
El cicle de vida de la dada

PID_00251618

Marcos Pérez González

Temps mínim de dedicació recomanat: 4 hores



Cap part d'aquesta publicació, inclòs el disseny general i la coberta, pot ser copiada, reproduïda, emmagatzemada o transmesa de cap manera ni per cap mitjà, sigui aquest elèctric, químic, mecànic, òptic, de gravació, de fotocòpia, o per altres mètodes, sense la prèvia autorització escrita dels titulars del copyright.

Índex

Introducció	5
Objectius	6
1. La dada com a actiu de valor per a l'organització	7
1.1. Estratègies competitives basades en actius	7
1.2. La dada com a actiu de valor	8
2. El cicle de vida de les dades	11
2.1. Què és el cicle de vida d'un actiu?	11
2.2. Com mesurar el valor de la dada?	12
2.3. Fases del cicle de vida de la dada	15
2.4. Com és possible detectar els canvis en una dada?	17
3. Govern de la dada	20
3.1. Orígens del <i>data governance</i>	20
3.2. Definició de <i>data governance</i>	21
3.3. Objectius del <i>data governance</i>	22
3.4. Fases del <i>data governance</i>	24
4. Components i maduresa del <i>data governance</i>	29
4.1. Components	29
4.1.1. Normes, estàndards i estratègia	31
4.1.2. Qualitat de la dada	31
4.1.3. Gestió de dades mestres	31
4.1.4. Gestió de la seguretat de la dada	32
4.2. Maduresa	32
4.3. Model de maduresa del <i>data governance</i>	33
4.3.1. Models de <i>data governance</i>	35
4.4. Quin model de maduresa fem servir?	37
Resum	39
Glossari	40
Bibliografia	41

Introducció

A mesura que les organitzacions fonamenten les seves decisions i iniciatives en dades, sorgeixen preguntes com: d'on provenen les dades?, tenen la qualitat necessària?, què sabem sobre la nostra informació?, estan aquestes dades alineades amb la nostra política d'empresa?, tota l'organització té la mateixa visió sobre la dada?

Dur a bon port iniciatives d'intel·ligència de negoci, *data warehousing* (en català, emmagatzematge de dades), analítica de dades o *big data* (en català, dades massives) requereix tenir un enfocament holístic respecte a la dada i poder respondre a les preguntes anteriors.

Per donar resposta a aquestes necessitats, emergeix el **govern de la dada**, que permet a una organització administrar de forma més eficient i eficaç la dada. De fet, aquesta iniciativa corporativa uneix persones, processos i tecnologia per canviar la forma en què les dades són adquirides, gestionades, mantingudes, transformades en informació, compartides en el context de l'organització com a coneixement comú i sistemàticament obtingudes per l'empresa per millorar-ne la rendibilitat.

Tot i que el govern de la dada es recolza sobre solucions de tecnologia, al llarg d'aquest mòdul comprendrem que està fonamentat en el desenvolupament de procediments i marcs d'actuació dins d'una organització.

En aquest mòdul estudiarem la necessitat i importància de conèixer el cicle de vida de la dada i en què consisteix el govern de la dada. Així mateix, comprendrem què aporta el govern de la dada i on impacta, quins són els seus components i, finalment, com avaluar la maduresa de la nostra pròpia organització.

Objectius

Aquest material didàctic està adreçat a:

- 1) Desenvolupadors i consultors que volen conèixer què significa el govern de la dada o *data governance*.
- 2) Desenvolupadors i consultors que volen ajudar al disseny i l'implementació d'estratègies de negoci que incloguin govern de la dada.
- 3) Gestors que estan interessats en la transformació digital de la seva organització i en la inclusió de govern de la dada com un dels seus pilars fonamentals.

En els materials didàctics d'aquest mòdul trobarem les eines indispensables per assimilar els objectius següents:

1. Entendre el concepte de *data governance*, les situacions en què és necessari desplegar una solució d'aquest tipus i els avantatges que proporciona.
2. Conèixer el cicle de vida de la dada i les necessitats en cadascuna de les fases del cicle.
3. Enumerar i donar a conèixer els diferents components del govern de la dada.
4. Saber què signifiquen les diferents etapes de maduresa en el govern de la dada i com analitzar la maduresa de la pròpia organització.

Si bé l'obra és autocontinguda en la mesura del possible, els coneixements previs necessaris són:

- 1) Coneixements bàsics sobre *business intelligence* i *big data*.
- 2) Coneixements sobre estratègia i gestió de les tecnologies de la informació (TI).

S'introduiran els conceptes necessaris per al seguiment d'aquest material.

1. La dada com a actiu de valor per a l'organització

1.1. Estratègies competitives basades en actius

Durant les últimes dècades, millorar el posicionament d'una organització ha estat objecte d'estudi tant dins com fora de l'acadèmia. Quan una organització disposa d'una millor posició que els seus rivals, diem que té un **avantatge competitiu** que permet assegurar els clients i defensar-se contra les forces competitives. Per exemple, Delta Airlines, aerolínia dels Estats Units, ha estat considerada durant molts anys un exemple en l'ús de l'eficiència operacional per destacar per sobre d'altres companyies en el seu mateix mercat. Un altre exemple és Apple, que ha destacat durant diversos anys per oferir productes altament integrats en un àmbit de programari i maquinari combinat (com iPod, iPhone o AirPods), amb dissenys orientats al client com a mecanisme de diferenciació.

Tradicionalment, els avantatges competitius s'han fonamentat en:

- **Lideratge en cost:** l'organització és capaç de crear un producte o proporcionar un servei a menor cost que els seus competidors.
- **Diferenciació:** l'organització és capaç d'oferir un servei o producte superior pels clients. La diferenciació és, freqüentment, el resultat d'optimitzar la cadena de valor del mercat en el qual s'opera o de dur a terme canvis en l'abast competitiu.

A mitjan anys vuitanta, emergeix un model que considera que els recursos i les capacitats que té l'organització són la clau per generar un rendiment superior al dels seus competidors. En aquest enfocament estratègic, és necessari poder determinar si els recursos i les capacitats són rellevants, cosa que proporciona el marc de referència que es mostra a la figura 1.

El marc anterior busca identificar i classificar els actius d'una organització (**financers, físics, humans, tecnològics, de reputació o organitzacionals**) en funció del seu valor. L'objectiu és identificar les fortaleeses i debilitats en relació amb els seus competidors i, d'aquesta manera, trobar noves oportunitats o un ús més eficient dels recursos. De fet, aquest mateix marc serveix per analitzar les capacitats d'una organització, ja que una capacitat es pot entendre com un **recurs intangible**.

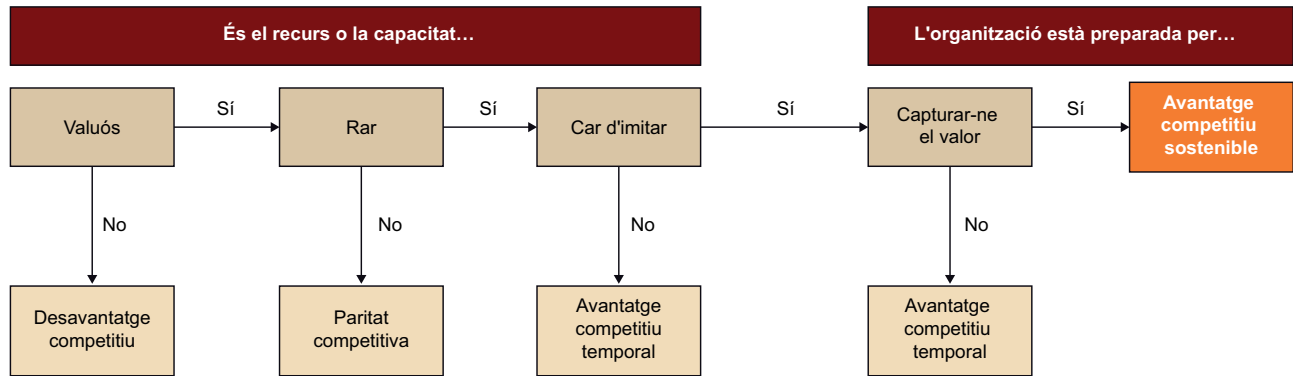
Lectura complementària

Porter, M. (1998). *Competitive Advantage Creating and Sustaining Superior Performance*. Free Press.

Lectura complementària

Wernerfelt, B. (1984). «A Resource-Based View of the Firm». *Strategic Management Journal* (núm. 2, vol. 5, pàg. 171-180).

Figura 1. Marc de referència



Font: Frank Rothaermel

Està clar que el principal objectiu d'aquesta estratègia és generar avantatges competitius sostenibles, fonamentats en:

- **Durada**, és a dir, la velocitat a la qual un recurs es deprecia o es converteix en obsolet.
- **Transparència**, és a dir, com menys transparent sigui el recurs, menys possibilitats que sigui imitat per un competidor.
- **Transferència**, és a dir, com menys capacitat de transferència de recursos —per exemple, no es poden adquirir—, menys capacitat per ser imitat per un competidor.
- **Replicació**, és a dir, com més complexitat de processos i capacitats d'organització, més gran serà la dificultat de replicació.

El fet que una organització tingui recursos i capacitats susceptibles de generar avantatges competitius no vol dir que realment els generi. En definitiva, cal dissenyar una estratègia que identifiqui els recursos i les capacitats de l'organització, mesuri el seu valor, maximitzi la seva explotació i, a més, es dugui a terme de manera contínua per augmentar aquest tipus de recursos dins de l'organització.

1.2. La dada com a actiu de valor

Vivim una època de profunda transformació de les organitzacions, fonamentada en tecnologia (com el núvol, impressió 3D, etc.), en la qual la gran majoria dels seus processos de negoci i la seva cadena de valor estan sent revisats i interpretats de nou.

Aquesta transformació té diferents efectes. Entre aquests efectes, destaca la generació de dades com a subproducte o com a intenció principal de la transfor-

mació. Quan, de manera conscient, l'organització gestiona la dada, és capaç de prendre decisions més eficaces i eficients, i de competir en el mercat d'una forma diferent.

És destacable, per exemple, el cas de Netflix, que fa servir les dades de client (el seu comportament i preferències) combinats amb algoritmes per evitar l'abandonament de clients. De fet, el seu coneixement i l'eficiència dels seus algoritmes incrementen a mesura que la companyia té més clients, el que genera un efecte de xarxa.

Fins a tal punt és important aquesta iniciativa que, segons Netflix, el seu impacte en el negoci és d'un bilió de dòlars.

Tal com apunten Higson i Waltho, aquestes organitzacions han canviat la seva percepció respecte a la dada. Han passat de considerar la dada com a un actiu tòxic, l'ús del qual és necessari controlar i limitar, a un actiu de valor, l'ús del qual s'ha de gestionar i maximitzar.

És a dir, aquelles empreses que s'estan transformant en organitzacions orientades a la dada contemplan la dada com un recurs que genera avantatges competitiu, seguint les idees de Wernerfelt.

El focus de la dada no només està en l'eficiència, com en el cas de Netflix. Ian Davis, analista de McKinsey, postula que les organitzacions han de ser àgils per adaptar-se als canvis del mercat. I defensa el paper de la dada com a font de l'agilitat i l'aprenentatge continu. Perquè una empresa sigui flexible davant el mercat, ha de poder adaptar les seves decisions de forma contínua i, per tant, serà tan flexible com ho siguin la seva dada i les seves capacitats d'anàlisi.

Des de la perspectiva d'un actiu, la dada té les seves peculiaritats si ho comparem amb un altre tipus d'actius:

- **Les dades freqüentment tenen una naturalesa transitòria**, és a dir, estan vinculades a un temps i a un moment. Per exemple, els valors borsaris, com a actiu, només tenen sentit si són capturats i analitzats de forma contínua. Això és molt diferent, per exemple, en el cas d'un edifici, el valor del qual presenta una durada més àmplia.
- **La dada requereix un manteniment continu**, és a dir, cal manipular-la per mantenir o fins i tot incrementar el seu valor.
- **La dada pot reutilitzar-se**, és a dir, a diferència d'altres actius que amb l'ús es desgasten, com ara una pilota, la dada té la possibilitat de ser utilitzada en diferents escenaris, a un cost marginal.
- **La dada posseeix un gran potencial com a actiu**, ja que té una baixa o nul·la transparència, transferència i replicació, i una llarga durada.

No obstant això, la visió de la dada com un actiu encara no està completament implantada a les organitzacions. El principal problema rau en el fet que

Lectura complementària

Higson, C.; Waltho, D.
(2010). *Valuing Information as an Asset*.

no es tenen mecanismes establerts per mesurar el cost i el valor de la dada, ni tampoc, freqüentment, processos per gestionar de forma adequada aquest actiu.

Si es posseïssin aquests valors, igual que es fa amb altres actius, la dada podria tenir una entrada en els informes financers, i seria factible dirimir el seu impacte en el negoci. A més, això permetria:

- Assegurar una cura i una administració apropiades dels actius de dades.
- Incentius per utilitzar aquests actius de forma eficient.
- Potenciar la recerca de rendibilitat sobre aquests actius, i classificar-los en funció de la seva rendibilitat.

No obstant això, aquest escenari no és possible de manera generalitzada. El valor de la dada està relacionat amb la capacitat de diferenciació i creació d'avantatges competitiu, com en el cas de Netflix, fet que suposa que cada organització podria tenir mecanismes diferents i propis per a mesurar-ne el valor.

Si es disposés d'un procés sistemàtic per gestionar les dades com a actiu, seria factible poder seguir l'exemple de Netflix i determinar què significa valor i com evoluciona.

Per tant, el nostre objectiu és respondre a les preguntes següents: com és possible identificar quines dades de l'organització són realment actius de valor? Com mesurem el valor d'aquests actius? Com determinem l'evolució del valor al llarg del temps?

En els apartats següents, discutirem aquestes preguntes, que s'articulen per mitjà del **govern de la dada**.

2. El cicle de vida de les dades

2.1. Què és el cicle de vida d'un actiu?

En el moment en el qual una organització identifica la dada com a actiu de valor, el pas següent és la gestió d'aquest actiu de forma precisa. Per a això, és necessari poder conèixer i associar a cada dada el que definim com el cicle de vida d'un actiu:

Entenem com a **cicle de vida d'un actiu** les diferents etapes per les quals passa un actiu des del seu naixement fins a la seva fi.

La visió del cicle de vida d'un actiu vincula directament l'actiu amb una activitat o un projecte. Per aquesta raó, en general, el cicle de vida d'un actiu pot estar vinculat a les etapes següents:

- idea inicial i estudis preliminars (del projecte);
- avaluació del projecte respecte a diferents perspectives, com la viabilitat tècnica o econòmica;
- definició i planejament de les etapes del projecte, i en quines d'elles l'actiu és rellevant;
- detall del projecte i disseny dels processos;
- execució del projecte d'acord amb la planificació;
- posada en marxa, prova de totes les instal·lacions i consegüent acceptació;
- operació de les instal·lacions, ús o consum dels béns o serveis;
- avaluació del projecte per identificar millores o canvis següents;
- descart, reciclatge o venda de la instal·lació.

És a dir, es tracta de saber tot el que passa amb la dada: des de com es crea o s'incorpora a un procés, fins al descart final, el seu reciclatge o la seva venda. Comprendre el cicle de vida de la dada significa analitzar les condicions de la dada en l'organització en vinculació amb el negoci i, per tant, gestionar aquest cicle. Això ens porta a definir el concepte següent.

Entenem com **gestió del cicle de vida de la dada** l'enfocament basat en polítiques per gestionar el flux de la dada mitjançant un o diversos sistemes d'informació al llarg del seu cicle de vida.

Una cosa que cal recalcar és que la dada és un actiu amb condicions molt definides que caracteritzen tot el seu cicle. Centrem-nos, per un moment, en la part final del cicle. La retenció de dades dins de l'organització pot estar definida per obligacions legals, per la legislació vigent o pel valor que la dada subministra a una organització. En països com Espanya, cal tenir en compte la llei de conservació de dades relatives a les comunicacions electròniques i a les xarxes públiques de comunicacions. També cal recordar que tot i que els costos d'emmagatzematge són cada vegada menors, hi ha la necessitat d'identificar quan les dades han de ser descartades, mitjançant la seva completa eliminació o desplaçant-les a un emmagatzematge extern per, així, mantenir una mida manejable a l'entorn de dades dins de l'organització.

Lectura complementària

Llei 25/2007, de 18 d'octubre, de conservació de dades relatives a les comunicacions electròniques i a les xarxes públiques de comunicacions.
Font: BOE-A-2007-18243.

2.2. Com mesurar el valor de la dada?

Com hem comentat, considerar la dada com un actiu obre la porta a preguntar-se quin valor té aquest actiu. Gestionar la dada no només consisteix a saber on és la dada, com ha evolucionat o què significa. També cal saber el valor que aporta una dada a l'organització i com mesurar-ne aquest valor.

Tot i que hem dit que de vegades pot resultar complicat determinar aquesta mesura, introduïrem un punt de partida general. D'acord amb la teoria de la comptabilitat, el valor d'un actiu prové de dues fonts:

- Ús de l'actiu, cas en el qual parlem de valor en ús.
- Venda de l'actiu, cas en el qual parlem de valor d'intercanvi.

A partir d'aquestes fonts, s'han fet servir tres mètodes per calcular el valor de la dada:

- **Fonamentat en el cost:** el valor de l'actiu està basat en el preu pagat per adquirir l'actiu. Per exemple, quan es compra una base de dades a Crunchbase*.
- **Fonamentat en el mercat:** el valor de l'actiu es fonamenta en el preu que està disposat a pagar un tercer per adquirir-lo.
- **Fonamentat en la utilitat:** el valor de l'actiu està basat en el valor actual dels beneficis esperats futurs.

Cadascun d'aquests mètodes té les seves limitacions.

- El valor fonamentat en cost pot ser inferior o superior al valor real per a un conjunt de dades. Pensem en el valor de la interacció d'un usuari a la

Lectura complementària

Moody, D.; Walsh, P. (1997). *Measuring The Value Of Information: An Asset Valuation Approach*.

* Més informació a:
www.crunchbase.com

pàgina d'Amazon, on el que està succeint té valor tan sols durant un curt període de temps.

- El valor fonamentat en el mercat requereix més temps per poder determinar quin preu és correcte per a la dada.
- El valor fonamentat en la utilitat presenta la complexitat que cal mesurar els fluxos de caixa en el futur, cosa que suposa treballar amb hipòtesis.

Com calcular aquest valor s'ha convertit en un tema candent, en transformar-se la dada en un actiu de valor per crear avantatges competitius o noves fonts d'ingressos, tal com s'ha comentat en l'apartat 1. Segons Gartner*, cal ampliar les formes en les quals determinem el valor de la dada, i n'hi ha de dos tipus: fundacionals i financeres. Vegem-les:

1) Mesures fundacionals, amb focus en la millora de la gestió de la informació.

- **Valor intrínsec**: avalua el grau de correcció de la dada i si és completa. Es calcula mitjançant la fórmula:

$$IVI = Validesa * Completesa * (1 - Escasses) * Cicle de vida$$

on:

- Validesa és el percentatge de registres considerats correctes.
- Completesa és el percentatge del total de registres respecte a l'univers de potencial o supòsit de registres.
- Escasses és el percentatge del seu mercat o dels competidors que també, probablement, tenen aquestes mateixes dades.
- Cicle de vida és la utilitat raonable i utilitzable de qualsevol unitat (registre) de l'actiu d'informació (per exemple, en mesos).
- **Valor de negoci**: avalua si la dada és rellevant per a un propòsit específic. Es calcula mitjançant la fórmula:

$$BVI = \sum_{p=1}^n Rellevància_p * Validesa * Completesa * Oportunitat$$

on:

Lectura complementària

Laney, D. (2015). *Why and How to measure the value of your information assets*. Gartner.

*Gartner és una de les principals empreses de prospecció de mercat.

- Rellevància és com d'útil pot ser la informació (o és) per a un o més processos de negoci (0 a 1).
- Validesa és el percentatge de registres considerats correctes.
- Completesa és el percentatge del total de registres respecte a l'univers de potencial o supòsit de registres.
- Oportunitat és amb quina rapidesa es capturen les instàncies noves o actualitzades de les dades i amb què és possible accedir-hi.
- **Valor de rendiment:** avalua com afecta la dada els factors clau del negoci (en anglès, Key Performance Indicator, l'acrònim és KPI). Es calcula mitjançant la fórmula:

$$PVI = \left[\frac{KPI_i}{KPI_c} - 1 \right] * T/t$$

on:

- i és el nombre d'instàncies de procés de negoci que utilitzen el recurs d'informació (grup informat).
- c són les instàncies de procés empresarial que no utilitzen la informació (grup de control).
- T és la mitjana de vida útil de qualsevol instància de dades.
- t és el temps durant el qual es va mesurar el KPI.

2) Mesures financeres, amb focus en la millora del rendiment financer:

- **Valor del cost:** avalua el cost associat a la pèrdua de la dada. Es calcula mitjançant la fórmula:

$$CVI = \frac{\text{Cost procés} * \% \text{ Atribuïble} * T}{t} \left\{ + \sum_{p=0}^n \text{Ingressos perduts}_p \right\}$$

on:

- Cost procés és el cost anual dels processos que permeten capturar la dada.
- % Atribuïble és el percentatge atribuïble a la captura de la dada.
- T és la mitjana de vida útil de qualsevol instància de dades.
- t és la durada del període sobre el qual es mesura el cost del procés.

- n és el nombre de períodes de temps fins que cal tornar a adquirir, o que la continuïtat de negoci ja no està afectada per la pèrdua d'informació.
- **Valor del mercat:** avalua el benefici que es podria obtenir en vendre o intercanviar la dada. Es calcula mitjançant la fórmula:

$$MVI = \frac{\text{Preu exclusiu} * \text{Nombre de } partner}{\text{Prima}}$$

- **Valor econòmic:** avalua com la dada contribueix al resultat final. Es calcula mitjançant la fórmula:

$$EVI = (\text{Ingressos}_i - \text{Ingressos}_c - \text{Costos}) * T/t$$

on:

- Ingressos_i són els ingressos generats amb la dada com a actiu.
- Ingressos_c són els ingressos generats sense la dada com a actiu.
- T és la mitjana de vida útil de qualsevol instància de dades.
- t és la durada de l'experiment.

2.3. Fases del cicle de vida de la dada

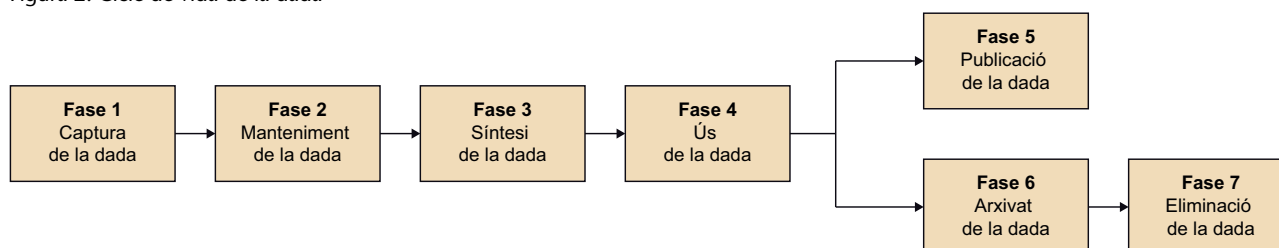
En el subapartat 2.1., hem comentat que la dada és un actiu diferent d'altres actius de l'organització, ja sigui per la seva naturalesa, la complexitat o fins i tot pel fet que sigui una metadada.

Per comprendre la dada i el seu valor, hem de fer-ho al llarg del seu cicle de vida, que té diferents fases, cadascuna de les quals té les seves pròpies característiques, com s'il·lustra a la figura 2.

Lectura complementària

Curto, J. (2015).
Fundamentos de big data.
UOC.

Figura 2. Cicle de vida de la dada



A continuació, entrarem en detall en cadascuna d'aquestes fases:

- **Fase 1. Captura de la dada.** Aquesta fase es pot considerar com l'acte de crear dades que no existeixen encara en l'organització i que mai no hi han existit. Trobem diferents mètodes per a la captura de dades, entre els quals destaquen:

- 1) **Adquisició de dades:** la ingesta de dades que han estat creades i existeixen fora de l'organització.
- 2) **Introducció de dades:** la creació de noves dades en l'organització per personal humà, o generades mitjançant dispositius dins de la pròpia organització.
- 3) **Recepció de senyals:** la captura de dades creades per dispositius, normalment en sistemes de control, però cada vegada més importants per als sistemes d'informació com l'internet de les coses.

Els tres mètodes identificats comporten importants reptes associats. Per exemple, l'adquisició de dades sol implicar la creació de contractes o normes que defineixen la manera en què l'organització pot fer servir les dades que obtingui de l'exterior. La introducció de dades implica una gestió detallada de la seguretat de la dada dins de l'organització, és a dir, qui té permisos per crear o modificar dades en l'organització. Finalment, un repte associat a la recepció de senyals és com captar i processar aquestes dades, les quals poden ser de qualsevol naturalesa i grandària, com ara les dades generades en temps real pels sensors de control en els jaciments petrolífers.

- **Fase 2. Manteniment de la dada.** L'objectiu d'aquesta fase és processar la dada, però sense que es generi encara un valor clar per a l'organització. Aquest processament inclou tasques d'integració, de neteja, d'enriquiment, i també de processos d'extracció, transformació i càrrega de la dada (conegudes en anglès com ETL - *extract, transform, and load*). A causa de la diversitat d'activitats en aquesta fase, hi ha nombrosos reptes associats, com per exemple, com millorar el procés d'enviament de la dada a la destinació final per la seva síntesi i el seu ús, prevenint que es generi un elevat nombre de moviments de dades durant tot el processament, d'inici a fi.
- **Fase 3. Síntesi de la dada.** Aquesta fase consisteix en la creació de dades de valor aplicant un processament o lògica inductiva determinada, fent servir altres dades com a font. Aquesta és l'àrea del processament analític on es fa servir el modelatge de dades, com, per exemple, el model de riscos d'una organització. La lògica inductiva requereix algun tipus d'experiència o coneixement com a part de la lògica de negoci, com ara la forma en què es creen els informes de crèdits bancaris. No obstant això, la lògica deductiva

Internet de les coses

És un concepte que es refereix a la interconnexió digital d'objectes quotidians amb internet. En anglès, *Internet of things*, abreujat IoT.

no formaria part d'aquesta fase, i quedaria relegada a la fase anterior, com crear un nou valor aplicant una equació simple:

$$\text{Valor net} = \text{Valor brut} - \text{Impostos associats}$$

- **Fase 4. Ús de la dada.** Una vegada que la dada ha estat capturada i transformada dins de l'organització i s'ha fet servir com a font a la fase anterior, s'utilitza per a benefici de la organització mateixa, en tasques que l'organització utilitza i gestiona. Encara que normalment es tracta de tasques que no pertanyen al cicle de la vida de la dada, la dada és cada vegada més considerada un actiu fonamental en els models de negocis de moltes organitzacions. En aquesta fase, també hi ha reptes importants, com per exemple l'ús permès de la dada o, el que és el mateix, si és legal o no l'ús de la dada de la manera en què els usuaris de negoci pretenen utilitzar-la.
- **Fase 5. Publicació de la dada.** Aquesta fase pot ser definida com l'enviament de la dada. Aquest enviament pot ser intern (a una intranet) o extern (a un lloc fora de l'organització, gestionat per tercers). Un exemple és una agència d'inversió que envia informes mensuals als seus clients. Una vegada que la dada ha estat enviada fora de l'organització, és impossible recuperar-la per a una posterior modificació. El govern de la dada ajuda a decidir com actuar amb les dades incorrectes o incompletes que han estat enviades fora de l'organització. Els accessos no autoritzats a les dades també s'inclourien en aquesta fase.
- **Fase 6. Arxivat de dades.** Aquesta fase consisteix a copiar les dades en un entorn on són emmagatzemades, en cas que es tornin a necessitar en el futur en un entorn actiu de producció, i també inclou la completa eliminació d'aquestes dades en tots els entorns actius. Un arxiu de dades és simplement un emmagatzematge de dades, però no de manteniment, d'ús o de publicació de les dades. En cas necessari, les dades poden ser recuperades en un entorn on és possible dur a terme qualsevol d'aquestes activitats.
- **Fase 7. Eliminació de la dada.** Aquesta és la fase final del cicle de vida de la dada, i consisteix en la destrucció de qualsevol còpia de la dada que hi hagi dins de l'organització. Idealment, aquest procés serà dut a terme mitjançant un arxivat de dades. Un repte, en aquesta fase, seria verificar que l'eliminació s'ha dut a terme de manera satisfactòria.

2.4. Com és possible detectar els canvis en una dada?

La dada no és un actiu estàtic durant el seu cicle de vida. Dins de les organitzacions, es creen noves fonts de dades contínuament i cal mantenir un registre de la dada al mateix temps que es mou per mitjà de diferents sistemes dins de

l'organització. Per això, necessitem establir el que es coneix com traçabilitat de la dada o llinatge de la dada.

S'entén com **llinatge de la dada** la capacitat de conèixer tot el cicle de vida d'una dada, des de la data i hora exacta en què va ser extreta, el moment en què es va produir la seva transformació, i fins a l'instant en què va tenir lloc la seva càrrega des d'un entorn font (servidor, fitxer, taula, camp, etc.) a un altre de destinació. En anglès, *data lineage*.

Mitjançant el llinatge de la dada, les organitzacions poden obtenir una millor comprensió del que succeeix amb la dada, el que en possibilita el rastreig i la identificació d'errors i, així, aplicar protocols més rigorosos de govern de la dada.

El llinatge de la dada proveeix als professionals de la dada una representació visual que permet visualitzar el flux de la dada. D'aquesta manera, es pot identificar com i quan la dada és modificada en l'organització. Per exemple, identificar els canvis efectuats en la dada pels diferents processos d'extracció, transformació o càrrega definits en l'arquitectura de dades de l'organització. Cal també destacar que, gràcies a la capacitat de monitoritzar la dada d'una manera continuada, els errors relatius a la dada poden ser detectats abans que apareguin i, per tant, corregits.

Un escenari d'ús comú del llinatge de la dada es dona en l'àrea de la intel·ligència de negoci (*business intelligence*). Per exemple, aquesta capacitat mostra com s'ha obtingut certa informació relativa al negoci i quin paper pot exercir en els diferents mètodes disponibles d'integració de dades en l'organització. Un altre escenari habitual té a veure amb la reducció de riscos i la protecció de dades. Els professionals de la dada poden fer servir el llinatge de la dada per gestionar d'una manera òptima la dada i, en tot moment, controlar on es troben les dades sensibles, per evitar així la seva exposició o minimitzar els efectes d'una possible violació a la seguretat en la organització.

Business intelligence

És el conjunt de metodologies, aplicacions, pràctiques i capacitats enfocades a la creació i administració d'informació, que permet als usuaris d'una organització de prendre decisions millors.

Altres casos d'ús inclouen:

- **Resolució d'errors o conflictes:** en la creació d'informes, la traçabilitat de la dada permet saber com s'han construït les mètriques que s'hi inclouen, quines transformacions s'han fet i d'on provenen.
- **Anàlisi de l'impacte:** en el desenvolupament i evolució de sistemes d'informació, el llinatge de la dada habilita comprendre per endavant què passarà amb la dada i quines mesures s'han de tenir en compte.
- **Informes de conformitat:** en certs sectors, s'han imposat regulacions i normes per a una transparència més gran. *Data lineage* habilita reduir els

errors humans i les bretxes en la conformitat respecte a la normativa vigent.

En essència, el llinatge de la dada és un pas necessari, però no suficient, cap a la gestió eficient de la dada.

3. Govern de la dada

Encara que pugui semblar el contrari, la gestió de la dada (i en la seva forma més moderna, el govern de la dada) no és un concepte nou, existeix des de fa més de vint anys. Per poder entendre la seva forma més moderna, cal comprendre els seus orígens.

3.1. Orígens del *data governance*

Com a concepte, el govern de la dada ha existit des de fa dècades i ha passat per diferents èpoques.

Des dels inicis de les tecnologies de la informació fins a finals dels anys vuitanta, les organitzacions van anar adoptant progressivament múltiples sistemes d'informació, de cara a suportar processos de negoci i reduir les tasques manuals. No obstant això, la dada era considerada un subproducte de poc valor, amb poca necessitat de ser compartida i gestionada. A les poques organitzacions en què se'n va identificar la necessitat, van mirar de governar la dada mitjançant el modelatge de dades corporatives, però amb escàs èxit, degut principalment a dos motius: d'una banda, la manca de suport en un àmbit corporatiu, ja que només el departament d'informàtica estava implicat en aquestes tasques; i, d'altra banda, la rigidesa de les aplicacions disponibles en l'època. En resum, en aquesta època el govern de la dada era considerat principalment un exercici merament acadèmic.

Durant les dues dècades següents, moltes organitzacions van començar a reconèixer que el valor de la dada anava més enllà de les transaccions. Es va iniciar en moltes organitzacions el llarg procés de prendre decisions fonamentades en dades. A poc a poc, també es va començar a identificar la necessitat de creuar les dades de diferents unitats de negoci i departaments.

En una primera instància, les organitzacions van abordar aquestes necessitats construint repositoris de dades consolidades, ja fos mitjançant sistemes ERP* o *data warehouse*. La implementació d'aquests sistemes, tot i que proporcionava clars beneficis, no estava exempta de reptes i problemes, com el seu cost, la durada i els conflictes generats per haver de respondre a múltiples necessitats de negoci al mateix temps. En aquesta època, també hem de destacar la caiguda de companyies com Enron, que va fer que el Govern dels Estats Units establís noves lleis per millorar la certesa i la fiabilitat de la informació dins de les organitzacions. En particular, ens referim a la Llei Sarbanes-Oxley, en an-

*ERP és l'acrònim de *enterprise resource planning*.

glès *Sarbanes-Oxley Act of 2002*, Pub. L. núm. 107-204, 116 Stat. 745, la finalitat de la qual és evitar frauds i riscos de fallida, per protegir l'inversor.

Per aquestes raons, es va començar a identificar el govern de la dada com a crític per al negoci. De fet, va ser un component clau en les mesures creades per la Llei Sarbanes-Oxley i altres reglaments, els quals obliguen els executius de les organitzacions a conèixer i ser personalment responsables de les dades que utilitzen en els seus negocis. El govern de la dada va començar a madurar ràpidament i, com a resultat, la tecnologia desenvolupada va començar a possibilitar l'automatització a l'hora de crear i gestionar lleis de negoci en un àmbit de dada. No obstant això, el govern de la dada estava habitualment centrat en diferents repositoris aïllats; per exemple, el govern de la dada per a un *data warehouse* o un sistema ERP. A més, al principi era un govern de la dada informal, que no tenia una estructura organitzativa ni processos ben definits.

En l'actualitat, els executius estan començant a tenir clara la necessitat de gestionar la dada com un actiu per a l'organització. Hi ha una proliferació de més i més regulacions vinculades amb la dada. Juntament amb l'explosió de la complexitat de la dada (*big data*), la transformació digital empeny a considerar formes més sofisticades d'explotar els processos de negoci i les dades (per exemple, mitjançant *machine learning*, conegut en català com aprenentatge).

Aquest escenari empeny les organitzacions a replantejar-se com gestionar la dada, en necessitar noves estructures d'emmagatzematge, de processament i d'anàlisi, i com afecten aquestes estructures als models de dades, a la qualitat, a la seguretat i al cicle de vida. Això ha portat a l'aparició de plataformes integrals de govern de la dada, en què es defineixen polítiques de dades en un àmbit empresarial.

El canvi actual també es fonamenta en el fet que aquestes iniciatives no són solament responsabilitat del departament de TI*, que aporta consistència i control, sinó també del negoci, que incrementa la seva responsabilitat sobre el dada i el reconeixement del valor que pot aportar. En alguna d'elles (com és el cas de Telefónica o Barclays), fins i tot hi ha un paper especialitzat en el consell de direcció, l'anomenat *chief data officer* (CDO).

3.2. Definició de *data governance*

La creixent necessitat de gestionar de forma eficient la dada ha generat el fet que múltiples institucions hagin dedicat recursos a formalitzar la definició del govern de la dada, a més de a definir els seus components. Això ha provocat que hi hagi nombroses definicions del concepte. Algunes s'exposen a continuació:

Llei Sarbanes-Oxley

És una llei dels Estats Units, també coneguda com la Llei de reforma de la comptabilitat pública d'empreses i de protecció a l'inversor, que busca monitoritzar les empreses que cotitzen en borsa de valors, i evitar que la valorització de les seves accions sigui alterada de manera dubtosa, mentre que el seu valor és menor.

Big data

És el conjunt d'estratègies, tecnologies i sistemes per a l'emmagatzematge, processament, anàlisi i visualització de conjunts de dades complexes.

Machine Learning

És un subcamp de la intel·ligència artificial amb focus en la creació d'algoritmes amb capacitats d'aprenentatge a partir de dades i dels resultats buscats.

*TI és l'acrònim de «tecnologies de la informació».

- Segons *The MDM Institute**, és la coordinació formal de persones, de processos i de tecnologia que permet a una organització utilitzar les dades com a actiu d'empresa.
- Segons l'Institut de Govern de la Dada**, és un sistema de decisions i responsabilitats sobre processos d'informació, executat d'acord a models preestablerts que descriuen qui pot prendre quines decisions, amb quina informació i quan, en quines circumstàncies i mitjançant quins mètodes.
- Segons Kik Consulting***, és l'execució i l'aplicació d'autoritat sobre la gestió d'actius de dada i el rendiment de les funcions de la dada.
- Segons el IBM Data Governance Council****, és una disciplina de control de qualitat per assessorar, gestionar, utilitzar, millorar, monitoritzar i protegir informació de l'organització.
- Segons Sunil Soares, expert en *data governance*, és l'elaboració de normes per optimitzar, protegir i fer ús de la informació com un actiu de l'empresa, mitjançant l'alineament d'objectius de múltiples funcions.

*Més informació a:
www.the-mdm-institute.com

**Més informació a:
<http://www.datagovernance.com>

***Més informació a:
<http://www.datagovernance.com>

****Més informació a:
<http://www.ibm.com>

En el present material, considerarem la definició de DAMA***** publicada al *Data Management Body of Knowledge (DMbok)*:

*****Més informació a:
<http://www.dama.org>

S'entén com **govern de la dada** (en anglès, *data governance*) l'exercici d'autoritat, de control i de presa de decisions compartida (planificació, vigilància i aplicació) sobre la gestió dels actius de dades.

És important destacar que totes les definicions són similars, i estableixen que el govern de la dada tracta sobre els processos que controlen la creació, l'accés, l'ús compartit, la utilització i la destrucció de la informació, i què passa quan hi ha un conflicte referent a això.

3.3. Objectius del *data governance*

Així doncs, el govern de la dada és un conjunt de normes i pràctiques ben definides per assegurar que la dada sigui:

1) **Accessible**, que significa que:

- a) L'usuari pot accedir a les dades que necessita i aquestes estan disponibles.
- b) El format de la dada coincideix amb el qual l'usuari demanda.

2) **Segur**, que significa que:

- a) Només poden accedir a la dada els usuaris autoritzats.
- b) S'impedeix l'accés a la dada als usuaris no autoritzats.

3) **Consistent**, que significa que:

- a) La informació és la mateixa quan diferents usuaris accedeixen a la mateixa dada.

4) **Alta qualitat**, que significa que:

- a) La dada és rigorosa i precisa.
- b) Compleix amb els estàndards preestablerts.

5) **Auditable**, que significa que:

- a) La traçabilitat de la dada és clara.
- b) El departament de TI sap qui fa servir la dada i per a quins propòsits.

L'objectiu final del govern de la dada no és solucionar tots els problemes de TI de l'organització, sinó centrar-se en tots els aspectes vinculats amb la dada. Per això, els principals objectius del govern de la dada inclouen:

- Definir, validar i comunicar les estratègies de la dada, normes, estàndards, arquitectura, procediments i mètriques.
- Monitoritzar les polítiques de la dada, els estàndards, l'arquitectura i els procediments.
- Patrocinar i supervisar el lliurament de projectes de gestió de dada i serveis.
- Gestionar i resoldre conflictes relacionats.
- Entendre i promocionar el valor de l'actiu de la dada.

En aquest sentit, implementar el govern de la dada en una organització comporta treballar en els punts següents:

- **Estratègia.** Desenvolupar una estratègia, és a dir, decidir quines dades es gestionaran; identificar les dades mestres. Determinar el valor de la dada basat en el cost de la captura, el manteniment, el valor de negoci, el risc de pèrdua o la certesa de la dada.
- **Comissió.** Establir una comissió de les línies de negoci que estan en la posició d'entendre la dada, d'entendre el negoci i de ser capaços de prendre una decisió.
- **Normes.** Establir un conjunt de normes per definir la integritat de les dades, la qualitat, la seguretat i l'utilització de la dada.

- **Estàndards.** Establir un conjunt d'estàndards per controlar la implementació de les normes, el modelat de les dades, les eines, les tecnologies i les metodologies.
- **Procediments.** Definir procediments per tractar conflictes de qualitat de la dada i de seguretat.
- **Conflictes.** Gestionar els conflictes i la seva resolució.
- **Supervisió.** Oferir una supervisió continuada de la gestió del projecte.
- **Incompliment.** Establir un conjunt de penalitzacions per incompliment dels estàndards i dels procediments.
- **Documentació.** Proporcionar documentació relativa a la dada i als requisits de metadades.

El govern de la dada no reemplaça la gestió de la dada, sinó que la complementa. Quan diferents usuaris no poden o han de prendre decisions independentment en l'organització, el govern de la dada facilita que equips que ocupen diferents funcions en el negoci s'uneixin per crear normes comunes, solucionar conflictes o donar nous serveis a les persones involucrades. La taula 1. presenta les diferències principals entre els dos conceptes.

Taula 1. Diferències entre govern i gestió de la dada

Dimensió	Govern de la dada	Gestió de la dada
Abast	Presa de decisions, monitoratge i execució que té l'autoritat en la gestió de la dada d'una organització.	Control de l'arquitectura de dades, qualitat, seguretat, normes, pràctiques i procediments.
Funció	Decidir què fer amb la dada i assegurar que es faci allò que s'ha decidit.	Implementar les arquitectures, processos, eines i normes per assolir els objectius del govern de la dada.

Gestió de la dada

És el desenvolupament i la implementació d'arquitectures, normes, pràctiques i procediments per gestionar eficaçment les necessitats del cicle de vida de les dades en una organització.

És bastant comú en les organitzacions el fet que diferents grups d'usuaris creïn diferents ubicacions de dades no sincronitzades entre elles. Això sol donar lloc a una pèssima eficiència operacional i un alt risc, cosa que produeix la necessitat que hi hagi una autoritat que resolgui els conflictes generats. Aquesta és la funció del govern de la dada. Per exemple, si hi ha multitud de repositoris sobre dades de clients, caldrà prendre decisions, sovint impopulars, per crear un repositori de dades mestres.

3.4. Fases del data governance

El disseny d'un programa de govern de la dada no és necessàriament una seqüència lineal. De fet, és probable que es prenguin decisions sobre cadascuna de les fases del programa a partir d'altres fases. Per tant, tot i que a continuació es detallen les fases en una seqüència lineal, s'haurà de tenir en compte un enfocament holístic a l'hora de fer el disseny.

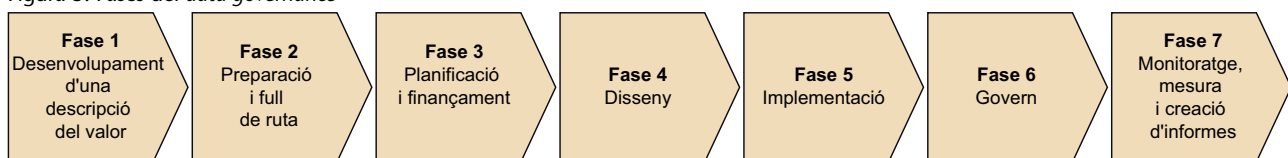
Segons l'Institut del Govern de la Dada, el cicle de vida del govern de la dada consta de set fases, com es mostra a la figura 3:

- 1) Desenvolupament d'una descripció del valor.
- 2) Preparació i full de ruta.
- 3) Planificació i finançament.
- 4) Disseny.
- 5) Implementació.
- 6) Govern.
- 7) Monitoratge, mesura, creació d'informes.

Full de ruta

Fa referència al pla detallat que reflecteix els objectius a curt i llarg termini i guia el progrés per assolir aquests objectius, fent servir, sovint, una solució tecnològica específica.

Figura 3. Fases del *data governance*



Font: Institut de Govern de la Dada

Com prerequisits, abans de crear el disseny del programa del govern de la dada hem de tenir en compte el següent:

- El missatge del valor del programa ha de ser clar.
- Hi ha d'haver un full de ruta que es proporcionarà a les persones involucrades, i que en detalli els diferents aspectes; en particular, ha de donar resposta a les preguntes qui, què, quan, on i per què.

Cal explicar el detall de les diferents fases:

Fase 1. Desenvolupament d'una descripció del valor. La primera fase se centra en definir el valor que el govern de la dada genera al negoci, a més dels estàndards per a un mesurament continuat després de la posada en marxa del programa. En aquesta fase, es revisa acuradament l'estratègia i els objectius del negoci i es defineix i es desenvolupa una connexió entre el govern de la dada i una millora en l'organització identificable des d'un aspecte financer. Per identificar el valor que el programa pugui generar al negoci, cal identificar què passa dins de l'organització en termes de la gestió de la dada, ja que és possible que hi hagi algun programa en execució, per exemple, relatiu a la qualitat de la dada o la gestió de dades mestres, ja com el govern de la dada suportarà el negoci, directament o indirectament per mitjà d'aquests programes.

A continuació, cal definir quin serà el criteri d'èxit per al programa del govern de la dada, és a dir, proveir uns estàndards de mètriques i mesures sobre

el tema que permetin identificar-ne l'èxit o el fracàs. Per exemple, es poden identificar mètriques financeres que permetin saber quin seria el cost d'una manca de govern de la dada o, el que és el mateix, l'ús de la informació d'una manera pobra i ineficaç dins de l'organització, en comparació d'un programa governat.

Fase 2. Preparació i full de ruta. En aquesta fase, es planifiquen els detalls sobre els esdeveniments al voltant de la data de llançament del govern de la dada, moment en el qual els actius de dades d'una organització passen de no estar governats a estar-ho. Sovint, el full de ruta implica un pla que es desenvolupa durant diversos anys, i que integra activitats amb altres projectes o iniciatives que s'executen dins de l'organització. Cal incloure punts de control freqüents i espais per recollir crítiques constructives al llarg del pla i, d'aquesta manera, modificar-lo si fos necessari i no arribar a un estat de bloqueig durant l'execució del pla. Les activitats d'aquesta fase impliquen:

- Integració del govern de la dada amb altres àrees. Atès que el govern de la dada està relacionat amb multitud d'àrees dins de l'organització, serà necessari tenir en compte les possibles relacions en el full de ruta. Per exemple, integració del govern de la dada amb la gestió de dades mestres dins d'una organització.
- Disseny de mètriques i requisits d'informes. Necessari per validar el que s'està gestionant i rebre informació per assegurar la continuïtat del pla.
- Definició dels requisits de manteniment o gestió del canvi. Cal planificar els possibles canvis que puguin ocórrer per poder gestionar-los. Així, s'identifiquen els possibles elements que tenen lloc durant una futura posada en execució o llançament del programa i es determina com estan coordinats entre ells.
- Definició del llançament del govern de la dada. S'especifiquen els detalls relatius a la posada en marxa del programa, incloent-hi informació sobre el personal relacionat que el gestionarà.

Fase 3. Planificació i finançament. El programa del govern de la dada sovint comporta una inversió significativa de temps i de recursos, tant humans com financers, per part de l'organització. No obstant això, aquests costos es compensen pel valor per al negoci que proporciona el programa. Mitjançant una descripció de valor ben definida, uns objectius de negoci clars i un suport o patrocini en un àmbit executiu, el programa del govern de la dada pot ser satisfactòriament executat i proveir un retorn d'inversió significatiu.

Fase 4. Disseny del programa. En aquesta fase, s'especifiquen i desenvolupen els detalls sobre com serà utilitzat el programa i la funcionalitat que tindrà. El resultat final d'aquesta fase són els principis, les normes i el disseny dels processos que definiran el programa. Les principals activitats que tenen lloc en aquesta fase són:

- Determinació dels procediments o principis centrals. Aquesta activitat resulta imprescindible per a l'èxit del programa, i consisteix a identificar, documentar i aprovar els procediments clau que seran utilitzats dins de l'organització per considerar la dada com a un actiu.
- Determinació dels processos estàndard per al negoci. Durant aquesta etapa, es determinen els processos genèrics que defineixen els objectius del programa del govern de la dada. Per exemple, com reacciona el negoci davant d'un procés d'identificació o de detecció d'incidències i les accions que es duen a terme en el procés per a la seva resolució.
- Determinació de les responsabilitats i el model de propietat. Així doncs, tots els processos definits prèviament no tenen utilitat fins que s'ha definit qui fa què i els diferents nivells de responsabilitat presents.
- Exposició del model de govern de la dada a l'equip executiu del negoci. D'aquesta manera, s'informa i educa l'equip de gestió de l'organització sobre les diferents responsabilitats i processos.

Fase 5. Implementació del programa. Aquesta fase representa l'execució de les activitats que s'han definit en el pas previ. És important destacar que aquesta fase no implica que hi hagi un definit començament i final de fase, ja que, en essència, una vegada que s'ha començat a suportar el govern de la dada, aquest procés mai no s'acaba. Les activitats principals d'aquesta fase són:

- Llançament (en anglès, *rollout*) del govern de la dada. Com el seu nom mateix indica, a partir d'aquest moment l'organització comença a governar les dades. Alhora, diferents equips dins de l'organització són educats i formats segons el que s'ha acordat en el full de ruta, i es publiquen diferents materials creats durant el disseny del programa, com ara guies i normes.
- Execució de canvis en el pla. En aquesta etapa, es recull tota l'activitat relativa al suport del govern de la dada. Per exemple, es pot afirmar que, durant l'execució del programa, diferents materials, com les guies o els materials de formació, necessitaran una actualització. Així mateix, el nou personal necessitarà una nova orientació sobre el programa implementat.

Fase 6. Govern de la dada. Una vegada que el disseny funcional està definit, el pas següent és plasmar aquest disseny en un marc organitzatiu. En aquesta fase, s'identifica el personal relatiu al programa del govern de la dada, com ara administradors o propietaris de la dada, i també algunes àrees o usuaris que tenen diferents activitats i rangs de responsabilitats relatius al govern de la dada. En aquesta fase, les activitats es basen en:

- Disseny del marc organitzatiu. És a dir, determinar on i quins nivells del marc són els encarregats d'executar, gestionar i ser responsables de gestionar els actius de dades.

- **Identificació de rols i responsabilitats.** Aquesta tasca, encara que sembli trivial, pot suposar obstacles dins de l'organització, com ara una percepció de donar poder sobre les dades a determinats usuaris que pot portar a problemes interns dins el marc organitzatiu, o la por de perjudicar la productivitat actual en afegir responsabilitats addicionals.

Fase 7. Monitoratge, mesura i creació d'informes. El programa del govern de la dada necessita ser examinat i revisat per saber si és eficaç per a l'organització després de la seva implementació. Normalment, un grup intern dins de l'organització durà a terme aquestes comprovacions, les quals impliquen la recollida de dades i la generació de mètriques que informen sobre diferents aspectes, com l'efectivitat de normes i estàndards, a més de l'activitat d'usuaris definits en la organització, com ara administradors de la dada.

4. Components i maduresa del *data governance*

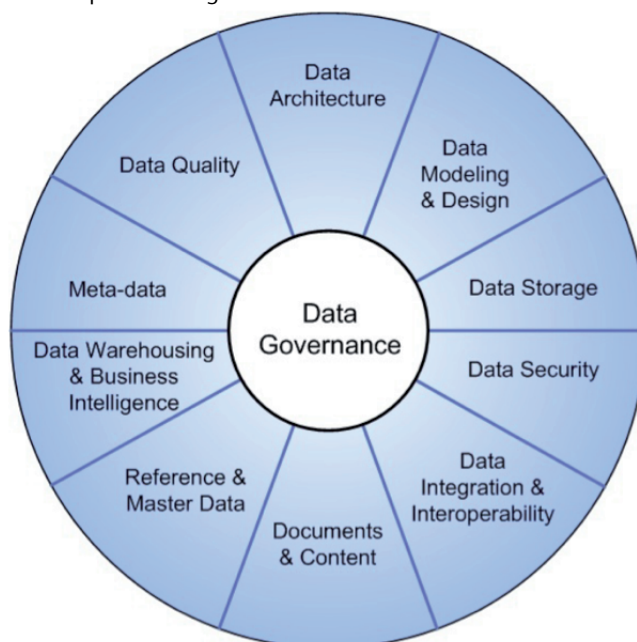
El govern de la dada és una àrea molt gran. Inclou la gestió d'aplicacions de negoci (des de la perspectiva de la dada), del govern TI (com la gestió de les bases de dades) i la gestió de la informació. En aquests materials, ens centrarem en aquells components de major rellevància per a l'anàlisi de la dada. Així mateix, discutirem com progressar el govern de la dada mitjançant un model de maduresa.

4.1. Components

Tal com hem comentat, el govern de la dada és un exercici coral. D'una banda, les diferents línies de negoci són responsables d'optimitzar l'ús i la comprensió de la informació; de l'altra, el departament de TI és responsable d'optimitzar l'accés i la protecció de la informació; i finalment, l'estructura organitzativa que suporta el govern de la dada és responsable d'optimitzar la disponibilitat i utilitat de la informació.

Aquestes responsabilitats s'entrellacen i es despleguen en les diferents àrees de la gestió de la informació, com es mostra a la figura 4.

Figura 4. Components del govern de la dada



Font: DAMA - DMBok

És a dir, l'enfocament actual del govern de la dada inclou múltiples funcions. Segons DAMA, aquestes funcions o components són:

- **Arquitectura de dades** (*data architecture*): consisteix en les estructures d'emmagatzematge i processament de la dada com a part de l'arquitectura empresarial.
- **Disseny i models de dades** (*data modeling and design*): consisteix en la gestió, l'anàlisi, el disseny, la construcció, el testeig i el manteniment dels models de dades.
- **Emmagatzematge de dades** (*database storage*): consisteix en la gestió de les infraestructures físiques d'emmagatzematge de dades.
- **Seguretat de dades** (*data security*): consisteix a assegurar la privacitat, la confidencialitat i l'accés en els casos apropiats.
- **Dades mestres i de referència** (*reference and master data*): consisteix en la identificació, el manteniment, la gestió, l'accés i la propagació de dades mestres i de referència.
- **Intel·ligència de negoci i magatzems de dades** (*data warehousing and business intelligence*): consisteix en la gestió dels processos analítics de l'organització i de l'accés de la forma correcta a la informació per a la presa de decisions.
- **Integració i interoperabilitat de dades** (*data integration and interoperability*): consisteix en la gestió de l'adquisició, l'extracció, la transformació, el moviment, la propagació, la replicació, la federació i la virtualització de dades.
- **Contingut i documents** (*documents and content*): consisteix en l'emmagatzematge, la protecció, l'indexació i l'habilitació d'accés de dades en documents, permetent la seva integració i interoperabilitat amb dades estructurades.
- **Metadades** (*metadata*): consisteix en la recopilació, categorització, manteniment, integració, control, gestió i distribució de metadades.
- **Qualitat de dada** (*data quality*): consisteix a definir, monitoritzar, mantenir i millorar la qualitat i la integritat de la dada.

I per sobre de totes elles, el govern de la dada, encarregat de la planificació, la vigilància i l'aplicació de la gestió i de l'ús de la dada.

D'aquesta manera, una plataforma tecnològica de govern de la dada ha de cobrir les funcions descrites en major o menor mesura. Freqüentment, és possible trobar plataformes de gestió de dades mestres i de qualitat de dades de manera independent. I amb més freqüència, la funció de la seguretat de dada cau sota la responsabilitat del departament de seguretat tecnològica. Així, avançar cap a un govern de la dada requereix coordinar totes aquestes iniciatives independents.

4.1.1. Normes, estàndards i estratègia

Normalment, constitueixen el pas inicial quan una organització inicia un procés de govern de la dada. Les principals activitats del govern de la dada amb focus en normes, estàndards i estratègia inclouen:

- identificar el personal involucrat, establiment de decisions i aclariment de responsabilitats;
- establir, revisar, aprovar i monitoritzar normes;
- establir, revisar, aprovar i monitoritzar estàndards;
- establir estratègies de dades corporatives.

Ens centrarem de manera detallada en tres àrees: la gestió de la qualitat de la dada, la gestió de dades mestres i la gestió de la seguretat de la dada.

4.1.2. Qualitat de la dada

Una part rellevant del govern de la dada és la necessitat de millorar la qualitat de la dada. Habitualment, la qualitat s'aplica a les dades mestres i hi involucra algun programari de gestió o neteja de dades. Les principals activitats del govern de la dada amb focus en la qualitat de la dada inclouen:

- identificar el personal involucrat, establiment de decisions i aclariment de responsabilitats;
- definir l'estratègia de la qualitat de la dada;
- monitoritzar la qualitat de la dada;
- informar sobre l'estat de les iniciatives de qualitat.

4.1.3. Gestió de dades mestres

Una altra de les funcions rellevants del govern de la dada és l'explotació i la gestió eficient de la informació. Aquesta eficiència es fonamenta en el fet que no hi ha visions divergents respecte a la dada, i es promou que hi hagi comunicació de les dades entre les diverses àrees de l'empresa. Les principals activitats del govern de la dada amb focus en la gestió de dades mestres inclouen:

- identificar les dades mestres i els de referència, tant interns com externs;
- consolidar i propagar l'ús de les dades mestres i els de referència;
- assegurar la qualitat de les dades mestres i de referència.

4.1.4. Gestió de la seguretat de la dada

En el context del govern de la dada, parlar de la gestió de la seguretat inclou també la privacitat i el compliment. No solament es tracta de protegir la dada, sinó de complir amb els requisits dels reguladors en el sector i el país en què s'opera. Aquesta funció inclou tecnologies per identificar informació sensible, protegir la dada, gestionar la seguretat i l'accés i també normes d'auditoria i control. Una arquitectura de seguretat de dades corporativa és essencial per protegir totes les dades sensibles en ús i emmagatzemats en l'organització. Les principals activitats d'aquesta funció inclouen:

- ajudar a protegir les dades sensibles mitjançant la gestió de l'accés i els requisits de seguretat;
- alineament entre l'arquitectura de seguretat de dades i iniciatives de negoci;
- ajudar a assessorar i gestionar el risc;
- ajudar a fer complir els requisits reguladors i contractuals;
- identificar el personal involucrat, l'establiment de decisions i la clarificació de responsabilitats.

4.2. Maduresa

El desenvolupament d'un programa de govern de la dada no està exempt de problemes. Trobem múltiples raons perquè no arribi a bon port. Per exemple:

- falta de suport executiu, és a dir, falta d'un clar lideratge en l'organització;
- manca de fonaments de gestió de la dada en l'organització, és a dir, manca de persones amb coneixements en l'organització;
- manca de mètriques de seguiment del programa, és a dir, manca d'indicators i objectius per a comprendre l'avanç del programa;
- baixa planificació en el programa i projectes, i no disponibilitat de pressupost i/o temps adequats, és a dir, manca d'organització respecte a aquestes iniciatives;
- manca de perfils adequats, tant tècnicament com en un àmbit de lideratge i propietat de la dada.

Avaluar l'estat de l'organització respecte al govern de la dada pot ajudar a comprendre en quina situació es troba l'organització, i quins són els següents passos.

4.3. Model de maduresa del *data governance*

Un bon punt de partida per desenvolupar una estratègia efectiva del govern de la dada és identificar quin és el grau de maduresa de la organització respecte a aquest govern.

Com succeeix amb altres models de maduresa, el grau de maduresa està centrat en els processos que tenen lloc dins de l'organització, la seva implementació i la manera d'operació. Aquests processos habiliten que l'organització funcioni eficaçment, i en regeixen el funcionament. Per tant, resulta essencial que aquests processos estiguin implementats d'una manera efectiva i que busquin una millora de l'optimització de forma continuada.

Actualment, trobem diferents models de maduresa relatius al govern de la dada, els quals no sempre coincideixen en les seves característiques. Entre ells, destaquem:

- MIKE2.0, que és un marc de referència de bones pràctiques relatives a la gestió de la informació, i el model és l'*information maturity model*.*
- 2IBM Data Governance Council, format per més de cinquanta institucions que col·laboren en l'àrea i el model és el *Data Governance Council maturity model*.**
- EDM Council, que és una associació per al desenvolupament i la implementació d'estàndards relatius a la dada i la publicació de bones pràctiques, i el model és el *data management maturity model* (DCAM).***
- DAMA, una associació internacional de professionals focalitzats en els conceptes i pràctiques relatius a la gestió de la dada i de la informació, i el model és el *DMBoK (data management body of knowledge)*.****
- Gartner, consultora de prospecció tecnològica, el model és el *Enterprise Information Management maturity model*.*****

Encara que hi hagi una manca de consens entre els models existents, en general es fonamenten tots ells en el *capability maturity model* (CMM). Com a norma general, és possible descriure cinc nivells o fases diferents. Cadascun d'aquests nivells consta d'uns objectius i d'unes pràctiques que descriuen les característiques presents, en un àmbit de processos i activitats necessàries per gestionar i millorar els processos relatius.

Maduresa

Fa referència a una mesura de l'habilitat d'una organització per a la millora contínua respecte a una disciplina particular.

*Més informació a:
<https://goo.gl/zCALoL>

**Més informació a:
<https://goo.gl/rUQ12i>

***Més informació a:
<https://goo.gl/12yEFS>

****Més informació a:
<https://goo.gl/EKG9FE>

*****Més informació a:
<https://goo.gl/wVvYUh>

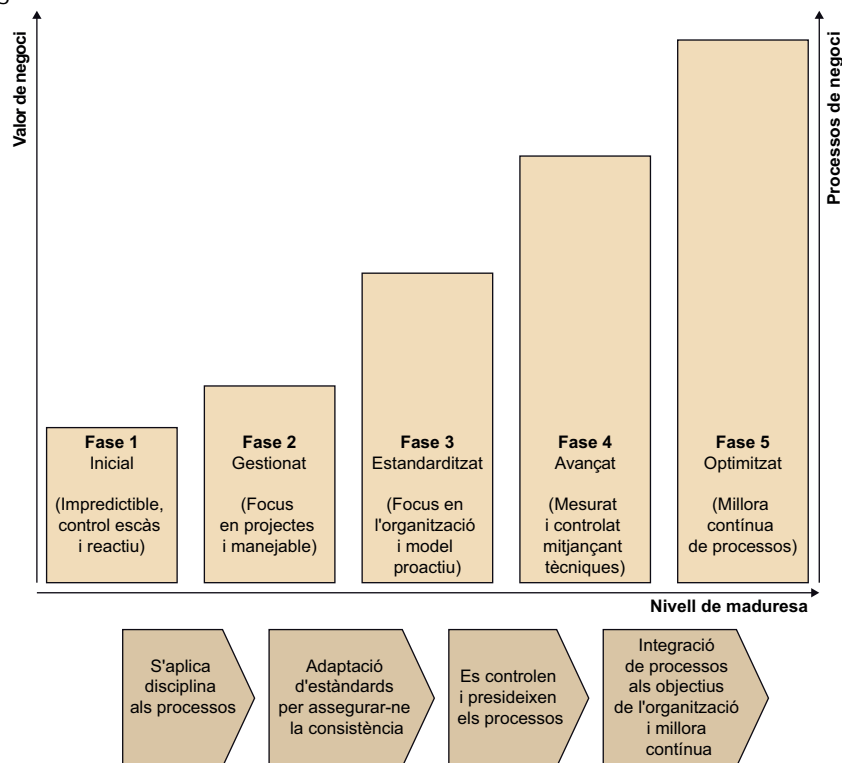
Això dona lloc a sis fases (incloent el cas que no hi ha absolutament res) respecte a la maduresa en el govern de la dada: cap, inicial, gestionat, estandarditzat, avançat i optimitzat. Així, com més alt sigui el nivell de maduresa de l'organització, més gran serà la seva eficiència i els seus costos i riscos seran menors.

- **Fase 0. Cap.** No existeix un procés de govern de la dada definit.
- **Fase 1. Inicial.** Hi ha una autoritat de la dada en el departament de TI, però té poca influència en els processos de negoci. La col·laboració entre àrees de TI i negoci no és consistent, i hi ha una total dependència de determinats experts de la dada en cada àrea de negoci. Com a resultat, els processos no estan integrats dins de l'organització.
- **Fase 2. Gestionat.** El cap i l'administrador solen estar presents en línies de negoci particulars. Hi ha processos poc definits en les aplicacions clau en les línies de negoci, i els problemes de la dada són gestionats d'una manera reactiva, sense identificar l'origen del problema. És una etapa primerenca a l'hora d'estandarditzar els processos en les diferents línies de negoci.
- **Fase 3. Estandarditzat.** El negoci està involucrat, hi ha un equip de diferents funcions i administradors de dades amb responsabilitats clares. Hi ha processos estandarditzats establerts i consistència en les línies de negoci. Es disposa d'un repositori central de polítiques de dades de fàcil accés, i la qualitat de la dada és regularment monitoritzada i mesurada.
- **Fase 4. Avançat.** L'estructura organitzativa del govern de la dada es considera crítica per a totes les funcions del negoci. El negoci és el propietari del contingut de la dada i de la creació de les polítiques de la dada. Es creen i es documenten models de dades corporatius.
- **Fase 5. Optimitzat.** El govern de la dada és un procés de negoci central i les decisions de negoci es prenen avaluant el benefici, el cost i el risc. Es creen objectius de millora de processos per a l'organització, els quals es revisen contínuament per reflectir possibles canvis en els objectius del negoci. Els costos es redueixen, a causa de l'automatització de processos.

Així doncs, les diferents fases mostren un govern de la dada indisciplinat que va evolucionant de models reactius al principi a proactius, cosa que al final del procés dona lloc a un govern que proveeix una visió única de l'organització. En aquest nivell, les organitzacions fan servir dades d'alta qualitat per suportar l'automatització de processos que no necessiten, o no haurien de necessitar, la intervenció humana.

A mesura que aquest nivell augmenta, el nivell de processos involucrats serà també més gran, com es mostra a la figura 5.

Figura 5. Fases del model de maduresa



Font: Marcos Pérez

4.3.1. Models de *data governance*

Així doncs, les diferents fases mostren un govern de la dada indisciplinat que va evolucionant de models reactius al principi a proactius, cosa que al final del procés dona lloc a un govern que proveeix una visió única de l'organització. En aquest nivell, les organitzacions fan servir dades d'alta qualitat per suportar l'automatització de processos que no necessiten, o no haurien de necessitar, de la intervenció humana.

Detallarem quatre models en funció de quatre dimensions —personal, normes, tecnologia i riscos, i recompenses—:

1) Model indisciplinat

- Personal
 - L'èxit depèn de la competència d'uns pocs empleats.
 - Els analistes de negoci no estan implicats en la definició de normes de qualitat de la dada.
- Normes
 - No hi ha qualitat de dada.
 - Dades i processos aïllats.
 - Recursos no optimitzats, a causa de la redundància de dades i la seva antiguitat.

- Tecnologia
 - Estandardització i neteja de dades en fonts de dades aïllades.
 - La millora de dades se centra en determinades aplicacions.
- Riscos i recompenses
 - El risc és molt alt. Els problemes de la dada suposen la pèrdua de clients o procediments erronis.
 - Les recompenses són baixes. L'organització rep pocs beneficis de la qualitat de dada.

2) Model reactiu

- Personal
 - L'èxit depèn d'un grup específic d'empleats d'IT.
 - No hi ha procediments estàndard que cobreixin diferents àrees funcionals, però sí que hi ha iniciatives de qualitat de la dada.
- Normes
 - El focus és corregir errors quan ocorren.
 - Els processos de gestió es centren en problemes existents.
 - S'estandarditzen rols i tasques en grups individuals i departaments.
- Tecnologia
 - Es disposa d'eines de control de qualitat.
 - No hi ha integració de dades en les unitats de negoci.
- Riscos i recompenses
 - El risc és alt, a causa de la falta d'integració i consistència de dades en l'organització.
 - Les recompenses són limitades i anecdòtiques, mitjançant processos individuals.

3) Model proactiu

- Personal
 - L'equip de gestió entén el paper del govern de la dada. S'assigna personal i recursos.
 - Es considera la dada com un actiu estratègic per a la presa de decisions.
 - Sorgeix la figura de l'administrador de la dada per implementar l'estratègia i treballar amb els diferents equips funcionals.
- Normes
 - Sorgeixen activitats en temps real i normes preventives de qualitat de la dada.
 - Es comparen les mètriques de la dada amb els estàndards de la indústria per obtenir informació per millorar el model.
 - Els objectius canvien de la correcció d'errades a la prevenció.
- Tecnologia
 - Un grup d'administradors de la dada manté les normes de negoci i les descripcions de les dades de l'organització.
 - Un continu monitoritzat de les dades ajuda a mantenir la integritat de l'organització.

- Es disposa de processos en temps real, i la qualitat de la dada és compartida per diferents maneres d'operació.
- Riscos i recompenses
- El risc és mitjà o baix. Els riscos es redueixen en disposar d'una major informació que incrementa la fiabilitat en la presa de decisions.
- Les recompenses són mitjanes o altes. Cada vegada més empleats adopten el model.

4) Model governat

- Personal
- El govern de la dada és suportat en un àmbit executiu en l'organització.
- Els usuaris de negoci tenen un paper actiu en l'estratègia de la dada.
- Hi ha un grup de govern o qualitat de la dada que treballa directament amb els administradors de la dada i altres usuaris.
- Normes
- Les noves iniciatives s'aproven després de considerar com afectaran la infraestructura de la dada existent.
- Hi ha normes automatitzades per assegurar la consistència i la fiabilitat de la dada en tota l'organització.
- Tecnologia
- S'estandarditzen les eines de qualitat i integració de la dada en l'organització.
- La dada és monitoritzada contínuament, se soluciona immediatament qualsevol desviació de l'estàndard.
- Els models de dades capturen la lògica del negoci i els detalls tècnics de totes les dades organitzatives.
- Riscos i recompenses
- El risc és baix. Les dades mestres estan controlades en l'organització, cosa que permet disposar d'una alta qualitat d'informació.
- Les recompenses són altes. Hi ha una millor comprensió sobre el negoci, que permet a l'equip de gestió una plena confiança en les decisions basades en la dada.

4.4. Quin model de maduresa fem servir?

En existir tants models de maduresa, les organitzacions no tenen clar quin triar ni per on començar. I, de vegades, això es tradueix a relegar aquesta elecció al proveïdor, que aplica el model de maduresa associat a la seva eina o servei, en lloc d'un controlat per l'organització.

Aplicar el model de maduresa requereix fer-se múltiples preguntes respecte a tots els components presentats i de les persones, les polítiques i les capacitats.

L'organització ha de prendre una decisió: si apostar per models amb propietat i de pagament (com el de DAMA), per models associats a una empresa (com el d'IBM) o, fins i tot, si apostar per un model obert i públic com el d'Stanford.

El model d'Stanford és una mica diferent dels anteriors, en el sentit que, tot i que es fonamenta en CMM, limita les àrees que avaluen persones, processos i capacitats i se centra en aspectes com les dades mestres i la qualitat de la dada, deixant de banda la seguretat de la dada. Depenent de l'organització i de la seva maduresa respecte a la seguretat de la dada, pot ser també un model per tenir en compte*. Per exemple, l'estat d'Oklahoma ha aplicat aquest model en lloc dels anteriors**.

Els models de maduresa són molt útils per a dominis de capacitats coneguts. A mesura que en la disciplina de l'anàlisi de la dada apareixen noves tecnologies, cal tenir en compte que aquests models no reflecteixen aquests canvis i si els reflecteixen, probablement, serà en una mesura de disminució de la maduresa. Amb això volem assenyalar que el model de maduresa, tot i que és útil com a eina de guia, no s'ha de convertir en un fre a la innovació en l'anàlisi i en noves tècniques i enfocaments.

Lectura complementària

Diversos autors (2013). *Stanford Data Governance Maturity Model*. Stanford University.

*Més informació a:
<https://goo.gl/wo4WUh>

**Més informació a:
<https://goo.gl/pw6jpg>

Resum

En aquest mòdul didàctic, hem presentat el concepte de govern de la dada per disposar d'una visió holística en la resta de l'assignatura.

Primer, hem explicat la dada com a actiu de valor per a l'organització, i hem vist que cal canviar la seva gestió.

A més, hem introduït el que es coneix com el cicle de vida de la dada, que descriu en essència totes les etapes per les quals passa com a actiu. En funció de l'etapa i del focus, el valor de la dada serà diferent i, per tant, s'han comparat diferents formes per al seu mesurament. La necessitat de conèixer i controlar aquestes fases ens ha portat a introduir el llinatge de la dada.

Això ens ha portat a introduir el govern de la dada com a punt central de la gestió de la informació d'una organització i a explicar en què consisteix, detallant els objectius que persegueix i les fases que inclou. En definitiva, el govern de la dada busca establir mecanismes de control suficients per al control i l'explotació eficient de l'actiu més important: la dada.

També s'han revisat en detall els components que formen el govern de la dada, per saber què cal tenir en compte en la implementació d'aquest tipus de programes. Aquests components cobreixen des de la qualitat de la dada, fins a la intel·ligència de negoci i el *data warehouse*.

Finalment, s'han revisat diferents models de maduresa per poder identificar la situació actual d'una organització i determinar els seus passos futurs.

Glossari

big data *m* Conjunt d'estratègies, tecnologies i sistemes per a l'emmagatzematge, processament, anàlisi i visualització de conjunts de dades complexes.

business intelligence *f* Conjunt de metodologies, aplicacions, pràctiques i capacitats enfocades a la creació i l'administració d'informació, que permet prendre millors decisions als usuaris d'una organització.

cicle de vida d'un actiu *m* Diferents etapes per les quals passa un actiu, des del seu naixement fins al final.

gestió del cicle de vida d'un actiu *m* Enfocament basat en polítiques per gestionar el flux de la dada gràcies a un o diversos sistemes d'informació al llarg del seu cicle de vida.

govern de la dada *m* Exercici d'autoritat, control i presa de decisions compartida (planificació, vigilància i aplicació) sobre la gestió dels actius de dades.

full de ruta *f* Pla detallat que reflecteix els objectius a curt i llarg termini i guia el progrés per assolir aquests objectius, fent servir sovint una solució tecnològica específica.

internet de les coses *f* Fa referència a la interconnexió digital d'objectes quotidians amb internet.

Llei Sabarnes-Oxley *f* Llei que busca monitoritzar les empreses que cotitzen en borsa de valors, per evitar que la valorització de les seves accions sigui alterada de manera dubtosa, mentre que el seu valor és menor.

llinatge de la dada *m* Capacitat de conèixer tot el cicle de vida d'una dada, des de la data i l'hora exacta en què va ser extreta i el moment en què es va produir la seva transformació, fins a l'instant en què va tenir lloc la seva càrrega des d'un entorn font a un altre de destinació.

machine learning *f* Subcamp de la intel·ligència artificial amb focus en la creació d'algoritmes amb capacitats d'aprenentatge a partir de dades i dels resultats buscats.

Bibliografia

Berson, A.; Dubov, L. (2011). *Master Data Management and Data Governance*, 2/E. Nova York: McGraw-Hill Education.

Brackett, M.; Earley, P. S. (2009). *The DAMA Guide to The Data Management Body of Knowledge (DAMA-DMbok Guide)*. Nova York: DAMA.

Diversos autors (2007). *The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance*. Nova York: IBM Data Governance Council.

Ladley, R. S. (2012). *Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program*. San Francisco: Morgan Kaufmann.

Mosley, M. (2009). *DAMA-DMbok functional framework*. Nova York: DAMA.

Plotkin, D. (2013). *Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance*. San Francisco: Morgan Kaufmann.

Seiner, R. S. (2014). *Non-Invasive Data Governance: The Path of Least Resistance and Greatest Success*. Basking Ridge: Technics Publications LLC.

