

**Máster universitario de Ciencia de Datos**

**Prueba de Evaluación Continua – PEC2**

**Fundamentos de la Ciencia de Datos – Bloque 2 Los peligros de no gobernar los datos: calidad, seguridad y ética**

Autor:

Mario Ubierna San Mamés

|  |
| --- |
| Índice de Contenido |

[Índice de Contenido 3](#_Toc69999316)

[Índice de ilustraciones 4](#_Toc69999317)

[1. Enunciado 5](#_Toc69999318)

[1.1. Pregunta 1 5](#_Toc69999319)

[1.1.1. ¿Cómo divulgarías internamente qué es el gobierno del dato en una HM Hospitales y su importancia en la transformación digital? 5](#_Toc69999320)

[2. Bibliografía 6](#_Toc69999321)

|  |
| --- |
| Índice de ilustraciones |

|  |
| --- |
| Enunciado |

## Pregunta 1

*En la transformación digital de HM Hospitales, además de los camios tecnológicos, se dieron fundamentalmente cambios culturales. Entre ellos, el gobierno del dato seguramente fue decisivo. Sitúate en el rol del CDO y a partir de ahí, imagina y describe:*

### ¿Cómo divulgarías internamente qué es el gobierno del dato en una HM Hospitales y su importancia en la transformación digital?

Antes de empezar a divulgar qué es el gobierno del dato indicaría el contexto en el que estamos, cabe destacar que este concepto ha ido tomando forma a lo largo de los años debido a la gran importancia que ha tenido y tiene.

Desde el comienzo de la informática siempre se ha buscado automatizar procesos y mejorar los procesos de negocio, inicialmente la capacidad de cómputo que se tenía era muy baja por lo que no se podía hacer un gran uso de los datos, sin embargo a medida que fuimos avanzando fue mejorando las tecnologías, y en consecuencia las organizaciones se dieron cuenta de que los datos revelaban mucha información.

Centrándonos en el tema de estudio de *HM Hospitales*, tal y como recordamos de la anterior práctica, su principal objetivo era que la organización se transformara digitalmente con el fin de mejorar los procesos de negocio, en este caso la mejora del servicio y automatizar tareas.

Este transformación digital se lleva a cabo solamente cuando está centrada en los datos, cuanto más datos tengamos y éstos sean de mayor calidad, mejores servicios vamos a poder dar, es aquí donde entra en juego el gobierno del dato.

Entendemos por gobierno del dato a todos aquellos procesos que se encargan de gestionar el dato en sí: el cómo/cuándo/dónde crearlos, cómo/cuándo/dónde gestionarlos, cómo/cuándo/dónde se accede a ellos, cómo/cuándo/dónde se usan, cómo/cuándo/dónde se destruyen… [1]

Tal y como vemos el dato hoy en día es de una gran importancia y toda trasformación digital requiere de datos, pero no de cualquier tipo de dato, sino que se busca la mayor calidad tanto del dato como en la gestión del mismo, es por ello que el gobierno del dato es de vital importancia para alcanzar la transformación digital en una organización.

### Si te basaras en DAMA, ¿qué objetivos principales perseguirías sobre los datos?

Lo primero que debemos tener claro es qué es el gobierno del dato según DAMA, ellos entienden como gobierno del dato el ejercicio de autoridad, control y toma de decisiones compartida sobre la gestión de los activos de datos [1].

El gobierno del datos tiene que ser capaz de: definir y validar las estrategias de los datos (procedimientos, normas, arquitectura…), monitorizar los datos (a nivel procedimental, normativo, arquitectónico…), dar valor al dato y saber gestionarlo...

Como podemos observar el gobierno del dato tiene diferentes objetivos con el fin de mejorar la toma de decisiones de una organización, y éstos solo se cumplen si se alcanzan los siguientes sobre los datos:

* El dato tiene que ser accesible, es decir, siempre que alguien quiera acceder al dato tiene que poder hacerlo, y dicho dato debe tener un formato que se corresponda con lo que el usuario quiere. Por ejemplo, en *HM Hospitales* diferentes médicos de un mismo hospital y cuya rama sea la misma tienen que poder acceder a la misma información.
* El dato tiene que ser seguro, con ello nos referimos a que no todo el mundo tiene que tener acceso al dato, solo aquellas personas autorizadas. En *HM Hospitales* nos tendríamos que asegurar que un enfermero no pueda tener el mismo nivel de acceso que un médico.
* El dato debe de ser consistente, es decir, al igual que debe tener un formato que se adapte al usuario, también si diferentes personas quieren acceder al mismo dato la información que se les proporcione tiene que ser la misma. Por lo tanto, diferentes médicos si quieren acceder a la misma información se les tiene que proporcionar el mismo resultado.
* El dato debe ser de una alta calidad, como bien hemos mencionado la calidad del dato es fundamental en la transformación digital, por lo que tenemos que garantizarnos que el dato es certero y para ello podemos hacer uso de estándares y normativas.
* Al dato tiene que ser auditable, es decir, hay que tener mecanismos para seguir la trazabilidad del dato y así ver cómo evoluciona. En *HM Hospitales* tendríamos que ser capaz de seguir la evolución de un paciente.

### ¿Cuál sería tu planteamiento para la consecución de estos objetivos, en función de las fases que tendrías que implementar para el gobierno del dato?

Para poder alcanzar estos objetivos trataría de seguir el ciclo de vida del gobierno del dato según *The Data Governance Institute* [1], cabe destacar que no se tiene que seguir cada una de las fases, pero para poder dar una explicación más general vamos a hacer uso de cada una de ellas aplicadas a *HM Hospitales*:

Lo primero que debemos realizar es definir qué ganamos si implementamos un gobierno del dato, es decir, qué valor nos da hacer uso de los datos para mejorar la toma de decisiones dentro de una organización. Una vez que tenemos claro este punto, debemos de tomar medidas para asegurarnos que vamos por el camino correcto. Respecto a *HM Hospitales*, nos preguntaríamos para qué nos sirven los datos y qué ganamos con ellos, y podríamos responder que conseguimos mejorar los servicios de los pacientes y tomar mejores decisiones. Por lo tanto, definiríamos una serie de métricas financieras para determinar cuál es el coste si hacemos un mal uso de los datos respecto a si implementamos un gobierno del dato.

Cuando ya tenemos claro lo que conseguimos al usar un gobierno del dato y cuál es su coste, al igual que sucede con cualquier tipo de proyecto es preparar la hoja de ruta, es decir, planificar los detalles de cada una de las fases que vamos a tener desde que comienza el proyecto hasta que “termina”, cuánto están previstas que duren, cómo vamos iterando sobre este proyecto, cómo se va involucrando poco a poco con otras áreas, cómo vamos a gestionar el cambio… En nuestro caso, *HM Hospitales* tendría que definir cuándo comenzó el proyecto, en qué punto están, definir métricas para saber cómo evoluciona el proyecto, cómo se involucra el gobierno del dato por ejemplo en cardiología o neumología…

Una vez que sabemos a priori cómo va a ser la hoja de ruta, todo proyecto tiene unos costes, tanto en tiempo como en recursos, por lo tanto, debemos definir en la siguiente fase la inversión tanto de tiempo como de recursos que se van a necesitar. Tal y como vimos en la anterior práctica *HM Hospitales* se había aliado con *Telefónica* [2] para llevar a cabo la transformación digital, y se habían marcado un límite temporal hasta 2024 pero desconocíamos la inversión en recursos, aunque podemos intuir que será elevada.

En la siguiente fase debemos diseñar el programa del gobierno del dato, es decir, sobre cómo se va a usar, qué funcionalidades tendrá… Por lo que *HM Hospitales* tendrá que definir los procedimientos para considerar el dato como un activo, definir procesos para el negocio (si tenemos un problema, por ejemplo no se está almacenando los datos de los pacientes que van a consulta, cómo vamos a solventarlo), debemos definir el acceso a la información (no todos los médicos tendrían que poder ver la misma información, no es lo mismo los pacientes de cardiología que los de neumología)…

Una vez definidos todos estos procesos, debemos implementarlos, es aquí donde entre en juego la fase de implementación del programa. *HM Hospitales* tendría que definir cuándo es el lanzamiento según la hoja de rutas y cómo va a ir evolucionando el proyecto.

Cuando ya tenemos los procesos implementados definimos el diseño del marco organizativo, es decir, el diseño del gobierno del dato. *HM Hospitales* lo que haría sería identificar básicamente roles y responsabilidades sobre la gestión de los datos, por ejemplo quién se encarga de actualizar los datos, bajo qué responsabilidad…

Finalmente, tenemos que poder comprobar el cómo evoluciona el proyecto, si necesita mejoras, si hay ramas en las que no está funcionando… Por lo tanto, monitoreamos y medimos la evolución del proyecto para así asegurarnos que el gobierno del dato dentro de *HM Hospitales* está funcionando de forma eficaz.

## Pregunta 2

*Después de la lectura del bloque 2, estamos seguros de que el gobierno del dato lo entendemos como un proyecto multifuncional.*

### Define y describe cada una de las funciones según DAMA.

Las funciones o componentes definidas según DAMA son nueve, tal y como vemos en la siguiente ilustración:

Diagrama

Descripción generada automáticamente

Ilustración - Funciones según DAMA.

* **Arquitectura de datos**: tal y como su proprio nombre indica, se encarga de todo lo que tiene que ver con las estructuras de datos, ya sean para almacenarlos, procesarlos…
* **Diseño y modelos de datos**: esta función está relacionada con la minería de datos, es decir, se busca crear y gestionar modelos de datos con el fin de mejorar la toma de decisiones.
* **Almacenamiento de datos**: es un rol parecido al de arquitecto de datos, ya que busca gestionar todas las estructuras de almacenamiento de datos, pero en este caso a nivel físico.
* **Seguridad de datos**: se basa en la gestión de la seguridad, privacidad, confidencialidad y acceso de los datos. Básicamente se tiene que garantizar que el uso que se vaya a dar a los datos cumpla con la legislación de cada país.
* **Datos maestros y de referencia**: esta función es la encargada de identificar y mantener los datos maestros y de referencia.
* **Inteligencia de negocio y almacenes de datos**: esta función está relacionada con el business intelligence ya que busca usar procesos analíticos a partir de bases de datos analíticas para mejorar la toma de decisiones.
* **Contenido y documentos**: su objetivo es poder almacenar los datos en documentos para así poder mejorar la capacidad de los sistemas de información para intercambiar información, compartir datos…
* **Metadatos**: se busca el identificar, mantener, integrar, básicamente gestionar los metadatos, los cuales son datos sobre los datos para describir mejor a estos segundos.
* **Calidad de dato**: se centra en definir procedimientos para garantizarnos la calidad del datos y la integridad del mismo.

Para terminar, hay que decir que según los apuntes hay una función más de las que hemos visto y que no queda reflejada en la anterior ilustración, esta es la integración e interoperabilidad de datos:

* **Integración e interoperabilidad de datos**: se basa en cómo obtener datos, cómo extraerlos, cómo transformarlos, cómo distribuirlos… Su objetivo es mejorar la compartición de datos.

### Argumenta para la iniciativa de *Covid Data Save Lives*, qué modelo de gobernanza de habrá adaptado mejor según sus dimensiones básicas.

A finales del año 2019 se detecta por primera vez en China una mutación del conocido virus *coronavirus*, éste se expande por todo el planeta causando numerosas muertes sin tener en cuenta la gran crisis social y económica que ha dejado a su paso.

*Covid Data Save Lives* es un proyecto creado por *HM Hospitales* el cual trata de reducir el impacto de este virus en nuestra sociedad, para ello se recoge en un *dataset* clínico información sobre los pacientes que tienen el *coronavirus* en hospitales de *HM Hospitales*¸ se obtiene información sobre diagnósticos, tratamientos, medicinas, resultados de laboratorio… [3], [4]

A todo esto cabe destacar que *HM Hospitales* está en un periodo de transformación digital, esto significa que se está centrando en los datos para mejorar las tomas de decisiones que se realizan dentro de la organización, y a su vez mejorar los servicios ofrecidos a los pacientes [2].

Una vez que tenemos un contexto podemos discernir qué modelo de gobernanza se ajusta mejor al proyecto, tal y como hemos visto hay cuatro modelos de gobernanza [1]: modelo indisciplinado, modelo reactivo, modelo proactivo y modelo gobernado.

Respecto a la dimensión personal, este proyecto se ha podido realizar porque el gobierno del dato está soportado por un ámbito ejecutivo dentro de la organización, es decir, *HM Hospitales* está en plena transformación digital y los ejecutivos apoyan la gestión del dato para mejorar la toma de decisiones, en nuestro caso poder solventar los problemas causados por el *coronavirus*.

Además, debe de existir un grupo que se encargue de la obtención, manipulación y calidad de los datos, por lo que dentro de esta perspectiva solo tenemos dos modelos que lo cumplen, tanto el modelo proactivo como el modelo gobernado, aunque es verdad que el modelo gobernado en este caso se adapta mejor.

Respecto a la dimensión normas, *Covid Data Save Lives* no tiene procesos y datos aislados, no hay antigüedad en los datos, sí que existe la calidad de los mismos, se busca predecir el comportamiento del *coronavirus* por lo que se centra en solucionar los problemas que pueden surgir, no que ya hayan surgido.

También se recogen diferentes métricas para saber como evoluciona el modelo, por lo tanto al igual que antes, tanto el modelo proactivo como el gobernado encajan en este punto. Es verdad, que es difícil saber si esta iniciativa se ha llevado a cabo después de considerar cómo afectaría a la organización y al dato, pero observando que es una empresa que se está transformando digitalmente podemos intuir que sí, es decir, que seguramente respecto a esta dimensión está más cerca de un modelo gobernado que proactivo.

Respecto a la dimensión tecnología, es difícil saber en qué modelo nos encontramos ya que no hay demasiada información por internet, pero teniendo en cuenta que se produce un monitorizado de los datos de forma continua podemos descartar los dos primeros modelos (éstos se centran en que sí que hay herramientas de control de calidad pero no existe una monitorización e integración continua). Por lo tanto, tanto el modelo proactivo como gobernado se adaptan a esta dimensión.

Respecto a la dimensión riesgos y recompensas, el riesgo es bajo ya que al disponer de una mayor información aumentamos la fiabilidad de los mismos y por lo tanto mejoramos la toma de decisiones, y la recompensas son elevadas ya que en la organización se comprende y usa mejor el gobierno del dato para mejorar servicios y la toma de decisiones.

En resumen, tanto el modelo proactivo como el modelo gobernado se ajustan perfectamente a las características del proyecto, pero teniendo en cuenta que la organización que hay detrás es *HM Hospitales*, y que dicha organización está transformándose digitalmente para así tomar mejores decisiones basadas en datos, considero que el modelo gobernado sería el ideal.

## Pregunta 3

*Los datos de la iniciativa Covid Data Save Lives:*

### ¿Cómo de privados los consideras?

a

|  |
| --- |
| Bibliografía |

[1] «MostraPDFMaterialAction.pdf». Accedido: abr. 22, 2021. [En línea]. Disponible en: http://cvapp.uoc.edu/autors/MostraPDFMaterialAction.do?id=246836.

[2] «Telefónica y HM Hospitales digitalizan el sector salud», *El Español*, abr. 30, 2019. https://www.elespanol.com/invertia/disruptores-innovadores/innovadores/20190430/telefonica-hm-hospitales-digitalizan-sector-salud/393962093\_0.html (accedido abr. 23, 2021).

[3] «Covid Data Save Lives». https://www.hmhospitales.com/coronavirus/covid-data-save-lives (accedido abr. 24, 2021).

[4] C. 31-12-2020 | 16:00 H. actualización 31-12-2020 | 16:00 H, «HM Hospitales aúna investigación y Big Data para cercar a la Covid-19», *La Razón*, dic. 31, 2020. https://www.larazon.es/salud/20201231/yzm7vnrllzgrbpvc4f76fbak7q.html (accedido mar. 06, 2021).