



Universitat Oberta
de Catalunya

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
Índice de ilustraciones	4
1. Enunciado	5
1.1. Pregunta 1	5
1.1.1. ¿Cómo divulgarías internamente qué es el gobierno del dato en una HM Hospitales y su importancia en la transformación digital?.....	5
1.1.2. Si te basaras en DAMA, ¿qué objetivos principales perseguirías sobre los datos?	6
1.1.3. ¿Cuál sería tu planteamiento para la consecución de estos objetivos, en función de las fases que tendrías que implementar para el gobierno del dato?	7
1.2. Pregunta 2	8
1.2.1. Define y describe cada una de las funciones según DAMA.	8
1.2.2. Argumenta para la iniciativa de <i>Covid Data Save Lives</i> , qué modelo de gobernanza de habrá adaptado mejor según sus dimensiones básicas.	10
1.3. Pregunta 3	12
1.3.1. ¿Cómo de privados los consideras?	12
1.3.2. Razona la influencia de la privacidad en la seguridad de los datos. .	13
1.3.3. ¿Qué elementos clave de seguridad y privacidad consideras que se han tenido en cuenta en la iniciativa de <i>HM Hospitales</i> ?	14
1.4. Pregunta 4	15
1.4.1. ¿Ha primado lo bueno, lo justo o lo correcto?	15
1.4.2. Qué principios han de orientar tu ejecución profesional como científico de datos desde la perspectiva de la ética.	16
2. Bibliografía	19

Índice de ilustraciones

Ilustración 1 - Funciones según DAMA.	9
--	---

1. Enunciado

1.1. Pregunta 1

En la transformación digital de HM Hospitales, además de los cambios 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:

1.1.1. ¿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.

Esta 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 transformación 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.

1.1.2. 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.

1.1.3. ¿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.

1.2. Pregunta 2

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

1.2.1. 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:



Ilustración 1 - Funciones según DAMA.

- **Arquitectura de datos:** tal y como su propio 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.

1.2.2. 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 cómo 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.

1.3. Pregunta 3

Los datos de la iniciativa Covid Data Save Lives:

1.3.1. ¿Cómo de privados los consideras?

Antes de definir lo que es privacidad, vamos a determinar qué es la seguridad del dato. Podemos entender por seguridad del dato la protección que se hace de los mismos para así evitar que personas no autorizadas tengan acceso y lo puedan modificar, como mínimo la seguridad del dato debe garantizar la confidencialidad, integridad y disponibilidad, aunque también podríamos añadir la autenticidad y la garantía de modificar/eliminar los datos [5].

Una vez que tenemos claro lo que es la seguridad del dato, vamos a definir la privacidad como la protección de la información personal, es decir, de los datos personales [5] que son sensibles.

Sabiendo lo que significa privacidad podemos responder a la pregunta, ¿cómo de privados son los datos de *Covid Data Save Lives*? Cabe destacar que no hay una escala para saber si los datos son más o menos privados, sino que vamos a hacer una observación general.

Recapitulando sobre *Covid Data Save Lives*, almacenaban información sobre los pacientes que acudían a hospital de *HM Hospitales* y presentaban patologías del *COVID-19*, se extraía información sobre el ingreso/alta médica, los tratamientos recibidos, los resultados del laboratorio, el paso por la UCI... [3]

Cuando hablamos de estos tipos de datos podemos apreciar que es información sensible ya que son datos muy personales, por lo que podemos considerar que dichos datos son privados, que la información sea privada no solo significa que impidamos el acceso a personas no autorizadas, sino que también el cómo se va a explotar esa información.

Por lo tanto, la tecnología ha avanzado mucho y hoy en día hay áreas que antes no existían, esto nos indica que los retos a lo que tenemos que hacer frente actualmente son mayores, basándonos que tenemos que garantizar la confidencialidad, integridad y disponibilidad, algunos de estos retos son:

- Cada día hay más cantidad de datos y muchos de ellos son distribuidos, por lo que garantizar la privacidad en todos ellos es una tarea compleja.
- La facilidad de duplicar la información también complica las labores de seguridad y privacidad.

- El almacenamiento en la nube permite que la información esté en diferentes centros remotos, por lo que si sufrimos un ataque en dichos centros nuestra información puede verse afectada.
- Gestión de datos, lo cual supone enriquecimiento de datos, es decir, que terceros puedan saber información personal como por ejemplo los bancos. También puede darse el caso de que la distribución de los datos se vea afectada, por no ser segura.
- O en el caso que estamos viendo, puede ser que la información del *dataset* haya alguna copia en el que los datos no son seguros porque no esté garantizada la confidencialidad, ya sea porque aparece el DNI del paciente...
- ...

1.3.2. Razona la influencia de la privacidad en la seguridad de los datos.

Haciendo un resumen de lo comentado en el punto anterior, la seguridad busca que la información como mínimo sea confidencial, íntegra y disponible. Por otro lado, el objetivo de la privacidad es garantizar que la información privada siga siendo privada tanto en el acceso como en el uso.

La privacidad aunque es un pilar fundamental de la seguridad, algunas veces se vulnera de forma voluntaria, aunque la gran mayoría se hace de forma inconsciente, ya que al hacer uso de esta información podemos vulnerar este principio de seguridad.

Que un dato sea seguro no implica que sea privado, por ejemplo, los estudios que realiza un estudiante no es información privada, sin embargo, la privacidad sí que implica seguridad, es decir, siguiendo con el ejemplo la renta de un estudiante sí que es de ámbito privado ya que es sensible, y dicho valor se puede usar para diferentes fines tanto positivos como negativos.

Por lo tanto, como podemos observar la privacidad va ligada a la seguridad, y su influencia tiene que ser la más alta posible aunque haya veces que nos olvidemos, ya que garantizar la privacidad de los datos sensibles es fundamental, y de no hacerlo puede conllevarnos que estemos incumpliendo diferentes leyes/normativas que haya en cada país.

Respecto al proyecto *Covid Data Save Lives*, los datos que se usan son de índole privada, ya que podrían caer en manos de terceros y sacar ventaja de ello, por ejemplo si se filtra los análisis de un paciente, una farmacéutica puede invertir en publicidad en *Google* o *Facebook* para vender sus medicamentos, o un seguro podría no concedernos un seguro de vida, o el banco rechazarnos un préstamo...

En resumen, la influencia de la privacidad en la seguridad es esencial, por lo que debemos garantizar siempre que los datos privados sigan siendo privados, a no ser que haya un acuerdo en este caso con el paciente para poder hacer uso de ellos, de ser así tendríamos que usar técnicas para anonimizar la información.

1.3.3. ¿Qué elementos clave de seguridad y privacidad consideras que se han tenido en cuenta en la iniciativa de *HM Hospitales*?

Cuando hablamos de elementos clave de seguridad y privacidad nos estamos refiriendo básicamente a tres elementos [5]: confidencialidad, integridad y disponibilidad. Todos ellos buscan garantizar la seguridad de los datos y mantener que la información privada siga siendo privada.

Por lo tanto, vamos a analizar cada uno de los elementos respecto al proyecto *Covid Data Save Lives*:

- **Confidencialidad:** entendemos por confidencialidad al elemento que nos permite garantizar que solamente las personas autorizadas tengan acceso a los datos, este punto *Covid Data Save Lives* sí que lo cumple, ya que como podemos observar en su página web, solamente aquellas instituciones académicas y del ámbito sanitario pueden acceder a dicho *dataset* [3].

Por otro lado, confidencialidad no se refiere solo a el acceso a los datos, sino que también a cómo están encriptados y anonimizados los mismos, este punto no lo podemos saber ya que no viene información por internet, sin embargo podemos intuir que la información al ser sensible y relacionada con el ámbito sanitario seguramente esté como mínimo anonimizada.

- **Disponibilidad:** los datos es verdad que tienen que ser seguros y privados, pero también debemos de garantizar que se puede acceder a ellos sin importar el cuándo se necesiten, desde dónde y en qué forma. Por lo tanto, buscamos que los datos estén accesibles todo el tiempo para así poder ser usados por todo el mundo, para ello podemos definir mecanismos de accesos a los datos, mecanismos para prevenir la pérdida de los mismos... Este punto también lo cumple *Covid Data Save Lives*, ya que podemos acceder a su página web las 24 horas del día los 365 días del año y solicitar el acceso a los datos.

Respecto a los mecanismos usados para el acceso del dato podemos ver que se centra básicamente en la solicitud de los datos de forma manual, y podemos suponer que luego *HM Hospitales* proporcionará algún tipo de fichero (CSV, Excel...) con los datos.

En cuanto a mecanismos usados para la prevención de pérdida de datos, carecemos de información suficiente para determinar qué mecanismos usan pero seguramente los tengan, ya que *HM Hospitales* es una empresa orientada al dato y este tipo de procesos estarán claramente definidos.

- **Integridad:** con este término nos estamos refiriendo a que los datos tienen que ser correctos y los más completos posible, es decir, de nada sirve tener datos que no tengan calidad suficiente para poder después analizarlos, explotarlos, etc., por ejemplo si se mide la capacidad respiratoria de un paciente y esos datos son números negativos entonces no tienen sentido.

La información tiene que ser lo más precisa y actualizado posible, de nada sirve tener registros de la mutación del *coronavirus* del año 1930, principalmente porque son mutaciones diferentes y la forma de actuar del virus también lo va a ser.

Para garantizar que los datos sean correctos, completos, precisos y actualizados la organización puede realizar las siguientes tareas: hacer un inventario de datos, identificar riesgos de daño personal, identificar riesgos de desajuste en el sistema...

Respecto a este elemento clave en la seguridad y privacidad, *Covid Data Save Lives* también lo cumple, como bien hemos mencionado *HM Hospitales* es una organización orientada al dato, por lo que podemos estar casi seguros de que existen procesos que garantizan la calidad de los datos y procesos que se encargan de actualizar/modificar/añadir/eliminar datos. Por lo tanto, aunque en su página web no especifique qué procesos usa, podemos intuir que este punto también se garantiza.

1.4. Pregunta 4

Si analizamos la iniciativa de Covid Data Save Lives:

1.4.1. ¿Ha primado lo bueno, lo justo o lo correcto?

Antes que nada debemos indicar que las teorías éticas nos permiten tomar diferentes puntos de vista a la hora de la toma de decisiones, con esto nos referimos a que no hay una teoría ética mejor que otra, sino que dependiendo de lo que nosotros valoremos podemos aplicar una u otra.

Desde el comienzo de la ética, podemos dividir estas teorías en dos grupos: aquellas en las que se prima lo bueno (éticas teleológicas), o aquellas en las que se prima lo correcto (éticas no teleológicas) [6].

Respecto a las éticas no teleológicas, buscan que toda acción moral lleve consigo el cumplimiento de alguna norma o regla, es decir, discernir entre lo que es correcto/justo y lo que no a partir de reglas, por lo que las consecuencias de una acción no van a afectar en la toma de decisiones.

Por lo tanto, actuar de forma correcta/justa es el cumplir una norma, pero dado el caso de *Covid Data Save Lives* vemos que el objetivo de dicho *dataset* no es discernir lo que es correcto de lo que no, es decir, no vamos a tener una norma o un conjunto de normas para saber si es correcto o no dicho *dataset*, esto no significa que no se tenga en cuenta este tipo de ética, ya que simplemente para recopilar los datos nos tenemos que centrar en esta ética, es decir, vamos a tener un conjunto de leyes en esta caso españolas/europeas en las que tenemos que garantizar la seguridad y privacidad de los datos, tal y como se explicó en la pregunta anterior.

Respecto a las éticas teleológicas buscan maximizar el bien, por lo tanto, tienen en cuenta las consecuencias de las acciones que se han tomado, es decir, en nuestro caso *Covid Data Save Lives* busca maximizar el bien, entendiendo por bien conocer mejor esta mutación del *coronavirus* y así reducir la mortalidad, predecir el cómo va a evolucionar, cómo se propaga...

A la hora de primar este *dataset* prima lo bueno antes que lo correcto, aunque tal y como hemos comentado esto no implica que carezca de ser correcto, ya por el mero hecho de almacenar información sensible la ética no teleológica está implícita, es decir, *Covid Data Save Lives* pone a disposición datos con el fin de mejorar y entender la situación actual (prima lo bueno), pero esta información se puede extraer/distribuir gracias a una serie de normas (implícito lo correcto).

En resumen, a nivel general este proyecto busca solventar o mejor dicho mejorar la situación actual debido al *coronavirus*, basándose en las consecuencias para maximizar el valor (salvar vidas), es decir, primando lo bueno.

1.4.2. Qué principios han de orientar tu ejecución profesional como científico de datos desde la perspectiva de la ética.

Cuando hablamos de los principios orientadores, nos estamos refiriendo a una forma de ayudar a evaluar cómo de buena o de mala es una solución, de esta forma conseguimos mejorar la toma de decisiones.

Respecto a los principios que nos permiten orientar la ejecución profesional, los podemos dividir en dos grupos [6]:

- EL principio de beneficencia y autonomía.
- El principio de no maleficencia y justicia.

La diferencia entre ambos grupos son la teoría ética en las que se basan, el primer grupo se centra en la ética teleológica mientras que el segundo grupo en la ética no teleológica, es decir, los primeros se centran en las consecuencias para determinar cómo de buena es una solución, y los segundos se basan en normas para discernir lo que es correcto de lo que no.

Tal y como vemos cada grupo está formado por dos principios, por lo tanto, tenemos cuatro principios que nos permiten orientar a cómo de profesionales somos a la hora de tomar una acción.

El principio de beneficencia se basa en hacer el bien, para ello identificamos los valores que tienen que ser maximizados. Que este principio se centre en hacer el bien no implica que no tenga en cuenta el daño ocasionado, de tal forma que hay que buscar el no hacer daño, o si lo hacemos que sea el menor posible para así maximizar el bien. Respecto al proyecto *Covid Data Save Lives* buscamos el maximizar el bien, que puede ser entender mejor cómo se comporta el *coronavirus*, salvar vidas...

El segundo principio es el de autonomía, este principio junto con el anterior suele confrontar, ya que es en este principio en el que consideramos la autonomía de los individuos, es decir, son los usuarios lo que tienen el derecho a tomar decisiones en los asuntos que afecten a su vida, en nuestro caso se corresponde con que cada persona pueda tomar decisiones libremente respecto a la privacidad, al uso y distribución de sus datos... Por lo tanto, este principio lo que nos permite definir es un límite, para así garantizar que es ético el proceso que estamos llevando a cabo.

El tercer principio es el de no maleficencia, se basa en no hacer daño, es decir, cuando entran en conflicto los dos anteriores principios, hacemos uso de este principio para garantizar que no haya consideraciones particulares, sino que buscamos el no hacer daño en un colectivo. Este principio es muy interesante, porque en el caso del *dataset Covid Data Save Lives* el principio de beneficencia sería el maximizar el número de vidas salvadas, el principio de autonomía nos dictaminaría cómo el paciente puede ejercer su derecho de que los datos no sean compartidos ni extraídos, y el principio de no maleficencia entraría en juego si un paciente ejerce su derecho de no compartir los datos, cómo afectaría al principio de beneficencia, es decir, cuál sería el daño causado.

Por último, tenemos el principio de justicia, éste permite poner límites al principio de no maleficencia a partir de la toma de decisiones siguiendo un criterio de justicia social.

En resumen, estos principios nos ayudan a tomar decisiones a partir del diagnóstico del problema en cuanto a valores y acciones, los cuales son usados diariamente incluso de forma inconsciente, en el ámbito de la ciencia de datos cada día van tomando más importancia, debido a la poca regulación que hay todavía sobre la explotación y difusión

de los datos, sin embargo, gracias a estos principios podemos seguir manteniendo una correcta alineación entre valores y acciones.

2. 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 aún investiga y Big Data para cercar a la Covid-19», *La Razón*, dic. 31, 2020. <https://www.larazon.es/salud/20201231/yzm7vnrlzgrbpvc4f76fbak7q.html> (accedido mar. 06, 2021).
- [5] «MostraPDFMaterialAction.pdf». Accedido: abr. 24, 2021. [En línea]. Disponible en: <http://cvapp.uoc.edu/autors/MostraPDFMaterialAction.do?id=246839>.
- [6] «PID_00243550.pdf». Accedido: abr. 25, 2021. [En línea]. Disponible en: https://materials.campus.uoc.edu/daisy/Materials/PID_00243550/pdf/PID_00243550.pdf.