguna persona mínimamente informada que no hubiese consultado en todos estos meses el mapa de evolución del virus por países, por comunidades autónomas, por municipios de su provincia. Incluso surgieron iniciativas particulares que, a partir de sencillas APIs, ofrecían gráficos y visualizaciones interactivas, actualizadas casi al instante. Todo esto ha sido gracias a contar con **datos demográficos en abierto**. Y lo anterior enfatiza una realidad evidente, pero en la que quizá no hemos

reflexionado aún lo suficiente: sin datos, no habría científicos de datos. Y, al definir las funciones de un científico de datos, me parece especialmente interesante una que se menciona en los materiales y que –a mi modo de ver– condensa muy bien una de las finalidades de su actividad: “crear historias a partir de datos” (Mora Cantallops, 2019, p. 19). Esto es justamente lo que recibimos durante meses de pandemia de un tema completamente desconocido para nosotros, y que continuamos consumiendo a diario, gracias a contar con datos abiertos.

En una esfera más científica, por su parte, se han pulverizado los tiempos medios de **fabricación de la vacuna**. Nuevamente se ha conseguido porque, como nunca antes, los distintos equipos científicos contaban con datos a gran escala, fruto de investigaciones compartidas, con experiencia previa acumulada en generación de vacunas que simulaban una infección y con prometedores avances en nuevas aproximaciones, como el ARN mensajero, en las que se basan las vacunas Pfizer/BioNTech y Moderna (El País, 2020). No han sido pocas las voces que han reclamado una mayor disposición colaborativa, incluso que se dejaran a un lado las patentes con el fin de salvar vidas cuanto antes y a la mayor escala posible. Esto último tal vez pecaba de cierta ingenuidad, incluso de serias dificultades para el requerido acopio económico, dentro del libre mercado y la libre competencia, que de una u otra manera alcanza también al sector de la salud.

Pregunta 2. Perfiles involucrados

2.1 Conocimientos previos y formación

Es habitual definir al científico de datos como *rara avis* o unicornio, dado el dominio en disciplinas, tan completas y exigentes en sí mismas, que se le exige: matemáticas y estadística, TI y programación, negocio específico de aplicación, comunicación (Mora Cantallops, 2019, pp. 15, 16). En febrero del 2020 se incorporó Alberto Estirado Bronchalo como director de Sistemas de Información y Transformación Digital de HM Hospitales, en un puesto homologable al CIO (*Chief Information Officer*) de la organización. Es interesante la coetilla de *Transformación Digital*, porque pone el foco en una de sus funciones primordiales y sobre la que se sustenta una inversión de 20 millones de euros: el desarrollo de un ambicioso plan

estratégico de “revolución digital en todos los niveles”. El CIO, sin duda, ha tenido que tener conocimientos de negocio, para liderar la implantación de un sistema de consultas telemáticas vía VPN para un grueso de 6.000 doctores y la “adaptación digital de todos los procesos asistenciales”. Ha requerido formación en ciberseguridad, con el fin de dotar de un ecosistema seguro a la organización, y superar con éxito los numerosos ataques en el ámbito de la salud, incrementados por el efecto pandemia. Y, además de su formación universitaria en Informática y Ciencias Empresariales, dispone de un posgrado en Ciencias del Dato del MIT que, sin duda, ha sido de utilidad en la consecución del proyecto *Covid data save lives* (ComputerWorld, 2020).

El grupo empresarial cuenta con un equipo de TI, conformado por 42 personas, además del apoyo externo derivado de acuerdos con *partners* tecnológicos, como Telefónica, la consultora TRC, Microsoft o Palo Alto. Disponen, asimismo, de un Departamento específico de I+D+i. Con esta amalgama, es de esperar que entre los perfiles involucrados se encuentren científicos de datos, especialistas en visualización, ingenieros del dato y arquitectos *cloud computing*.

2.2 Habilidades

Al abordar la esencia de las organizaciones orientadas al dato, en los materiales de la asignatura se menciona continuamente la necesidad de tener un pensamiento analítico, **cultura analítica**, incluso se reclama que el empleado sea un “trabajador analítico” (Curto Díaz, 2017, p. 38). Esta es una habilidad clave que debe tener y aplicar en su día a día todo perfil involucrado en la transformación digital y, sin duda también, en el proyecto transformador de HM Hospitales.

Es preciso tener visión estratégica para identificar problemas, descubrir oportunidades y decidir en base a datos y hechos, sin perder de vista, incluso, que “el científico de datos construye nuevos modelos y busca conocimiento sobre indicadores que la compañía todavía no sabe que son importantes” (Mora Cantallops, 2019, p. 19). Y, por supuesto, debe fomentarse una relación colaborativa, sabiendo que el fin último no es aplicar un modelo, sino mejorar el negocio. A este respecto, una habilidad imprescindible es no perder la **curiosidad** por aprender, y aprovechar el espacio de trabajo para adquirir el debido lenguaje de negocio, para **saber entenderse** con los perfiles del resto de áreas no propiamente tecnológicas.

2.3 Buenas prácticas

Un **proyecto es finito**, está acotado en el tiempo y dispone de una asignación de recursos concreta. Un proyecto no puede concebirse sin fechas ni entregas. Parece una obviedad, pero acostumbra ser el caballo de batalla de todo jefe de proyecto. Una buena práctica obliga a una planificación realista y a un cumplimiento acorde a la palabra dada. A este respecto, me ha parecido muy interesante la mención a “establecer un modo ordenado y estructurado de pedir información” (Curto Díaz, 2017, p. 40). Los proyectos ambiciosos y de gran alcance suelen implicar muchas iteraciones e ineludiblemente obligan a recurrir a personal de gestión, que tiene que compatibilizar las nuevas líneas de negocio con el día a día, con mantener a flote lo que ya funciona. No se puede estar requiriendo datos a todas horas ni por entregas, debe estar pautado y deben respetarse los cauces y los tiempos.

No dejarse deslumbrar por los *hypes* del momento, sin perder el entusiasmo por el lado disruptivo de la tecnología, es también vital en todo perfil transformador. Es fundamental prestar atención a la **implantación**, sabiendo que una cosa es que exista tecnología con un nivel de madurez suficiente y que funcione, y otra muy distinta es que esta tecnología se integre en la organización, para ser utilizada por personas que no somos nosotros y que se utilice como nosotros esperamos. En este sentido, nunca debe perderse de vista el *talento oculto* de los usuarios finales para no dejar de sorprendernos.

Pensar y fomentar la **multidisciplinariedad** permite contar con distintas perspectivas, hallar aproximaciones complementarias, ver un problema desde otra óptica y, tal vez, formularse “las preguntas adecuadas” y con ellas, las soluciones que –de otra manera– posiblemente no hubiésemos alcanzado (Curto Díaz, 2017, p. 28).

La **formación** a usuarios está siempre presente en todo diagrama de Gantt. Pero formar a usuarios no es dar varias charlas y entregar documentos PDF. La formación debe ser dirigida, motivada, consensuada, adaptada, inmersiva y pensada para los distintos perfiles que conforman la organización. A este respecto, y muy especialmente en el ámbito de la transformación digital y la IA, debe prestarse atención a la **explicabilidad** de los modelos. Los resultados alcanzados tienen que ser explicables, asimilables por el razonamiento humano, con independencia de la complejidad del problema a tratar, los algoritmos empleados y las

múltiples variables y pesos asociados. Existen casos recientes en el ámbito de la salud que apuntan justamente en la dirección contraria, y que suscitan serias dudas en la comunidad científica².

Comunicar es pensar también en la **comunicación interna**. Cuesta mucho recuperar la confianza de responsables que conocen cuestiones de negocio de su competencia en foros no oficiales. Pensar en *stakeholders*, ‘a quién le afecta’, empieza por pensar dentro de la propia casa común, que es la organización empresarial. La pregunta es muy pertinente: “¿cómo es que las organizaciones siguen teniendo tantos problemas en la implantación de este tipo de proyectos?” (Curto Díaz, 2017, p. 5). Y junto a la comunicación, **paciencia y empatía** son dos cualidades fundamentales en todo proyecto en el que intervenga un equipo humano, procurando encontrar siempre el camino de menor resistencia para llegar al lugar previsto.

Pregunta 3. Proyecto *Covid data save lifes*

Covid data save lifes pone a disposición de la comunidad científica “un dataset anonimizado con información clínica de los pacientes atendidos por la pandemia en HM Hospitales en Madrid, Barcelona y Galicia”. Dispone de más de 3.000 historias clínicas completas de pacientes tratados por el virus SARS-CoV-2 (La Razón, 2020). Es un *dataset* clínico único que “recoge las distintas interacciones en el proceso de tratamiento del COVID-19, incluyendo información pormenorizada sobre diagnósticos, tratamientos, ingresos, pasos por UCI, pruebas diagnósticas por imagen, resultados de laboratorio, alta o deceso, entre otros muchos registros” (HM Hospitales, 2020).

3.1 Ciclo de vida

3.1.1 Captura

En *Covid data save lifes* la creación de los *datasets* está integrada en el propio proceso de generación de los datos. Los datos son propios, generados en el lado del cliente. No es

² ‘En Estados Unidos una IA analizó dos millones de electrocardiogramas y diseñó una red neuronal que es capaz de predecir mejor que cualquier cardiólogo el riesgo de muerte en un horizonte temporal de un año. Qué hizo y qué descubrió la máquina para ser capaz de hacer eso aún no se sabe’ (La Voz de Galicia, 2021).

necesario aplicar técnicas de extracción a medida que van apareciendo en fuentes externas. Este es el escenario más eficiente por definición, porque desde la propia concepción del dato ya se dispone de un mecanismo para capturarlo y consolidarlo adecuadamente. A su vez, se cuenta con indicadores precisos de calidad, ya que se conoce exactamente cómo se han recogido los registros, qué mecanismos de detección de errores se han desplegado y que variabilidad pueden presentar.

3.1.2 Almacenamiento

Dada la amplitud de fuentes de información comentadas arriba, es de esperar que el almacén de datos cuente con formatos diversos de almacenamiento, en función de si los datos son introducidos por teclado, lectores QR o de código de barras, etc., o si su origen son sensores biométricos, pruebas diagnósticas de imagen, entre las que se encuentra el PACS [sistema de almacenamiento y archivo de imágenes médicas], etc. A este respecto, será necesario documentar técnicamente los formatos propietarios de almacenamiento, sobre los que suelen apalancarse la maquinaria clínica y las tecnologías muy verticales, con el fin de disponer de formatos de intercambio abiertos, que posibiliten la integración con el resto de repositorios.

3.1.3 Preprocesado

La organización dispone de un CRM en modelo pasivo, que buscan integrar en el *core* del servicio integral al paciente. En esta línea radica también el eje del denominado “dato único”, con el fin de vincular toda la información relativa a una misma persona, basándose en la estandarización de sistemas.

Atendiendo al filtrado de variables y a la selección de datos de interés, se abre un poderoso dilema que seguramente ha sido resuelto acudiendo a más espacio de almacenamiento y capacidad de cálculo. Entre las variables a tener presente estarán, sin duda, las habituales: edad, masa corporal, deficiencias respiratorias, patologías severas, etc. Pero, ante un virus desconocido, letal en algunos casos y con una asombrosa capacidad de contagio, sería muy difícil prescindir de otros indicadores que, *a priori*, podrían no estar relacionados pero que, en fases posteriores de descubrimiento, resultaran de interés.

Previo al análisis masivo, debería practicarse y auditarse el anonimizado de los datos. La razón es evidente: tendrá acceso a esta información personal distinto al cuadro médico que trata al paciente, y las motivaciones para el tratamiento de estos datos ya no serán las propias de la práctica clínica ordinaria. Por tanto, como paso previo al análisis, debería asegurarse que no existe ninguna marca personal que identifique, o permita identificar, a una persona física.

3.1.4 Análisis

En las entrevistas concedidas y en la documentación ofrecida en relación al *dataset* final, se nos da el dato de N : 3.000 pacientes. Sabemos también que, como punto de partida, se recomienda que N sea, al menos, “de un orden de magnitud que d ”, que será cada una de las observaciones asociadas a cada paciente (es decir, $N > 10 \times d$) (Minguillón, 2016, pp. 27, 28). Por tanto, con todas las salvedades posibles, y en un plano teórico inicial –asumiendo también que la proporción es aconsejable que sea incluso mayor–, con el volumen del *dataset* podría justificarse una suficiencia de muestreo que contemplara hasta 300 dimensiones diferentes.

3.1.5 Visualización

El CIO menciona en la entrevista que pretenden “bajar una capa la inteligencia de negocio”, con el fin de alcanzar el escalón de los supervisores (ComputerWorld, 2020). Esto pasa, sin duda, por mejorar la visualización de los datos y comunicar mejor. Si admitimos la cosificación, “los humanos son, principalmente, máquinas de procesado visual”³ (Minguillón, 2016, p. 30).

3.1.6 Publicación

Se distinguen dos claras áreas en esta última fase del ciclo de vida, como producto final accionable: la publicación de datos de consumo interno, para gestionar y posibilitar la dinámica del propio negocio y, en una segunda vertiente, la difusión de datos en abierto. Como se ha indicado arriba, se ha puesto a disposición de la comunidad médica y científica un *dataset* anonimizado, pionero en todo el mundo, que ha propiciado la visualización en el mapa mundial del grupo hospitalario, y que ha facilitado cientos de investigaciones en búsqueda de tratamientos contra el COVID-19.

³ A este respecto, realizaría un apunte -muy a pie de página, pero a mi entender necesario. Para visualizar ideas como “revolución digital”, “hospital líquido”, “telemático”, etc., no resulta del todo conveniente que el CIO elija, como fondo de la vídeo-entrevista, una pizarra de rotuladores a medio borrar, propia de tiempos pretéritos (ComputerWorld, 2020).

3.2 Pirámide D-I-K-W

En el nivel más alto de la pirámide D-I-K-W (*data, information, knowledge y wisdom*), la sabiduría se define como ‘la habilidad para aplicar el conocimiento en beneficio propio’ o, en este caso, del propio proyecto. Es la aplicación del conocimiento, extraído de la información y definida esta última a partir de los datos, lo que permite “alcanzar la sabiduría” (Minguillón, 2016, pp. 6, 11). Y *Covid data save lifes* alcanza este nivel por su propia concepción y esencia.

HM Hospitales, con la debida planificación de un esmerado proyecto, recopila infinidad de datos provenientes de los pacientes con COVID-19 que atiende. Estos datos se traducen inmediatamente en información, al estar dotados del mayor contexto posible y adquirir significado concreto, asociados a cada paciente tratado y medidos con unidades precisas, que responden a preguntas concretas sobre datos propios de ingreso, constantes de seguimiento y constantes vitales, medicación, resultados de laboratorio, diagnósticos, procedimientos, etc.

La información adquirida se convierte en conocimiento al desencadenar inmediatamente acciones concretas en favor del paciente, en aras de su estabilidad y pronta recuperación, fin primigenio de todo centro hospitalario al que uno acude con una dolencia. Y en el siguiente y último escalón, la aplicación de este conocimiento, gracias al análisis masivo de datos, la búsqueda de patrones, la correlación con variables de interés y, sin duda, la apertura del *dataset* a la comunidad médica y científica de todo el mundo, permite el “grado más alto del conocimiento” y alcanzar un “conocimiento profundo” en la materia en cuestión, tal y como define la RAE a la “sabiduría” en su primera y tercera acepción (Real Academia Española, 2020). En base al *dataset* compartido se pueden obtener modelos predictivos de la evolución del COVID-19, modelos epidemiológicos, respuesta a tratamientos, parámetros sociodemográficos (HM Hospitales, 2020).

Pregunta 4. Transformación digital

4.1 Orientación al dato

La transformación que debe realizar cada organización para orientarse al dato es “un proceso único” (Curto Díaz, 2017, p. 16). Lo avala sin duda la experiencia singular de HM Hospitales y su encomiable situación actual:

- Nivel de **competencias analíticas**: acercamiento al análisis masivo de información genética, reconociendo precios competitivos.
- Dominio de la **gestión de la información**: integración del sustrato genético del paciente, la exposición medioambiental y su estilo de vida. Implantación del modelo de “hospital líquido” donde lo importante es el paciente y el servicio, videoconsulta y adaptación digital de todos los procesos asistenciales. HIS [sistema de información hospitalario] único.
- **Cultura orientada al dato**: el grupo hospitalario cuenta con departamento propio de I+D+i y se constata la implicación del área directiva al punto que se crea una nueva posición, el de CIO en la organización. El puesto pertenece a la junta directiva y al consejo de administración, palancas clave para abordar el cambio cultural que conlleva la orientación al dato.

(ComputerWorld, 2020).

“Una organización orientada al dato es aquella que se gestiona fundamentada en hechos y datos” (Curto Díaz, 2017, p. 15). En mi opinión, y con la información disponible, entiendo que HM Hospitales es una organización con un **proceso avanzado de transformación** hacia la orientación al dato. Es loable el plan de inversión en esta línea y el liderazgo desde las capas organizativas más altas, notablemente en la figura de su presidente, Juan Abarca Cidón. En la entrevista concedida a Opinno, el directivo concibe la innovación como una aliada vital, y recuerda que su grupo hospitalario ha nacido y crecido en la era de la transformación digital (cuentan con poco más de 25 años en el mercado), lo que les ha permitido incorporarla en su ADN fundacional. “Hemos apostado desde nuestros inicios por tecnología disruptiva”, sentencia, y se propone objetivos que nos trasladan incluso una visión futurista, con atisbos de ciencia ficción: “queremos llegar antes de que la enfermedad aparezca. No queremos curar, simplemente no queremos que enfermen”. La frase ha servido, de hecho, para crear el titular de la entrevista.

El presidente de HM Hospitales es licenciado en Medicina y licenciado en Derecho, y doctor en Medicina y Cirugía⁴. En la conversación se constata su apuesta por la innovación, como quedó indicado arriba, pero también se intuye claramente su lógica defensa por la formación y profesión médica, lo que marca ineludiblemente el orden de prioridades en la aproximación y fijación de estrategia: “En ocasiones [...] basta una simple regla de tres [...] nuestro enfoque se basa en el sentido común⁵ sobre el que se hacen las preguntas y en el racional de la búsqueda de la respuesta [...] La irrupción de la tecnología de la información es un proceso que lleva años, pero ahora se habla más de ello [...] no deja de ser conceptualmente una transformación de la organización muy pequeña, dado que el núcleo del negocio sigue siendo el mismo”. Me ha parecido muy oportuna la entrevista porque, desde mi óptica subjetiva, he tenido la sensación de que el entrevistador no quería abandonar la tecnología, mientras el entrevistado se esforzaba en recalcar que, obviamente, sólo con tecnología no es suficiente (Oppino, 2019).

En esta línea, entiendo que es pertinente indicar también que la creación y cobertura del puesto de CIO en la organización es de hace tan sólo poco más de un año. Es de esperar, por tanto, que los distintos niveles que posibilitan la orientación al dato **requieran aún de cierto recorrido** en sus respectivas fases de madurez. Es llamativo también que el proceso de elección del *partner* tecnológico, para “potenciar e impulsar el uso de nuevas tecnologías dentro de los procesos asistenciales y de relación con pacientes”, y la posterior firma del acuerdo de colaboración finalmente con Telefónica, en abril de 2019 (El Español, 2019), fueran previos a la elección del CIO en la organización, realizada varios meses después, en febrero de 2020⁶ (CIO España, 2020).

Como fuerza tractora, el **entorno competitivo**, en el que se desarrolla la actividad de la organización, favorece y reclama la transformación digital iniciada. “Junto con el marketing y

⁴ CV disponible en: <https://es.linkedin.com/in/juan-abarca-cidon-b7b72122>

⁵ Se entiende el necesario término “sentido común” pero, en mi opinión, una organización orientada al dato debe vigilar muy de cerca expresiones de este tipo que, en ningún caso, pueden servir de pretexto para obviar la necesaria reflexión en torno al dato. La premisa es ineludible: “análisis basado en datos en lugar de la intuición” (Curto Díaz, 2017, p. 13).

⁶ En la entrevista concedida a ComputerWorld, Alberto Estirado comenta que, en realidad, siempre había trabajado para HM Hospitales, pero desde fuera; en su última etapa, como CEO de la consultora TRC. El nivel de crecimiento experimentado desencadenó en el grupo hospitalario la necesidad de “un directivo de alto perfil tecnológico” (ComputerWorld, 2020).

las finanzas, el sector de la salud es probablemente el que más uso de la ciencia de datos está haciendo hoy en día” (Mora Cantallops, 2019, p. 28).

4.2 Modelo de madurez

Siguiendo el modelo de madurez de Davenport, Harris y Morison, se referencian las distintas situaciones en las que se encuentra HM Hospitales, procurando encuadrar cada una de ellas en la fase de madurez correspondiente, según la matriz DELTTA. Para este análisis, será pertinente diferenciar previsión, planificación, declaración de intenciones publicadas en prensa generalista y especializada, etc., de los resultados actualmente alcanzados, que es el factor objetivable a medir. El punto de partida, en todo caso y como se puede apreciar a simple vista, es alentador. Y la pandemia, muy especialmente en el sector de la salud, ha acelerado notablemente la adopción de políticas activas centradas en el dato.

Factores de éxito	Fase 1	Fase 2	Fase 3	Fase 4	Fase 5
Datos				HIS [sistema de información hospitalario] único. Datos genéticos + ambiental + estilo de vida	
Organización			Análisis de eficiencia		
Liderazgo				Apuesta directiva. Inversión de 20M€	
Objetivos				Adaptación digital de todos los procesos asistenciales	
Tecnología				Modelo de hospital líquido	Omnicanalidad. Cloud híbrido
Personas				Cobertura CIO	Departamento I+D+i

Se aprecia un modelo de madurez avanzado, en el entorno de la denominada fase 4, situando a HM Hospitales como una compañía analítica, que se dirige hacia una estrategia de transformación digital y orientación al dato coordinada.

Referencias bibliográficas

Andreu Martínez, María Belén (2017). "Open Data en el ámbito sanitario y su compatibilidad con la privacidad del paciente" [en línea]. *IMODEV*. París: Institut du Monde et du Développement pour la Bonne Gouvernance Publique. [Fecha de consulta: 19 de marzo de 2021]. Disponible en: <https://ojs.imodev.org/index.php/RIGO/article/view/200/330>

CIO España (2020). "HM Hospitales elige a Alberto Estirado como director de Sistemas de Información y Transformación Digital". *CIO España*. Madrid: IDG Communications. Actualizada: 4 de febrero de 2021. [Fecha de consulta: 13 de marzo de 2021]. Disponible en: <https://www.ciospain.es/sanidad/hm-hospitales-elige-a-alberto-estirado-como-director-de-sistemas-de-informacion-y-transformacion-digital>

ComputerWorld (2021). "Juan Abarca, de HM Hospitales, Premio ComputerWorld 2020 al presidente del año" [en línea]. *ComputerWorld*. Madrid: IDG Communications. Actualizada: 4 de febrero de 2021. [Fecha de consulta: 13 de marzo de 2021]. Disponible en: <https://www.computerworld.es/entrevistas/juan-abarca-de-hm-hospitales-premio-computerworld-2020-al-presidente-del-ano>

ComputerWorld (2020). "En HM Hospitales hemos vivido una revolución digital a todos los niveles" [en línea]. *ComputerWorld*. Madrid: IDG Communications. Actualizada: 9 de diciembre de 2020. [Fecha de consulta: 14 de marzo de 2021]. Disponible en: <https://www.computerworld.es/entrevistas/en-hm-hospitales-hemos-vivido-una-revolucion-digital-a-todos-los-niveles>

Curto Díaz, Josep (2017). *Organizaciones orientadas al dato*. Barcelona: Universitat Oberta de Catalunya.

Data Saves Lives (2020). "The future of healthcare: is it all about data?" [en línea]. *Data Saves Lives*. Actualizada: 15 de octubre de 2020. [Fecha de consulta: 20 de marzo de 2021]. Disponible en: <https://datasaveslives.eu/blog/isitallaboutdata15102020>

El Español (2019). “Telefónica y HM Hospitales digitalizan el sector salud” [en línea]. *El Español*. Madrid: El León de El Español Publicaciones. Actualizada: 30 de abril de 2019. [Fecha de consulta: 13 de marzo de 2021]. Disponible en:

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

El País (2020). “ARN, la molécula que puede sacarnos de esta pandemia” [en línea]. *El País*. Madrid: Ediciones El País. Actualizada: 29 de noviembre de 2020. [Fecha de consulta: 15 de marzo de 2021]. Disponible en: <https://elpais.com/ciencia/2020-11-28/arn-la-molecula-que-puede-sacarnos-de-esta-pandemia.html>

HM Hospitales (2020). “Covid data save lives” [en línea]. *HM Hospitales*. Madrid: HM Hospitales 1989. Actualizada: Abril de 2020. [Fecha de consulta: 13 de marzo de 2021]. Disponible en: <https://www.hmhospitales.com/coronavirus/covid-data-save-lives>

La Razón (2020). “HM Hospitales aúna investigación y Big Data para cercar a la Covid-19” [en línea]. *La Razón*. Madrid: Lagarsa. Actualizada: 31 de diciembre de 2020. [Fecha de consulta: 13 de marzo de 2021]. Disponible en: <https://www.larazon.es/salud/20201231/yzm7vnrlzgrbpvc4f76fbak7q.html>

La Voz de Galicia (2021). “El Gobierno elige Galicia para presentar el Plan Nacional de Inteligencia Artificial” [en línea]. *La Voz de Galicia*. A Coruña: La Voz de Galicia. Actualizada: 18 de febrero de 2021. [Fecha de consulta: 10 de marzo de 2021] Disponible en: <https://www.lavozdeg Galicia.es/noticia/sociedad/2021/02/17/gobierno-elige-galicia-presentar-plan-nacional-inteligencia-artificial/00031613583116917917880.htm>

Minguillón, Julià (2016). *Fundamentos de data science*. Barcelona: Universitat Oberta de Catalunya.

Mora Cantallops, Marçal (2019). *Los roles, ámbitos y nombres de la ciencia de datos*. Barcelona: Universitat Oberta de Catalunya.



Oppino (2019). “No queremos curar, queremos llegar antes que la enfermedad” [en línea]. Madrid: Entrepreneur Capital. Actualizada: 28 de octubre de 2019. [Fecha de consulta: 19 de marzo de 2021]. Disponible en: <https://opinno.com/es/insights/no-queremos-curar-queremos-llegar-antes-que-la-enfermedad>

Real Academia Española (2020). *Diccionario de la lengua española* [en línea]. DLE 23.4. Madrid: Real Academia Española. [Fecha de consulta: 20 de marzo de 2021].
Disponible en: <https://dle.rae.es/sabidur%C3%ADa?m=form>