



GESTIÓ DE DADES EN CENTRES D'R+D+I



XUNTA
DE GALICIA



Institució
CERCA
Centres de Recerca
de Catalunya



GUIA DE GESTIÓ DE DADES EN CENTRES D'R+D+I

PROGRAMA FORMATIU EN GESTIÓ ESTRATÈGICA I OPERATIVA DE CENTRES I UNITATS DE RECERCA D'EXCEL·LÈNCIA

*Una iniciativa de la Conselleria de Cultura, Educació
i Universitat de la Xunta de Galicia en col·laboració amb
la Fundació Empresa-Universitat Gallega (FEUGA)*

Aquest document constitueix un lliurable del Programa formatiu en gestió estratègica i operativa d'unitats i centres de recerca d'excel·lència (GESCI), finançat per la Conselleria de Cultura, Educació i Universitat de la Xunta de Galicia i per la Institució CERCA.

Ha estat elaborat per la Fundació Empresa-Universitat Gallega (FEUGA) amb la participació de diferents experts: Ignasi Labastida, Nadia Tonello i Lluís Rovira. També hi han col·laborat Faustino Infante i M. Jesús Tallón, de la Secretaria General d'Universitats de la Xunta de Galicia, i la consultora Estratega Consulting Global, SL.

En cas que en el text d'aquest document apareguin substantius de gènere gramatical masculí per referir-se a col·lectius, s'ha d'entendre que s'utilitzen per designar de manera genèrica individus d'ambdós sexes, sense que aquest ús comporti cap intenció discriminatòria. Aquesta opció lingüística té com a única finalitat facilitar la lectura del document i aconseguir més economia en l'expressió.

Aquesta guia no representa l'opinió ni la posició oficial de la Xunta de Galicia en aquesta temàtica.

Dipòsit legal C2121 – 2021

Copyright © 2021. Xunta de Galicia. Institució CERCA

Obra subjecta a la llicència de Reconeixement de Creative Commons, disponible a <https://creativecommons.org/licenses/by/4.0/es/>



Imprès el novembre del 2021

www.gesci.es

ÍNDEX

1	INTRODUCCIÓ	5
2	GESTIÓ DE DADES DE RECERCA: CONCEPTES BÀSICS	7
2.1	Les dades de recerca	8
2.2	Els principis FAIR	11
2.3	El cicle de gestió de dades de recerca	14
2.4	El rol del gestor de dades	16
3	LA PLANIFICACIÓ	19
3.1	Pla de gestió de dades	19
3.2	Aspectes que cal tenir en compte amb les dades personals	23
3.3	La gestió de la propietat intel·lectual	24
4	TREBALLAR AMB DADES	27
4.1	Reutilització de dades existents	27
4.2	Formats i organització	29
4.3	Emmagatzematge i accés	30
5	COMPARTIR, PUBLICAR I PRESERVAR DADES	33
5.1	Compartir dades	34
5.2	Repositoris	39
6	EL COST DE GESTIONAR DADES	43
7	UN MODEL DE POLÍTICA DE GESTIÓ DE DADES	45
7.1	Aspectes tècnics	46
8	GLOSSARI	49

Capítol

1

INTRODUCCIÓ

1 INTRODUCCIÓ

La **GUIA DE GESTIÓ DE DADES EN CENTRES D'R+D+I** és el cinquè lliurable de la «Col·lecció de Guies per a la Gestió de Centres R+D+I», desenvolupada en el marc del Programa formatiu en gestió estratègica i operativa de centres i unitats de recerca d'excel·lència (GESCI).

L'objectiu fonamental d'aquesta cinquena guia és **revisar, des d'una perspectiva pràctica, els elements fonamentals que cal tenir en compte per a la gestió de les dades de recerca a les institucions de recerca**, i proporcionar pautes d'actuació clares.

Aquesta publicació té com a **principals destinataris** els col·lectius següents:

- **Membres d'òrgans de govern i equips de direcció** de centres d'R+D+I responsables de traçar l'estratègia i definir polítiques de gestió en tots els àmbits.
- **Gestors de dades de recerca**, així com personal de gestió no especialitzat que assumeix íntegrament o parcialment funcions lligades a la gestió de les dades de recerca.
- **Personal investigador amb responsabilitats en matèria de gestió de projectes** de recerca, a qui les entitats finançadores requereixen, cada cop més, plans de gestió de dades de recerca.

La guia s'organitza internament en **capítols independents**, la qual cosa n'afavoreix una lectura individualitzada. Això no obstant, es recomana llegir tota la guia per obtenir una comprensió integral i més profunda de la gestió de dades de recerca. El document finalitza amb un **glossari** en què s'inclouen els principals conceptes relacionats amb la matèria.

L'estructura dels capítols és homogènia i inclou una explicació de cada temàtica que cal tractar, una sèrie de recomanacions associades i, addicionalment, altra informació rellevant.

Partint del que s'entén per dades de recerca i el seu cicle, i utilitzant com a marc de referència els principis FAIR per a la gestió de les dades, aquest document desenvolupa els

continguts següents:

- **La gestió de dades de recerca en un centre i els conceptes bàsics.** Aquest capítol s'inicia descrivint les dades de recerca, les diferents classificacions, els principis FAIR i les seves implicacions, el cicle de gestió de les dades de recerca i el rol del gestor de dades al centre.
- **La planificació.** Aquest capítol se centra en la gestió de dades de recerca, pren com a referència el pla de gestió de dades d'un projecte, en defineix els continguts principals i n'identifica alguns exemples i eines disponibles per elaborar-lo i fer-ne el seguiment. A continuació, es defineixen els protocols com a base per generalitzar els principis de gestió de dades en una institució, més enllà de la gestió que es faci en projectes concrets, i s'aborden dues dimensions principals que cal tenir en compte en tota planificació: la gestió de dades personals i la gestió de la propietat intel·lectual.
- **El treball amb dades.** En aquest bloc s'inclouen, des d'un punt de vista pràctic i transversal a qualsevol disciplina, aspectes com les fonts de les dades de recerca, les condicions d'accés a les dades, així com les d'ús i publicació, o la citació de fonts, a més del format de les dades i les seves implicacions pel que fa a principis FAIR, temps o cost que comporta el seu ús. Finalment s'aborden aspectes pràctics relacionats amb l'organització i l'emmagatzematge de dades.
- **Compartir, publicar i preservar dades.** Aquest capítol aborda aquestes baules del cicle de gestió, i aprofundeix en la dualitat de dades obertes i tancades i en els diferents tipus de llicències en la gestió de dades i de repositoris existents.
- El capítol final tracta l'**estimació del cost** de gestionar dades i el desenvolupament d'un model de **política de gestió de dades**.

Aquesta guia ofereix un ventall de coneixements, eines i recursos amb què es pretén consolidar la **cultura de la gestió de les dades de recerca** com a fonament de l'excel·lència dels centres de recerca de Galícia, en línia amb la filosofia de l'espai europeu de recerca i del programa GESCI en què s'emmarca.



Capítol 2

GESTIÓ DE DADES DE
RECERCA: CONCEPTES
BÀSICS

2

GESTIÓ DE DADES DE RECERCA: CONCEPTES BÀSICS

La **gestió de dades de recerca** (o *research data management*, RDM, en anglès) comprèn les decisions relatives a les activitats que tenen a veure amb el **cicle de vida de les dades de recerca**, és a dir, amb la **recopilació, organització, processament, anàlisi, preservació i publicació de les dades utilitzades en un projecte de recerca**. La possibilitat de reutilitzar les dades creades

en un projecte i per a altres projectes és la característica que reactiva el procés de gestió i li confereix el caràcter de cicle. En aquest context, els **principis FAIR** (de l'anglès *Findable, Accessible, Interoperable, Reusable*) van sorgir com a resposta davant la necessitat d'homogeneïtzar la bona gestió de les **dades de recerca**. A la figura següent es mostra el cicle de vida de les dades de recerca.

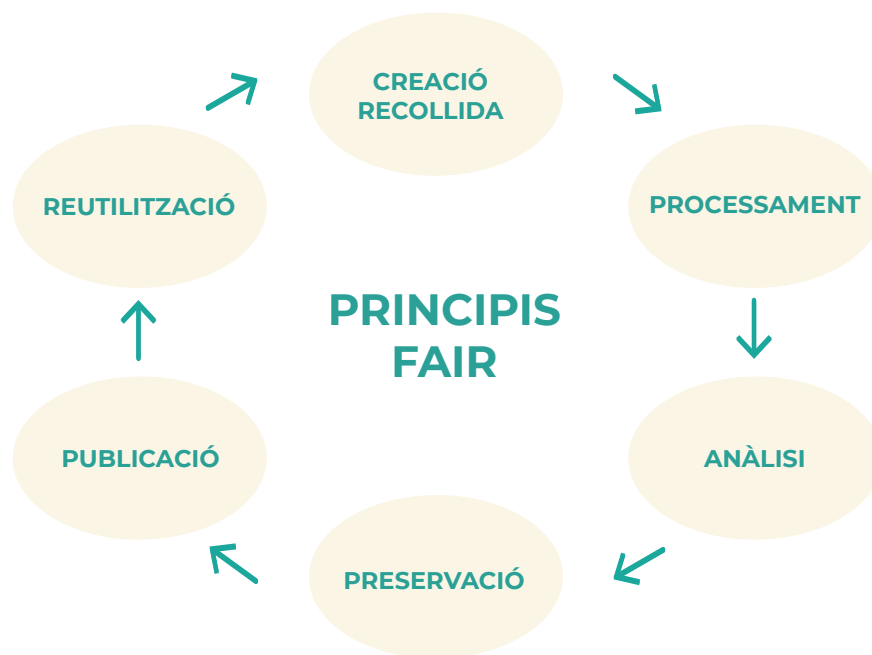


Figura 1. Cicle de vida de les dades de recerca. Font: elaboració pròpia.

La diferent tipologia de dades i la varietat de formats, repositoris o disciplines dificulta que hi hagi una única estratègia recomanable de maneig de les dades. Cada centre ha de desenvolupar la seva pròpia estratègia per aconseguir una gestió adequada de les seves dades que redundi en una millora de la recerca. Fer-ho els permetrà:

- Disposar de més transparència per validar els resultats de la recerca.
- Assegurar que les dades siguin localitzables, accessibles, interoperables i reutilitzables.

- Millorar el perfil de l'investigador, l'impacte i la visibilitat dels projectes.
- Afavorir la protecció de dades i minimitzar el risc de pèrdua d'aquestes dades.
- Complir la legislació sobre la matèria i els requeriments de les entitats finançadores.
- Estalviar temps i fer un ús eficient dels recursos disponibles.

En la gestió de dades de recerca d'una institució d'R+D+I **es poden distingir tres figures clau**, que tenen aquestes responsabilitats principals:

● **El gestor de dades (o *data steward*)** d'un centre és la persona que **coordina la política de gestió de dades de la institució i que fa d'engrenatge entre els grups de recerca**, les polítiques del centre i les de les entitats finançadores, i gestiona l'aplicació de lleis o normatives i les infraestructures a disposició dels investigadors. Les seves responsabilitats en la gestió de dades de recerca van des d'analitzar les necessitats de gestió de dades del col·lectiu a què serveixen, proporcionar assessorament, coordinar els interessos de *stakeholders* (parts interessades) entorn de les dades, formar i difondre bones pràctiques de gestió, desenvolupar polítiques de gestió de dades de recerca del centre, etc.

● **L'analista o científic de dades (*data scientist*)** és la persona que, **dins d'un equip d'investigadors, s'encarrega de les dades de recerca**. Marca les pautes sobre com recollir-les, processar-les i preparar-les per analitzar-les. Les seves responsabilitats en la gestió de dades de recerca van des de col·laborar amb altres investigadors en la recollida i l'anàlisi de dades, fer vigilància tecnològica, crear estratègies d'anàlisi i visualització d'informació, identificar problemes i proporcionar solucions d'anàlisi de dades, utilitzar tècniques estadístiques i models de mineria de dades amb aquestes finalitats, avaluar fonts i tècniques de recollida de dades, mantenir-se actualitzat respecte a tecnologies, tècniques i mètodes, etc.

● **L'investigador** és la persona que duu a terme l'activitat investigadora i que es responsabilitza, dins de l'àmbit dels seus propis projectes, de la **recollida de dades** i de la **implementació de bones pràctiques** de maneig, el **compliment de requisits ètics i legals** o la presa de decisions d'acord amb la política de gestió de dades. Així mateix, **en la reutilització de dades ha de tenir en compte les condicions** a les quals estan subjectes les

dades, i **ha de respectar el dret de citació dels creadors**. Finalment, en absència dels perfils anteriors, ha de gestionar les dades de recerca al llarg de tot el cicle, i assumir la resta de funcions esmentades.

2.1 Les dades de recerca

Les dades de recerca **són un conjunt d'informació, digitalitzada en fitxers o qualsevol suport físic, que en el procés de recerca contribueixen a la producció d'un resultat científic**. Segons el consorci CASRAI, que proposa establir estàndards per descriure la recerca, són l'evidència en un procés de recerca per validar les seves troballes i resultats.¹

Les dades de recerca **poden ser xifres que s'han obtingut mitjançant un experiment o una simulació**, però també poden ser **imatges, enregistraments o transcripcions d'entrevistes**. Per tant, les dades de recerca **poden tenir orígens i formats diversos**, i poden ser dades observades, simulades, recollides, processades, etc.

D'altra banda, en el marc d'un procés de gestió de dades, les dades s'agrupen en col·leccions o bases de dades. Al llarg d'aquesta guia els termes *dades* i *conjunts de dades* s'utilitzen indistintament, encara que hi ha diferències quant a la gestió de la propietat intel·lectual, com s'explicarà més endavant.

Hi ha **moltes maneres de classificar dades**, segons la disciplina i el tipus de projecte de recerca. A la taula següent es presenten **alguns exemples de les classificacions de dades més comunes**, amb una breu referència als principals requisits que comporta la seva gestió.

Informació addicional rellevant:

¹ <https://casrai.org/term/research-data/>

CRITERI DE CLASSIFICACIÓ	TIPUS DE DADES	DESCRIPCIÓ	EXEMPLES	REQUERIMENTS DE GESTIÓ
Segons el suport en què es troba la informació	Físiques	Són totes les dades de recerca en formats com el paper o un altre suport analògic	<ul style="list-style-type: none"> • Documents de text i taules de dades en paper • Fotografies, gràfics o mapes en paper • Mostres de materials inorgànics o de teixits orgànics • Col·leccions, biblioteques, fonoteques, materiotèques, pinacoteques, etc. 	Requereixen un sistema d'emmagatzematge específic i especialitzat, segons el tipus de material de suport en qüestió
	Digitals	Són les que han estat produïdes per ordinador o un altre sistema amb resultat digital	<ul style="list-style-type: none"> • Documents de text en suport digital • Vídeos, àudios o fotos en suport digital • Gràfics o mapes en suport digital • Bases de dades • Fitxers de valors numèrics generats per instruments (mesuraments) en suport digital • Fitxers de valors derivats del processament o l'anàlisi 	Requereixen nomenar i organitzar el sistema d'emmagatzematge en carpetes i fitxers, si bé poden dependre del format, del contingut, de la fase de processament de les dades, etc.

CRITERI DE CLASSIFICACIÓ	TIPUS DE DADES	DESCRIPCIÓ	EXEMPLES	REQUERIMENTS DE GESTIÓ
Segons el nivell de processament de la informació	Primàries	Constitueixen el punt de partida d'una recerca. Són dades que no han patit modificacions	Dades resultants de: <ul style="list-style-type: none"> • Mesuraments • Enquestes • Observacions • Experiments 	La producció de dades primàries s'ha de justificar en el PGD. S'han d'emmagatzemar juntament amb la informació de la font de procedència
	Secundàries	Són dades que han patit algun tipus de processament (dades intermèdies, resultats)	<ul style="list-style-type: none"> • Dades processades • Resultats d'anàlisis • Representacions gràfiques • Resums, anotacions, etc. 	Els autors són responsables del compliment dels principis FAIR i de proveir informació suficient per reproduir-les
Segons el tipus de contingut	Personals	Es refereixen a característiques de persones i en permeten la identificació	<ul style="list-style-type: none"> • Noms i cognoms • Adreces postals, números de telèfon, IP de dispositius electrònics • Dades genètiques, clíniques • Característiques físiques, edat • Vídeos, fotos, imatges 	Aquestes dades estan subjectes a mesures especials de seguretat, segons l'RGPD i les lleis del país d'origen
	No personals	No es refereixen a persones, i si són dades de persones, no en permeten la identificació	<ul style="list-style-type: none"> • Dades anonimitzades • Càrrecs o noms d'empreses • Qualsevol dada de plantes, animals o coses 	Les mesures de seguretat i accés d'aquestes dades s'estableixen segons la seva confidencialitat

Font: elaboració pròpia.

ACCIONS RECOMANADES

- Identificar les dades de recerca que ha de gestionar cada institució i proporcionar-ne la definició i classificació al personal investigador de cada unitat de recerca, basant-se en la disciplina científica.

2.2 Els principis FAIR

Els **principis FAIR** tenen l'origen en l'article «FAIR Guiding Principles for scientific data management and stewardship» publicat el 2016 a la revista *Scientific Data*. Són un **conjunt de directrius que cal seguir perquè les dades de recerca siguin accessibles, es puguin entendre i es puguin reutilitzar**, també en estudis d'altres disciplines. En definitiva, els principis FAIR van sorgir com a **resposta davant la necessitat d'homogeneïtzar i millorar la gestió de les dades de recerca**, que generalment depenia del criteri del propietari de les dades. Per tant, les dades es consideren FAIR quan estan curades segons aquests principis. En aquest article, es proposa que les dades de recerca han de complir els quatre principis que s'expliquen a continuació:

- **Ser localitzables:** les dades són localitzables **quan tenen assignat un identificador global únic i permanent**. Com a identificador es pot utilitzar un DOI (*Digital Object Identifier*) o un identificador del sistema Handle, que es fa servir generalment en repositoris per a publicacions. La majoria dels proveïdors de serveis de

publicació de dades ofereixen l'assignació de DOI, mentre que el seu ús també està molt generalitzat per a publicacions.

En cas que un centre vulgui fer aquesta assignació en un repositori propi, s'ha de donar d'alta en una agència de registre, com ara DataCite, especialista en el registre de DOI per a dades. Hi ha altres agències de registre, com Crossref, que s'especialitza en publicacions.

Tant les dades com les metadades s'han de registrar i indexar perquè siguin localitzables. Per això, en cas que es faci servir un repositori propi, cal donar-lo d'alta als registres internacionals, especialment al portal de referència re3data.² En cas que el centre publiqui les dades en un servei extern, possiblement aquest registre ja s'hagi fet.

A més, en el cas de la indexació de les metadades, cal seguir pautes i estàndards, com per exemple els que ofereix l'esquema de DataCite, on s'indiquen les metadades obligatòries per descriure un conjunt de dades, així com els valors o les maneres de descriure-les.

Informació addicional rellevant:

² <https://re3data.org/>

● **Ser accessibles:** les dades són accessibles quan les metadades es poden recuperar a través del seu identificador mitjançant un protocol de comunicació estàndard. Aquest protocol ha de ser obert, gratuït i d'accés universal, i ha d'incloure la possibilitat d'establir un procediment d'accés amb autenticació i autorització quan sigui necessari. De nou, aquesta accessibilitat pot estar ja establerta per defecte en un dipòsit extern, però caldrà activar-la si es tracta d'un dipòsit propi. Finalment, quan les dades ja no estiguin disponibles, s'ha de mantenir l'accés a les metadades.

● **Ser interoperables:** la interoperabilitat ve determinada per l'ús d'un llenguatge formal, accessible, compartit i àmpliament aplicable per representar el coneixement, tant en el cas de les dades com de les metadades. Les metadades utilitzen vocabularis que segueixen els mateixos principis FAIR i inclouen referències condicionades a altres metadades.

● **Ser reutilitzables:** les dades són reutilitzables quan les metadades fan

servir en la seva descripció diversos atributs precisos i rellevants, es divulguen amb una llicència d'ús clara i accessible, es vinculen de manera detallada amb la seva procedència i compleixen els estàndards de la comunitat de la disciplina corresponent.

Els conceptes de **dades gestionades segons els principis FAIR** i **dades obertes** no són equivalents ni excloents, sinó **complementaris**. No obstant això, en aquest sentit hi ha iniciatives com el moviment Open Science per a la ciència oberta i la bona gestió de les dades, que té com a objectiu la maximització de dades obertes i FAIR.

Els principis FAIR no garanteixen que les dades siguin obertes, igual que les dades obertes no són necessàriament dades curades sota els principis FAIR. El nivell d'accessibilitat és el que determina si les dades són obertes o restringides a determinats usuaris. El grau de compliment d'aquests principis és el que determina si les dades són FAIR. A la figura següent es pot veure exemplificat:

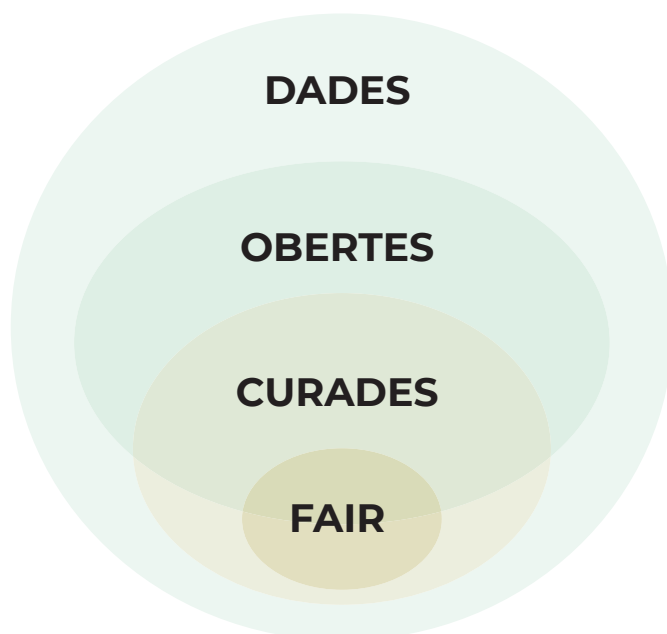


Figura 2. Dades obertes, dades curades i dades FAIR. Font: elaboració pròpia.

2.2.1 Com es pot aconseguir que les dades segueixin els principis FAIR

Un dels **principals objectius que cal plantejar per a una bona gestió de les dades** de recerca consisteix a aconseguir **que les dades de la institució compleixin els principis FAIR al màxim nivell possible**. Aquest objectiu s'ha de preveure des de l'inici de qualsevol **projecte** de recerca, quan **caldrà decidir el següent**:

- Quins són els **formats més adequats** per treballar amb les dades i preservar-les.
- Quin és el **repositori o la infraestructura** que convé utilitzar.
- Quines **metadades** s'han de fer servir per descriure les dades.

- Quins **identificadors permanents** es faran servir en publicar les dades.

En general, el personal investigador no té coneixements específics en gestió de dades per adaptar-les als principis FAIR, per la qual cosa, si **el centre** adopta aquests principis, **haurà de proporcionar al seu personal**:

- **Formació** en la matèria.
- **Accés a les eines** que puguin ser necessàries.
- **Posada a la seva disposició de personal especialitzat en la gestió de dades** que, a més d'oferir assessorament sobre tot el procés, pugui aportar el seu coneixement per a l'elecció de metadades, formats, repositoris, etc.

ACCIONS RECOMANADES

- Declarar, al document de política de gestió de dades del centre, la seva intenció i nivell de compromís amb el seguiment dels principis FAIR, per tal que les dades de recerca gestionades siguin localitzables, accessibles, interoperables i reutilitzables.

- Proporcionar els ajuts pel que fa a mitjans, personal, informació, etc. que el personal investigador necessiti per complir l'aplicació dels principis FAIR i informar sobre les plataformes i els repositoris genèrics, institucionals o temàtics que té a la seva disposició, així com sobre les eines que pot utilitzar per publicar i compartir dades seguint els principis FAIR.

Informació addicional rellevant:

- «FAIR Guiding Principles for scientific data management and stewardship»: <https://www.nature.com/articles/sdata201618>
- Podeu trobar una explicació més detallada sobre els principis i subprincipis FAIR al portal GoFAIR: <https://www.go-fair.org/fair-principles/>
- FIP mini-questionnaire. Build your FAIR Implementation Profile: <https://www.go-fair.org/wp-content/uploads/2021/06/FIP-mini-questionnaire.pdf>
- Esquema de metadades més utilitzat en repositoris de dades, DataCite: <https://schema.datacite.org/>
- EU-LIFE recommendations on Research Data Management in the life sciences: <https://eu-life.eu/newsroom/news/eu-life-recommendations-research-data-management-life-sciences>

2.2.2 Exemples de dades FAIR

Tot i que sempre resulta difícil identificar exemples genèrics que siguin d'utilitat a causa de les peculiaritats de cada disciplina o àmbit científic, a continuació es destaquen exemples de portals que ofereixen les dades seguint els principis FAIR.

UK Data Service³

És un portal que ofereix un catàleg de dades amb una àmplia varietat de metadades, i inclou l'ús de DOI com a identificador i l'ús de vocabularis controlats. En molts casos, l'accés és restringit i requereix tenir un compte, encara que això no impedeix que les dades segueixin els principis FAIR.

Omics DI⁴

Aquesta plataforma ofereix la localització de conjunts de dades de genòmica o proteòmica, entre altres àmbits. No fa servir un DOI, sinó un identificador propi o l'identificador que tingui el conjunt de dades en origen, encara que, novament, les dades estan àmpliament descrites i s'ofereix informació sobre la seva procedència. D'altra banda, el portal inclou referències de les publicacions lligades a les dades i ofereix l'accés mitjançant un protocol de comunicació o servei web que permet obtenir el resultat a una consulta directament, sense necessitat de descarregar-se les dades.

Mendeley Data⁵

Algunes plataformes o repositoris destaquen la qualitat de determinades dades, com és el cas d'aquest portal que ofereix una llista dels conjunts de dades que compleixen millor els principis FAIR seleccionats per un comitè d'investigadors extern.

ACCIONS RECOMANADES

- Disposar d'una llista d'exemples de dades FAIR en les disciplines concretes i difondre-la convenientment entre el personal investigador.

2.3 El cicle de gestió de dades de recerca

La gestió de dades de recerca es duu a terme en paral·lel al procés de recerca, per la qual cosa comporta la realització de tasques durant la planificació, l'execució i la difusió de resultats de l'R+D, tal com es mostra a la taula següent.

Informació adicional rellevant:

³ <https://beta.ukdataservice.ac.uk/datacatalogue>

⁴ <https://www.omicsdi.org/>

⁵ <https://data.mendeley.com/collections/46769202-3a48-4097-a6fb-1d056cf0d99d>

FASE DE RECERCA	FASE DE GESTIÓ DE DADES DE RECERCA	ACTIVITATS DE GESTIÓ DE DADES
Inici del projecte d'R+D	Inici del projecte d'R+D	<ul style="list-style-type: none"> • Elecció de l'eina d'edició • Planificació de revisions i actualitzacions • Identificació de les dades existents i dels permisos o els drets per accedir-hi i utilitzar-les
	Processament	<ul style="list-style-type: none"> • Elecció de la infraestructura per a la preservació de les dades creades o recollides (activitat FAIR) • Definició de l'estructura de carpetes i convencions per a noms de fitxers (FAIR)
Recerca (Anàlisi)	Anàlisi	<ul style="list-style-type: none"> • Elecció i documentació de dades, formats i metadades (FAIR)
	Preservació	<ul style="list-style-type: none"> • Definició de mecanismes de còpia de seguretat i preservació. Seguretat i control d'accessos: repositoris institucionals (FAIR)
Publicació de resultats	Publicació	<ul style="list-style-type: none"> • Publicació de dades en repositoris temàtics o multidisciplinaris • Elecció de llicències d'ús, protecció de dades i copyright
Avaluació	Gestió de dades	<ul style="list-style-type: none"> • Avaluació dels costos de gestió • Avaluació i revisió del pla de gestió

Font: elaboració pròpia.

Posteriorment, un cop finalitzat el projecte i publicats els resultats, cal continuar projectant la gestió de les dades conforme als principis FAIR, per tal de garantir-ne el com-

pliment efectiu (emmagatzematge segur, accessibilitat per a altres investigadors, disponibilitat per a la reutilització, etc.).

2.4 El rol del gestor de dades

La funció principal d'un **gestor de dades** consisteix a **alinejar**, a través de la gestió de les dades, **els interessos dels diferents stakeholders que intervenen en el procés investigador**. Així, hi ha **tres àmbits d'influència principals** que el gestor de dades ha de coordinar:

- **L'àmbit polític o normatiu:** el gestor de dades ha de tenir presents normatives i polítiques que li venen donades (per exemple, RGPD, estratègia o política institucional, de centre, etc.).
- **L'àmbit de la recerca pròpiament dit:** el gestor de dades ha de tenir en compte **com són els fluxos d'informació** perquè es dugui a terme el procés investigador **segons la disciplina** de què es tracti (per exemple, la necessitat de compartir informació amb agents externs), les **eines necessàries** (per exemple, programari per al processament de

dades o maquinària de laboratori específica), **formats estàndard d'ús comú a la disciplina**, etc.

- **L'àmbit de la infraestructura i els serveis disponibles:** el gestor de dades ha de **conciliar les disponibilitats** d'infraestructura i serveis tant amb les necessitats del col·lectiu investigador com amb els condicionants polítics o normatius identificats (per exemple, la manera de tractar la informació confidencial, com ara l'anonimització o la gestió d'accessos, entre d'altres; l'elecció del tipus de repositori per a la publicació de les dades; el tipus de llicència, etc.).

La il·lustració següent mostra les tres àrees d'influència principals que conflueixen en la gestió de les dades de recerca: política, activitat investigadora i infraestructura. L'estudi que ha generat aquesta il·lustració ha estat realitzat en l'àmbit de les ciències de la vida, però es pot estendre a altres camps de coneixement.

ÀREES D'IMPLEMENTACIÓ PER A LA GESTIÓ DE DADES



Figura 3. Àrees de competència del gestor de dades. Font: <https://doi.org/10.5281/zenodo.3460552>

Per tant, a causa de la seva comesa, el **gestor de dades integra característiques de tres perfils diferents:**

- De l'**investigador**, ja que necessita un cert grau de coneixement sobre la disciplina científica per poder entendre'n el llenguatge, així com conèixer els formats i les tipologies de les dades amb què es treballa.
- De l'**informàtic**, per poder tenir les dades emmagatzemades de manera segura en infraestructures adequades, per compartir-les durant el desenvolupament d'una recerca.
- Del **documentalista**, per poder descriure les dades d'acord amb els esquemes de metadades existents.

La creació de la figura del gestor de dades en l'àmbit universitari s'està implementant en alguns països. És el cas dels Països Baixos, on la Universitat de Tecnologia de Delft (TU Delft) es pot considerar un bon exemple, ja que ha implantat un equip de gestió de dades a cada facultat integrat per professionals especialitzats. Aquestes unitats assessoren el personal investigador, imparteixen formació, fan un se-

guiment de les polítiques, es coordinen amb els equips tècnics, etc. Són la peça clau per a la gestió de les dades a la institució.⁶

Per **fer una bona gestió de dades, un centre de recerca ha de disposar de personal qualificat per dur a terme les diferents tasques** que inclou aquest procés. Amb la formació adequada, algunes tasques les pot fer el personal investigador del centre mateix, però d'altres necessitaran la intervenció de professionals especialitzats. Per això, alguns grups disposen d'analistes o enginyers de dades, encara que **l'ideal és que el centre tingui un curador o gestor de dades.**

Per tant, **el centre ha de programar una formació estable** que es pot **complementar amb formacions més específiques** en determinats aspectes depenent de les característiques del centre o dels projectes, per exemple, com s'han d'abordar els aspectes ètics o com s'han d'anonimitzar dades. Aquestes formacions les pot fer personal especialitzat, però també es poden organitzar trobades entre investigadors per intercanviar bones pràctiques i experiències que permetin anar capacitant tota l'organització.

ACCIONS RECOMANADES

● **Disposar de personal preparat per a la gestió de dades, capaç de formar i orientar el personal investigador en la presa de decisions de gestió de dades de recerca dels seus projectes, la qual cosa implica:**

- Identificar els rols necessaris per a la gestió de dades.
- Crear perfils de treball nous o assignar rols de gestió de dades als perfils existents.

● **Disposar d'un programa de formació en gestió de dades.**

Informació addicional rellevant:

Per aprofundir sobre les àrees de competència dels gestors de dades en el context europeu de ciència oberta, es recomana consultar l'informe *Digital skills for FAIR and open science: Report from the EOSC Skills and Training Working Group* (<https://op.europa.eu/s/pnPe>).

⁶ <https://openworking.wordpress.com/2020/02/14/a-year-in-the-life-of-a-data-steward/>

A man in a dark blue shirt is looking at a tablet in a server room. The room has blue lighting and server racks with glowing lights. A large teal shape is overlaid on the left side of the image.

Capítol 3

LA PLANIFICACIÓ

3

LA PLANIFICACIÓ

Abans de començar qualsevol recerca cal planificar com s'han de gestionar les dades que s'utilitzaran. Per això cal elaborar un pla de gestió de dades (PGD, o DMP, per les sigles en anglès de *data management plan*) que inclogui la descripció i el tractament de les dades al llarg del cicle de vida de les dades de recerca.

En aquesta secció es presenten les pautes perquè el personal investigador pugui elaborar un PGD i perquè els centres ofereixin el suport necessari per fer-ho. Així mateix, es revisen altres aspectes rellevants que cal tenir en compte abans de començar la recerca, com ara la gestió de la propietat intel·lectual o la necessitat de dissenyar la recerca considerant el possible impacte que pugui tenir l'ús de dades personals en la privacitat de les persones.

Pel que fa a la confecció del PGD, quan un centre disposa d'uns procediments definits i incorporats en un protocol de gestió de dades, la descripció de les activitats ha de prendre com a base allò que ja està escrit. En cas contrari, cal definir què s'ha de fer amb les dades en cadascun dels PGD dels projectes en què es participi. En aquest sentit, és convenient tenir present que cada cop hi ha més entitats finançadores que requereixen l'elaboració d'un PGD i que els centres també ho incorporen com un requisit en el marc de les seves polítiques de tractament de dades de recerca.

3.1 Pla de gestió de dades

El pla de gestió de dades (PGD) és un document viu, actualitzable i versionable, que recull la descripció de les dades, les eines i les decisions de gestió escollides abans i durant l'execució d'un projecte de recerca. La seva elaboració i revisió pot recaure en l'investigador principal o en un altre membre del grup de recerca que sigui responsable d'assumir aquesta tasca, per exemple, analistes o enginyers de dades, en el cas que hi hagi aquestes figures.

El PGD no s'ha de considerar un simple document que cal emplenar per exigències de l'entitat finançadora, sinó que ha de ser un document dinàmic que ajudi el personal investigador a millorar la seva recerca, dissenyant com es faran servir les dades, identificant possibles problemes que puguin sorgir i ideant alternatives. Com qualsevol altra planificació, és possible que al llarg del procés s'hagi de modificar —per exemple, quan no es puguin obtenir les dades previstes—, per la qual cosa és recomanable fer revisions periòdiques i una revisió final un cop hagi conclòs el projecte.

Els PGD han de ser validats per qui en sol·licita l'elaboració, ja sigui l'entitat finançadora o el centre mateix. Des del punt de vista d'un centre, se'n pot establir la utilització com a política institucional, però continua sent recomanable assignar personal al seu seguiment, i idealment ha de ser responsabilitat del gestor de dades.

El PGD ha d'incloure el següent:

● **Informació mínima administrativa**, inclo-sa la identificació del projecte (títol, codi i entitat finançadora) i de l'investigador principal. S'hi poden incloure altres dades, com ara la identificació d'altres membres de l'equip investigador i altres membres del consorci, si escau.

● **Descripció de les dades implicades en el projecte**, amb els formats i una estimació de la mida. Les dades poden ser creades en el procés de recerca, ja sigui per simulació, experimentació o observació, o bé es poden utilitzar dades ja existents a les quals tingui accés el personal investigador. En aquest darrer cas, cal especificar com s'accedeix a aquestes dades i quins són els usos autoritzats.

● **Descripció del sistema d'emmagatzematge** que es farà servir durant el projecte, així com els **sistemes que s'empraran per analitzar o processar les dades**. Cal tenir en compte qui ha de tenir accés a aquest espai, tant si és personal intern com extern. En cas que s'hagi de treballar amb dades personals, cal informar el delegat de protecció de dades i obtenir la validació del comitè d'ètica que correspongui, respectivament.

● **Política d'accés: qui pot accedir a les dades, i com i quan ho pot fer**. Aquest accés pot ser escalonat al llarg del temps, o ser restringit i establir unes condicions d'ús determinades. A més, **s'ha d'identificar** qui és el responsable de les dades i, per tant, **qui pot concedir accés a la informació**.

● Finalment, cal indicar on **s'han d'emmagatzemar i preservar les dades** un cop hagi conclòs el projecte. En cas que les dades es publiquin, cal cercar un repositori de confiança i incloure les metadades corresponents.

És habitual que les entitats finançadores de la recerca proposin els seus propis PGD, amb plantilles que els beneficiaris han d'emplenar i enviar. Per exemple, la

Comissió Europea i el Consell Europeu de Recerca (ERC) proposen l'esquema següent:

- Resum de les dades: descripció de les dades que s'utilitzaran, format i mida.
- Dades FAIR: explicació dels passos que cal seguir perquè les dades compleixin els principis FAIR.
- Assignació de recursos: identificació dels recursos assignats a la gestió de dades. Poden imputar-se com a costos del projecte.
- Seguretat de les dades: descripció de procediments per al control d'accessos a les dades, especialment rellevants quan es treballa amb dades sensibles o confidencials.
- Aspectes ètics: identificació de qualsevol aspecte ètic relacionat amb la recerca, normatives o lleis aplicables.
- Altres aspectes: qualsevol altre aspecte que es consideri important destacar.

L'estructura i el contingut que suggereix la *Fundació «la Caixa»* en les seves convocatòries de recerca es compon de cinc apartats:

- Resum de les dades.
- Responsabilitats.
- Dades FAIR.
- Seguretat de les dades.
- Altres temes.

En aquest model cal destacar l'apartat «Responsabilitats», on s'ha d'indicar qui és la persona responsable de la gestió de les dades i de la concessió d'accessos i titular dels possibles drets de propietat intel·lectual.

A escala estatal, els darrers plans *d'R+D+I* recullen la recomanació d'elaborar PGD, sense que hi hagi cap requeriment —com sí passa a escala europea— ni plantilla oficial que serveixi de guia.

ACCIONS RECOMANADES

- Proporcionar suport al personal investigador perquè pugui planificar la gestió de dades dels projectes d'acord amb els requisits de les principals entitats finançadores.
- Establir com a bona pràctica un PGD propi del centre alineat amb les plantilles existents de les entitats finançadores, a fi d'afavorir que els investigadors estiguin familiaritzats amb els aspectes principals que cal tenir en compte.
- Publicar a la intranet del centre el model de PGD, així com una secció de preguntes freqüents sobre la seva elaboració.
- Organitzar formacions periòdiques per a investigadors sobre l'emplenament del PGD, utilitzant tant el model propi del centre com les plantilles de les principals entitats finançadores.
- Fer revisions sistemàtiques del PGD, així com una revisió final, una vegada conclòs el projecte, amb personal assignat a aquestes tasques.

3.1.1 Eines per elaborar un PGD

La confecció d'un PGD **no requereix una aplicació específica**, ja que es pot generar com a informe utilitzant un programa de tractament de textos. No obstant això, **hi ha eines en línia** que permeten elaborar el pla —fins i tot de forma col·laborativa—, compartir-lo amb altres usuaris i fer un seguiment de les diferents versions, a fi de controlar els canvis introduïts durant el projecte de recerca. A més, aquestes aplicacions solen recollir els continguts fixats per les entitats finançadores a través de plantilles i, en cas que s'instal·lin

als servidors propis del centre, permeten també la incorporació de plantilles pròpies. Tot seguit es destaquen dues de les eines més conegudes.

DMP Online⁷

És l'eina més coneguda i utilitzada. Creada pel Digital Curation Centre (DCC) escocès, ha estat adaptada i traduïda per diferents institucions com, per exemple, el *Consorti Madroño*⁸ o el *Consorti de Serveis Universitaris de Catalunya (CSUC)*⁹. El DCC ofereix el codi de l'aplicació perquè es pugui instal·lar localment i gestionar les plantilles que es vulguin oferir als usuaris.

Informació addicional rellevant:

- *H2020 templates: Data management plan v2.0 – 15.02.2018*: https://ec.europa.eu/research/participants/data/ref/h2020/other/gm/reporting/h2020-tpl-oa-data-mgt-plan-annotated_en.pdf
- *Model de pla de gestió de dades per a projectes de la Fundació Bancària «la Caixa»*: https://fundacionlacaixa.org/documents/10280/1198289/modelo_de_plan_de_gestio_n_la_caixa_es.pdf/685ac764-bc0b-9511-8e88-2768919637a8?t=1593414644950

⁷ <https://dmponline.dcc.ac.uk/>

⁸ <https://pgd.consorciomadrono.es/>

⁹ <https://dmp.csuc.cat/>

ARGOS¹⁰

Oferta pel consorci Open AIRE, és semblant a l'anterior, encara que presenta com a particularitat que és *machine readable* (llegible per màquina), és a dir, permet estandarditzar els continguts del pla i que siguin més interoperables, a més de facilitar l'extracció de la informació més rellevant.

ACCIONS RECOMANADES

- Proporcionar al personal investigador eines de suport per a l'elaboració dels PGD, així com la formació pràctica sobre com s'han de fer servir.
- Facilitar al personal investigador accés a un repositori amb exemples reals de PGD i plantilles adaptades a la casuística del centre.

3.1.2 Els protocols de gestió de dades

Els PGD estan associats a projectes de recerca **concrets**. Tanmateix, si se'n vol generalitzar l'ús en un centre, cal elaborar un protocol de gestió de dades que, alhora, pot formar part d'una política o estratègia general de la institució.

Un protocol de gestió de dades **ha d'incloure:**

- Les normatives i les lleis aplicables.
- Les responsabilitats de la gestió de les dades i de l'accés a les dades.
- La infraestructura i els serveis que el personal investigador té a la seva disposició durant tot el cicle de les dades de recerca.
- Una llista de formats preferents on figurin els més comuns entre les comunitats investigadores dels membres del centre.
- Informació sobre la gestió de dades personals i la gestió de la propietat intel·lectual.
- Les opcions de preservació i publicació de les dades que ofereix el centre.

Malgrat que l'elaboració d'un protocol és una de les atribucions del responsable de gestió de dades del centre, és recomanable disposar d'un grup de treball on participin tots els perfils que intervenen en les diferents activitats d'aquest procés, ja sigui directament, com els investigadors, científics de dades, etc., o indirectament, com els delegats de protecció de dades personals o els responsables de sistemes, entre d'altres.

ACCIONS RECOMANADES

- Elaborar i publicar a la intranet del centre un protocol de gestió de dades alineat amb la política o l'estratègia general del centre, per tal que tot el personal conegui les directrius que cal seguir en les activitats d'aquest procés.

Informació addicional rellevant:

Proposta de protocols de gestió de dades per disciplines de Science Europe:
https://www.scienceeurope.org/media/nsxdyvqn/se_guidance_document_rdmmps.pdf
¹⁰ <https://argos.openaire.eu/splash/>

3.2 Aspectes que cal tenir en compte respecte de les dades personals

Segons el **Reglament general de protecció de dades (RGPD)**,¹¹ **són dades personals totes les dades que permeten la identificació d'una persona**, com, per exemple: noms, dates, característiques físiques, imatges personals, adreces de correu físiques o electròniques, adreces IP (*Internet Protocol*) de dispositius, números d'identificació associats a una base de dades a través de la qual es poden trobar dades que identifiquen una persona, etc.

Entre aquestes dades personals hi ha també les **«especialment protegides»** o sensibles, com **les relatives a origen ètnic, creences, opinions polítiques, orientació sexual o afiliació sindical d'una persona**, per a les quals cal disposar d'infraestructures segures, adequades per treballar i preservar aquesta informació. També hi ha les dades **«de categoria especial»**, que es refereixen a **informació genètica, dades biomètriques o de salut**, que també s'han de tractar amb cautela per no posar en risc els drets de les persones (són dades tancades, de manera que s'ha de prestar especial atenció als forats de seguretat i vigilar on s'emmagatzemen).

Moltes vegades els **projectes de recerca inclouen l'ús de dades personals. És molt important que el centre estableixi els procediments que cal seguir** en aquests casos per complir la normativa i el codi de conducta propi. A més, el personal investigador ha de disposar de la informació necessària per utilitzar aquestes dades correctament.

Quan es tracten dades personals cal implicar-hi el delegat de protecció de dades des que es comença a treballar en la planificació, ja que aquest perfil supervisa els processos previstos **per protegir els**

drets fonamentals de les persones i pot proporcionar assessorament per mitigar o eliminar el possible impacte en la seva privacitat. Es tracta d'un perfil diferent del de gestor, ja que té caràcter jurídic i és expert en la gestió de les dades personals en l'àmbit de la recerca, però també en altres àmbits com les enquestes de qualitat, les enquestes a associats, etc.

D'altra banda, cal **analitzar si totes les dades que es demanen són imprescindibles per a la recerca i informar adequadament** els participants sobre les dades que s'utilitzaran i com seran processades.

Així mateix, **el centre ha de tenir registrats els diferents processos que duu a terme amb les dades personals** i, en cas que **estigui previst que en algun moment aquestes dades passin a ser obertes, prèviament s'han d'anonimitzar**. Si es detecta que es pot posar en risc algun dret fonamental, el delegat de protecció de dades, en col·laboració amb el personal investigador, ha de fer una avaluació de l'impacte en la privacitat de les persones (Privacy Impact Assessment, o PIA).

Finalment, un altre aspecte que cal tenir en compte és la **validació ètica que pugui necessitar una recerca** quan es faci amb persones. En aquest cas, encara que no es tractin directament dades personals, es necessita una validació (com en el cas d'enquestes, enregistraments, etc. fets a persones). Aquest tipus de validacions **les fan els comitès ètics, interns o externs**, en resposta als mateixos codis de conducta institucionals, a sol·licituds d'entitats finançadores, o per imperatiu legal, com passa, per exemple, en l'àmbit de la recerca biomèdica¹² o de l'experimentació amb animals. La legislació determina el requisit d'avaluació per part d'un comitè ètic reconegut, però no estableix requisits formals.

Informació addicional rellevant:

¹¹ <https://gdpr.eu/>

¹² <https://www.boe.es/eli/es/l/2007/07/03/14>

Aquests comitès ètics poden facilitar ajuda i exemples a l'hora d'elaborar documents de consentiment informat. De nou, és important que el personal investigador del centre conegui els procediments establerts a aquest efecte.

ACCIONS RECOMANADES

- Establir procediments i personal (delegat de protecció de dades i comitès ètics) per minimitzar impactes i riscos en tractar amb dades personals i sensibles. El personal investigador ha de tenir presents, en tot moment, les indicacions dels centres al respecte.

3.3 La gestió de la propietat intel·lectual

La **propietat intel·lectual es refereix als drets sobre qualsevol producció**, com les obres artístiques i científiques, així com les bases de dades, **que corresponen als autors**, fruit de la seva intel·ligència i creativitat, **i a altres titulars**. El terme *propietat intel·lectual* sovint s'estén al concepte de **propietat industrial**, que inclou dissenys industrials, patents i models d'utilitat, marques i noms comercials, així com topografies de semiconductors.¹³ **A Espanya aquests dos conceptes estan regulats per lleis** diferents que atorguen condicions i durades diferents als titulars dels drets.

La gestió de la propietat intel·lectual s'ha de plantejar des de la planificació, i cal identificar quins drets¹⁴ es generen (d'exploració, de reconeixement, etc.), **qui n'és el titular**, si es podran **reutilitzar les dades generades** i sota quines **condicions**.

En molts casos, les dades, de forma individual, no són protegibles per aquesta via pel fet que no es tracta de creacions originals que es puguin considerar obres, per exemple,

Informació addicional rellevant:

- Avaluacions d'impacte de protecció de dades, Agència Espanyola de Protecció de Dades (AEPD): <https://www.aepd.es/es/derechos-y-deberes/cumple-tus-deberes/medidas-de-cumplimiento/evaluaciones-de-impacto>
- Guia pràctica per a les avaluacions d'impacte en la protecció de les dades subjectes al RGPD, Agència Espanyola de Protecció de Dades (AEPD): <https://www.aepd.es/sites/default/files/2019-09/guia-evaluaciones-de-impacto-rgpd.pdf>
- *Ethics and data protection*. Comissió Europea, 14 de novembre de 2018: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-data-protection_en.pdf
- *Consentiment informat del participant (model orientatiu)*. Comissió de Bioètica de la Universitat de Barcelona: http://www.ub.edu/comissibiocetica/sites/default/files/documents/pagina-basica/model_orientatiu_consentiment_es.pdf
- *El consentimiento informado en la investigación en enfermedades raras*. Teresa Pàmpols, Carmen Ayuso i Guillem Pintos: <https://www.ciberisciii.es/media/602614/14-cap-9.pdf>
- Xarxa de Comitès d'Ètica d'Universitats i Organismes Públics de Recerca (RCE): <https://redcomitesetica.es/>

¹³ https://www.oepm.es/es/propiedad_industrial/propiedad_industrial/

¹⁴ Vegeu el capítol III, articles 14 a 25, del Reial decret legislatiu 1/1996, de 12 d'abril, pel qual s'aprova el text refós de la Llei de propietat intel·lectual, en què es regularitzen, s'aclareixen i s'harmonitzen les disposicions legals vigents sobre la matèria.

quan es tracta d'uns valors sorgits d'un experiment, unes dades de temperatura observades o unes dades. Pel que fa a les **compilacions o bases de dades, la llei de propietat intel·lectual** vigent¹⁵ a Espanya, així com la majoria de les lleis vigents a Europa,¹⁶ inclouen el **reconeixement del dret sui generis**, que atorga el dret exclusiu a extreure i reutilitzar els continguts al creador d'una base de dades. Per tant, aquest titular del dret pot concedir els permisos necessaris per reutilitzar les dades contingudes en una base de dades independentment de si les dades, individualment, també estan

protegides per propietat intel·lectual. Aquest aspecte és important per als centres, tant pel seu caràcter generador de noves dades com per la reutilització de dades ja existents en les seves investigacions. A més, cal tenir en compte que si la recerca que s'està desenvolupant pot donar lloc a resultats objecte de propietat industrial, caldrà establir els mecanismes necessaris per no revelar cap informació. Així, les dades que puguin contribuir a generar una patent han de romandre tancades durant un temps prudencial que eviti posar en risc la seva concessió.

ACCIONS RECOMANADES

- **Proporcionar suport en matèria de propietat intel·lectual al personal investigador, orientar sobre la política del centre i posar a la seva disposició personal preparat per analitzar situacions i prendre decisions.**
- **Establir una política clara respecte a la titularitat dels drets i els ingressos que pugui generar la creació de dades, així com la seva compilació, i difondre aquesta política entre el personal investigador, per tal que conegui les condicions a què està subjecte (si pot compartir dades a la seva voluntat, etc.).**
- **Definir clarament qui és el titular dels drets en cada situació (el personal investigador, el centre o la seva entitat matriu, el consorci o l'entitat finançadora), ja que és imprescindible per aplicar llicències i permetre possibles usos de les dades.**

Informació addicional rellevant:

És habitual que les institucions de recerca tinguin polítiques, normes i procediments clars per tractar la propietat industrial, però no sempre n'hi ha per al cas de la propietat intel·lectual i, quan n'hi ha, en alguns casos, només inclouen alguns tipus d'obres, com ara el programari i les bases de dades. En la majoria de casos, la titularitat dels drets d'explotació recau sobre la institució i qualsevol ingrés derivat de la seva explotació comercial es divideix entre els investigadors i la institució a què estan adscrits.

En cas que el centre o el seu personal investigador participi en un projecte finançat per una organització externa, cal estudiar si hi ha algun requisit respecte a la propietat, mirant de buscar sempre l'equilibri entre aquests requisits i les normes del centre. D'altra banda, la contractació puntual de personal investigador per col·laborar en un projecte que genera dades i que és susceptible de produir drets ha d'incloure un apartat que remeti a la normativa del centre, ja que es pot donar el cas que un investigador reculli dades i després les faci servir en altres projectes sense que el centre en tingui coneixement.

15 Reial decret legislatiu 1/1996, de 12 d'abril, pel qual s'aprova el text refós de la Llei de propietat intel·lectual, en què es regularitzen, s'aclareixen i s'harmonitzen les disposicions legals vigents sobre la matèria.

16 Directiva 96/9/CE del Parlament Europeu i del Consell, d'11 de març de 1996, sobre la protecció jurídica de les bases de dades.

Capítol 4

TREBALLAR
AMB DADES



4

TREBALLAR AMB DADES

Quan es dissenya una recerca, en primer lloc s'identifica quin tipus de dades es necessiten i s'analitza la font més adequada, la reutilització de dades existents o la generació pròpia de dades. Un cop presa aquesta decisió, cal considerar **quins formats es faran servir, com s'organitzaran les dades, i on i com s'emmagatzemaran i es publicaran.**

4.1 Reutilització de dades existents

De vegades el personal investigador inicia la seva activitat a partir de dades ja existents, que **poden ser d'accés lliure o requerir un permís d'accés específic.** Aquesta pràctica pot ser habitual en determinades disciplines, per exemple, en les ciències socials, quan s'utilitzen dades estadístiques, o en la realització d'estudis epidemiològics a partir de dades sanitàries.

El primer pas **quan s'accedeix a dades externes** és, per tant, **analitzar quines són les condicions d'accés, ús i publicació.** En aquest sentit, és important conèixer la diferència entre *dades públiques* i *dades obertes*, ja que són termes que solen induir a confusió. **Les dades són públiques si s'hi pot accedir per consultar-les sense cap tipus de registre**, per exemple, quan estan publicades en un portal de dades. No obstant això,

perquè les dades puguin ser reutilitzades cal que hi hagi una llicència o un avís que indiqui **quins usos se'n permeten i en quines condicions.** És aleshores quan es pot parlar de *dades obertes*. **L'absència de llicència o avís s'ha d'entendre en termes de «tots els drets reservats».** D'altra banda, encara que les dades individualment poden no ser protegibles per propietat intel·lectual, sí que ho pot ser la base de dades que les conté, com ja hem vist. L'extracció i la reutilització de la totalitat o d'una part substancial dels continguts d'una base de dades requereix una autorització específica, llevat que ja s'hagi concedit prèviament a través de la llicència d'ús.

Els portals de dades obertes que ofereixen les administracions públiques són un clar exemple de dades obertes. Aquestes dades solen estar publicades sota llicències que permeten reutilitzar-les àmpliament, com les llicències CC BY¹⁷ o CC0¹⁸, que s'analitza en el proper capítol. És el cas de l' Institut Nacional d'Estadística (INE), que permet consultar dades sense cap tipus de registre (dades públiques) i disposa, a més, d'una secció de dades obertes¹⁹ en què s'especifiquen les condicions de reutilització. La Xunta de Galícia²⁰ o el Govern d'Espanya²¹ també disposen dels seus respectius portals de dades obertes.

Informació addicional rellevant:

¹⁷ <https://creativecommons.org/licenses/by/4.0/>

¹⁸ <https://creativecommons.org/share-your-work/public-domain/cc0/>

¹⁹ <https://www.ine.es/datosabiertos>

²⁰ <https://abertos.xunta.gal/>

²¹ <https://datos.gob.es/>

En altres supòsits, l'accés a les dades està més restringit i la reutilització i la difusió d'aquestes dades, a priori, no està permesa o presenta limitacions. És el cas de les microdades d'Eurostat, que només són accessibles per a propòsits científics i exigeixen registrar-se.²² En qualsevol cas, una mateixa institució pot estar oferint dades sota diferents condicions d'ús simultàniament, tal com fa l'INE, que especifica al seu lloc web que la informació procedeix de múltiples fonts, i autoritza la reutilització de la informació si la font original és la institució mateixa.

Quan es generen dades amb vista a la seva futura reutilització o quan es reutilitzen dades existents, un aspecte que cal tenir en compte és la seva citació a les fonts de la recerca. A iniciativa de FORCE11, una comunitat d'acadèmics, bibliotecaris, editors i finançadors de la recerca, s'han proposat uns principis²³ per a la citació correcta de les dades, que inclouen unes bones pràctiques per al reconeixement de l'autoria, l'accés a les dades i la seva verificació. Aquesta comunitat també reivindica que les dades tinguin la mateixa consideració que qualsevol resultat de recerca. Tot seguit es mostren exemples de citació que segueixen aquestes recomanacions:

Exemples

- King, G.; Zeng, L. (2007). «Replication data for: When Can History be Our Guide? The Pitfalls of Counterfactual Inference». <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/EK886K>
- Barnett, C. L.; Wells, C.; Thacker, S.; Guyatt, H. J.; Fletcher, J. M.; Lawlor, A. J.; Winfield, I. J.; Beresford, N. A. (2015). «Elemental concentrations in fish from lakes in Northwest England». NERC Environmental Information Data Centre. (Dataset). <https://doi.org/10.5285/ed90df1b-462c-46bb-afbd-59794fb03f6b>

ACCIONS RECOMANADES

- Analitzar com es poden reutilitzar les dades i si es podran publicar a posteriori (quins usos estan permesos i en quines condicions).
- Incloure els elements principals següents a la citació de dades: autors, títol del conjunt de dades, repositori o arxiu, any, versió i identificador permanent.

Informació addicional rellevant:

²² <https://ec.europa.eu/eurostat/web/microdata/overview>

²³ <https://www.force11.org/datacitationprinciples>

4.2 Formats i organització

Si es vol que les dades siguin FAIR cal triar bé els formats que s'utilitzaran. De vegades els formats estan condicionats per l'aparell o el dispositiu amb què es treballi en la recerca, però altres vegades hi ha la possibilitat d'escollir el format que més convingui. La **recomanació sempre és utilitzar formats estàndard i no dependents de programes privatis ni de programari especialitzat**, si és possible.

L'elecció dels formats és, per tant, un aspecte crucial a l'hora de compartir les dades i fomentar-ne la reutilització, ja que pot obligar a transformar posteriorment les dades si no compleixen els requisits de preservació del repositori on s'han d'arxivar. En cas que només sigui possible

desar les dades en un format propietari, es recomana acompanyar-les d'una indicació del programa utilitzat per generar-les, així com la versió específica del programa.

A priori **no es pot establir una llista tancada de formats perquè cada disciplina pot utilitzar diferents formats establerts i consolidats per la comunitat**. Tanmateix, es pot oferir una llista amb recomanacions o formats de preferència, que poden estar condicionats per la política de preservació. En el cas de Data Archiving and Networked Services (DANS), el centre nacional neerlandès de dades de recerca, s'estableixen uns formats de preferència per a diferents tipus de dades²⁴. A la taula següent s'indiquen alguns dels formats de dades recomanats per DANS.

TIPUS	PREFERÈNCIA	NO PREFERITS
TEXT	PDF/A, ODT	DOC, DOCX, PDF (no PDF/A)
LLENGUATGE DE MARQUES	XML, HTML	SGML, MD
FULL DE CàLCUL	ODS, CSV	XLS, XLSX, PDF/A
DADES ESTADÍSTIQUES	DAR, SPS, DO, R	PER, SAV, DTA
IMATGES VECTORIALS	SVG	AI, EPS, WMF

Font: DANS.

Informació addicional rellevant:

24 <https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depositing/file-formats>

D'altra banda, hi ha formats que poden ser recomanables per a determinats tipus de dades, però que en poden dificultar la reutilització (per exemple, una taula en format PDF). Per això, **a l'hora de seleccionar un format s'ha de valorar també tant el temps que requereix com els problemes que en pot comportar l'ús.**

Per la seva banda, l'organització de les dades es refereix a la manera com aquestes s'emmagatzemen en l'espai de treball, que ha de ser estructurada i consistent. Aquest aspecte és important quan les dades s'han de compartir o cal facilitar l'accés a espais comuns o carpetes compartides a persones de la mateixa organització o d'organitzacions diferents.

A més, per facilitar l'organització de les dades, els noms dels fitxers han de ser explicatius. En el cas de les **dades dinàmiques**, també es recomana seguir un control de versions i indicar de quina versió es tracta. Aquesta indicació es pot mantenir quan es vagin publicant les dades als repositoris corresponents.

ACCIONS RECOMANADES

- **Seleccionar i difondre formats recomanats pel centre, així com la metodologia d'organització de les dades, de denominació dels fitxers i d'assignació de versions.**

4.3 Emmagatzematge i accés

Quan es treballa amb dades és molt important **determinar els tipus de dispositius que s'utilitzaran per emmagatzemar-les durant el desenvolupament de l'activitat investigadora.** Una elecció equivocada podria comportar la pèrdua de les dades, cosa que tindria conseqüències fatals per a la recerca.

Un centre de recerca ha d'oferir als investigadors un espai de treball on emmagatzemar les dades amb control d'accés, que permeti **establir els permisos i configurar,** per tant, **els perfils d'accés** per treballar amb unes dades que encara no són aptes per a la difusió. L'accés compartit s'ha de plantejar tant per a usuaris interns com externs, ja que és habitual que en els projectes de recerca es treballi en col·laboració amb altres entitats.

També s'ha d'**establir un procés de còpies de seguretat** per assegurar que no es perdin les dades, fins i tot en circumstàncies excepcionals.

Així mateix, el personal investigador ha de conèixer com ha de procedir amb el maneig de les dades per evitar les pèrdues d'informació i ser conscient que els dispositius com els llapis de memòria USB o els ordinadors personals són eines ràpides i fàcils d'utilitzar per emmagatzemar les dades, però no són les més adequades pel que fa a la seguretat de la informació.

Informació addicional rellevant:

Per a més informació sobre l'organització de dades:

- *UK Data Service*: <https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/organising/>
- *Research Data Netherlands (RDNL)*: <https://datasupport.researchdata.nl/en/start-the-course/iii-research-phase/organising-data>

D'altra banda, els espais d'emmagatzematge que s'utilitzen durant la recerca poden estar connectats amb els repositoris on es dipositaran les dades un cop finalitzada la recerca, que es poden fer servir per publicar i preservar les dades.

En aquest sentit, cal esmentar l'exemple de la Universitat d'Utrecht, que ofereix

als investigadors diferents solucions per emmagatzemar les dades, tant gratuïtes com de pagament. Per ajudar-los a triar, han elaborat un qüestionari dinàmic²⁵ que, a mesura que es va emplenant, va descartant opcions fins a arribar, finalment, a la solució més apropiada.

ACCIONS RECOMANADES

- Determinar els dispositius que es faran servir per treballar amb les dades.
- Disposar d'un espai de treball on emmagatzemar les dades amb control d'accés, que permeti establir permisos i determinar rols d'actuació per treballar amb les dades.
- Definir un sistema de còpia de seguretat de les dades.
- Definir pautes d'actuació per al maneig de les dades. Difondre, formar i conscienciar el personal del centre sobre la importància del seguiment.
- Seleccionar possibles solucions alternatives per emmagatzemar dades i facilitar al personal investigador informació sobre com pot seleccionar la més adequada en cada cas.

Informació addicional rellevant:

²⁵ <https://tools.uu.nl/storagefinder/>



Capítol 5

COMPARTIR,
PUBLICAR I
PRESERVAR DADES

5

COMPARTIR, PUBLICAR I PRESERVAR DADES

Aquest capítol analitza la **darrera fase del cicle de gestió de les dades: com s’han de compartir, publicar i preservar**, i aborda de manera detallada l’elecció i el paper dels diferents tipus de repositori existents.

La figura següent reflecteix el procés de les dades primàries fins a la seva publicació en un repositori en accés obert o accés restringit:

- Es parteix d’unes **dades primàries** o de la reutilització de dades existents.
- Després de la seva obtenció, es generen unes **dades secundàries** mitjançant el tractament o el processament de les dades.
- Les dades secundàries es poden **descartar, arxivar o bé publicar** en un repositori.

- El repositori pot ser **temàtic, institucional o consorciat**.
- La publicació de les dades en un repositori ha d’exposar **sempre les metadades en obert** perquè la informació pugui ser identificada. En cas que les dades s’ofereixin en obert, s’han d’acompanyar d’una **llicència d’ús**. En canvi, **quan l’accés a les dades és restringit**, s’ha **d’afegir informació sobre com es poden utilitzar**, en cas que se segueixin els principis FAIR. Una possibilitat és oferir un document estàndard d’acord d’ús de dades (*data user agreement*, o DUA) que l’usuari ha de signar i enviar al titular de les dades per poder-les fer servir. En qualsevol cas, la publicació en un dipòsit hauria de comportar l’assignació d’un identificador permanent (PID, de l’anglès *persistent identifier*) com el DOI.

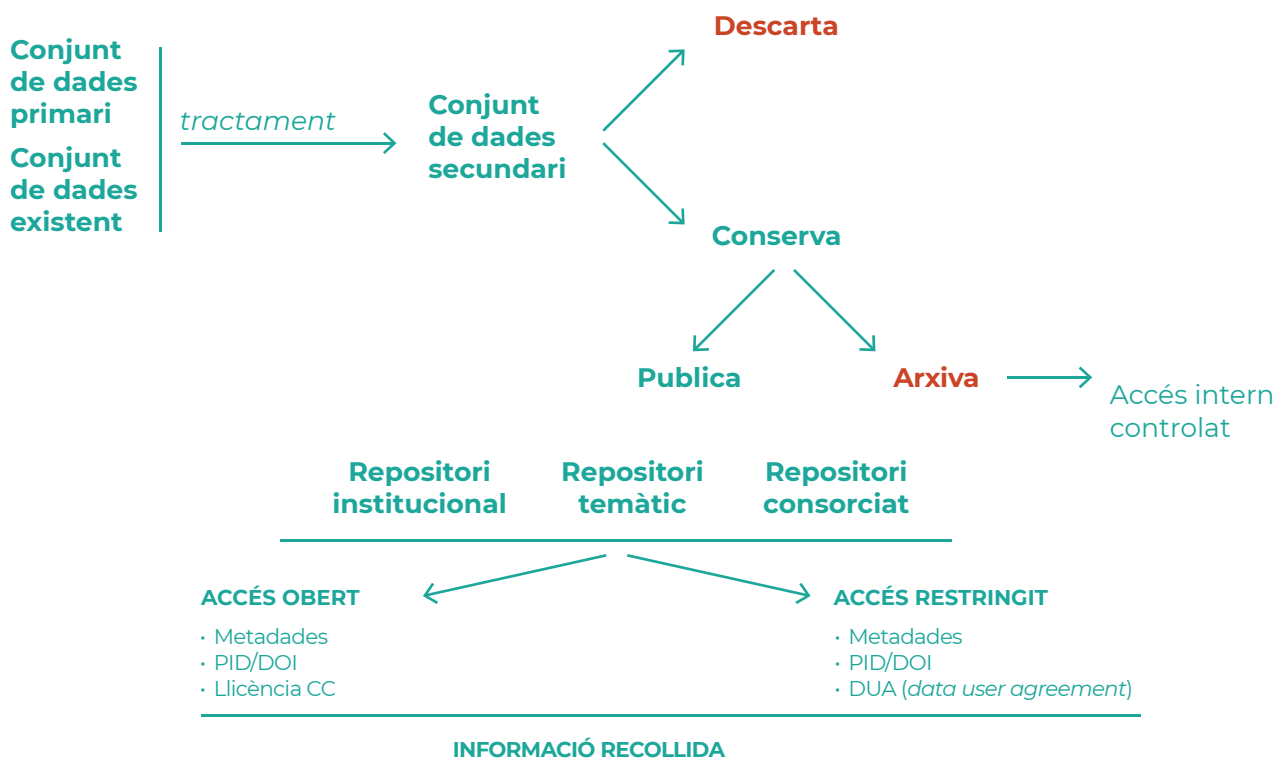


Figura 4. Procés de les dades primàries fins a la seva publicació en un repositori en accés obert o accés restringit.

Font: elaboració pròpia.

Com ja hem comentat al llarg d'aquesta guia, els principis FAIR serveixen com a orientació per a una bona gestió de les dades de recerca. **En cas que les dades es vulguin compartir públicament cal triar un repositori que també segueixi els principis FAIR**, és a dir, que entre altres característiques ofereixi les següents:

- Emprar un identificador únic.
- Fer servir un esquema de metadades estàndard.
- Permetre l'ús de llicències per a dades.
- Utilitzar un sistema de preservació.

L'**aplicació del principi de reutilització** que estableixen els principis FAIR **requereix un tractament diferenciat en el cas de les dades obertes i tancades**. Per això, els primers epígrafs del capítol es dediquen a aquest aspecte. D'una banda, **s'analitzen les diferents llicències que es poden utilitzar per permetre aquesta reutilització de les dades** i, de l'altra, **s'explica com es poden tractar les dades tancades per compartir-les de manera agregada o anònima i com es pot informar d'una reutilització restringida sotmesa a condicions**.

5.1 Compartir dades

5.1.1 Dades obertes

Com s'explica més amunt, **les dades obertes són dades que no requereixen autorització d'accés, exploració, anàlisi ni reutilització**. També es pot considerar que les dades són

obertes **quan les condicions d'accés són mínimes** (per exemple, mitjançant un registre previ o l'acceptació d'uns termes d'ús molt laxos). **El tipus de control o autorització necessaris per accedir a les dades** determina quan es poden considerar «obertes». Aquest concepte **no implica necessàriament garanties sobre la qualitat o el valor científic de les dades**, ni sobre la seva possible utilitat futura.

Si es volen compartir les dades de manera oberta, és a dir, **permetent-ne a priori la reutilització sense que els usuaris hagin de demanar permís, cal fer servir algun tipus de llicència**. Possiblement les llicències per a continguts oberts més conegudes són les que ofereix l'organització Creative Commons.²⁶ Des del 2002 aquesta organització ofereix un conjunt de llicències que permet autoritzar la reutilització amb diferents opcions. Són llicències que no es van crear específicament per a dades, però que ja des de la seva versió 4 contenen una clàusula que indica la seva aplicació als possibles drets que hi hagi sobre una base de dades, de manera que:

- Si la llicència limita l'explotació comercial, s'ha d'entendre que es poden reutilitzar els continguts d'una base de dades per crear una altra base de dades que no es pot comercialitzar.
- Si es limiten les obres derivades, no es poden crear bases de dades amb els elements de la base de dades subjecta a la llicència. S'entén com a obra derivada aquella que conté la totalitat o una part significativa dels elements d'una base de dades existent.

Informació addicional rellevant:







Per aprofundir sobre les característiques tècniques del servei de generació de DOI o PID, vegeu aquest article:

<https://www.biblogtecarios.es/rafaelibanez/pid-identificadores-persistentes/>

²⁶ <https://creativecommons.org/licenses/>

● Si s'hi inclou la condició *copyleft*, és a dir, que es mantingui la llicència en potencials obres derivades, s'ha de mantenir la llicència original en les bases de dades que es creïn.

A la taula següent s'indiquen les llicències estàndard de Creative Commons (CC).

	<p>Reconeixement</p>	<p>Permet la reproducció, la distribució i la comunicació pública de l'obra original, així com la creació d'obres derivades, sense cap restricció excepte l'obligació de reconèixer-ne l'autoria i mantenir l'avís de la llicència.</p>
	<p>Reconeixement Compartir Igual</p>	<p>Permet la reproducció, la distribució i la comunicació pública de l'obra original, així com la creació d'obres derivades, sense cap restricció excepte l'obligació de reconèixer-ne l'autoria i mantenir l'avís de la llicència. En el cas de les obres derivades, s'han de difondre amb la mateixa llicència.</p>
	<p>Reconeixement sense obra derivada</p>	<p>Permet la reproducció, la distribució i la comunicació pública de l'obra original sense cap restricció excepte l'obligació de reconèixer-ne l'autoria i mantenir l'avís de la llicència. No es permet crear obres derivades.</p>
	<p>Reconeixement No Comercial</p>	<p>Permet la reproducció, la distribució i la comunicació pública de l'obra original, així com la creació d'obres derivades, sempre que no es faci amb finalitats comercials. Cal reconèixer-ne l'autoria i mantenir l'avís de llicència.</p>
	<p>Reconeixement No Comercial Compartir Igual</p>	<p>Permet la reproducció, la distribució i la comunicació pública de l'obra original, així com la creació d'obres derivades, sempre que no es faci amb finalitats comercials. Cal reconèixer-ne l'autoria i mantenir l'avís de llicència. En el cas de les obres derivades, s'han de difondre amb la mateixa llicència.</p>
	<p>Reconeixement No Comercial Sense Obra Derivada</p>	<p>Permet la reproducció, la distribució i la comunicació pública de l'obra original sempre que no es faci amb finalitats comercials. Cal reconèixer-ne l'autoria i mantenir l'avís de llicència. No es permet crear obres derivades.</p>

Font: elaboració pròpia.

Un altre instrument legal que ofereix Creative Commons és CC0.²⁷ Aquesta eina normalment es presenta com una renúncia als drets de propietat intel·lectual, però en realitat és una mica més complexa. Si s'analitza el text legal s'observa que hi ha tres nivells d'actuació:

1. El primer és la renúncia: el titular que té algun dret hi renuncia, però com que en la legislació espanyola (i en altres) hi ha drets irrenunciables i intransferibles, s'indica que el titular dels drets renuncia a tots els drets als quals pot renunciar.
2. Aquí intervé el segon nivell: el titular concedeix a l'usuari allò que es pot denominar una llicència sense condicions sobre tots els drets a què no ha pogut renunciar.
3. Finalment, intervé el tercer nivell: el titular es compromet a no perseguir l'incompliment. Per tant, el titular dels drets no ha d'emprendre cap acció legal si no se'n reconeix l'autoria, encara que es tracti d'un dret moral irrenunciable. D'aquesta manera, es posa més el focus en les normes de la comunitat investigadora que no pas en un

estricte compliment de la llei de propietat intel·lectual.

Una altra opció per llicenciar dades és utilitzar el conjunt de llicències que ofereix la iniciativa Open Data Commons (ODC).²⁸ Aquesta iniciativa impulsada per l'Open Knowledge Foundation ofereix tres llicències:

- 1. Open Data Commons Open Database License (ODbL):** inclou el requeriment *copyleft*, de manera equivalent al que actualment es requereix amb la llicència CC BY-SA.
- 2. Open Data Commons Attribution License:** només requereix un reconeixement, com la llicència CC BY.
- 3. Open Data Commons Public Domain Dedication and License (PDDL):** cessió al domini públic equivalent a l'opció CC0 de Creative Commons.

A la taula següent es poden veure les equivalències entre les llicències ODC i les llicències de Creative Commons.

Llicència ODC	ODC ODbL	ODC Attribution License	ODC PDDL
Equivalència a CC	CC BY-SA	CC BY	CC0

Font: elaboració pròpia.

Finalment, cal destacar que alguns països han optat per generar les seves pròpies

llicències nacionals, com és el cas de França²⁹ o el Regne Unit³⁰.

Informació adicional rellevant:

27 <https://creativecommons.org/choose/zero/>

28 <https://opendatacommons.org/>

29 <https://www.etalab.gouv.fr/wp-content/uploads/2017/04/ETALAB-Licence-Ouverte-v2.0.pdf>

30 <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

ACCIONS RECOMANADES

- Analitzar la titularitat i les llicències associades a totes les dades que s'utilitzen en una recerca.
- Definir clarament qui és el titular de tots els drets (un tercer —reutilització—, l'investigador, el centre o el finançador), ja que serà el titular qui podrà autoritzar l'ús d'aquestes dades.

5.1.2 Dades tancades

Tot i que hi ha la tendència de promoure l'obertura de dades de recerca, també **hi ha moltes situacions en què les dades s'han de mantenir tancades**, per exemple, **quan es treballa amb dades personals, confidencials o sota normes de seguretat**. És important insistir que els principis FAIR no requereixen que les dades siguin obertes, sinó que s'informi sobre com es poden reutilitzar.

En aquests casos **es pot establir un acord d'ús**, com s'ha esmentat més amunt, **en què s'indiqui quins usos estan permesos i en quines condicions**, tenint presents els possibles aspectes legals o normatius d'aplicació, com els referents a la utilització de dades personals (Reglament europeu de protecció de dades³¹ i legislació nacional³²). Per concedir l'accés a aquestes dades que no són públiques cal esperar a tenir l'acord signat per ambdues parts.

Altres opcions per compartir dades que no es poden fer públiques són l'**anonimització**,

l'**agregació** o l'ús de tècniques com la **privacitat diferencial**, que s'han de dur a terme correctament per no posar en perill la confidencialitat, la seguretat o la privacitat de les dades i els subjectes implicats:

- L'**anonimització de dades personals** és el procés a través del qual **es trenca la connexió entre la dada personal i la persona física a què es refereix**, per evitar-ne la identificació. Les dades resultants es denominen *anonimitzades* i no entren dins dels casos protegits pel Reglament general de protecció de dades o altres normatives d'aplicació. Quan l'element (programari, taula de valors, etc.) que permet la reidentificació de la persona a què es refereixen les dades apareix simplement aïllat i allunyat de les dades, i no destruït, es parla de *pseudonimització*, i les dades segueixen considerant-se dades personals. En aquest sentit, es pot consultar la guia creada per l'Agència Espanyola de Protecció de Dades per orientar i garantir els procediments d'anonimització de dades personals.³³

Informació addicional rellevant:

31 Reglament general de protecció de dades (Reglament UE 2016/679): <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32016R0679&from=EN>

32 En el cas espanyol, la Llei orgànica 3/2018, de 5 de desembre, de protecció de dades personals i garantia dels drets digitals: <https://www.boe.es/eli/es/lo/2018/12/05/3>

33 <https://datos.gob.es/es/documentacion/orientaciones-y-garantias-en-los-procedimientos-de-anonimizacion-de-datos-personales>

- **L'agregació** consisteix en la compilació de dades eliminant la informació individual de cada dada. En pot ser un exemple l'obtenció d'informació sobre quanta gent ha fet un desplaçament determinat, sense revelar qui l'ha fet individualment.

- **La privacitat diferencial**,³⁴ finalment, consisteix en un conjunt de sistemes i pràctiques que **permeten extreure conclusions sobre un grup mantenint en tot moment la privacitat de les dades de cadascun dels seus membres**. Un exemple seria introduir errors o soroll a les dades d'un cens sense alterar les dades totals.³⁵

ACCIONS RECOMANADES

- Establir models d'acords d'ús restringit de dades.
- Publicar les metadades corresponents a les dades tancades, així com donar la possibilitat d'accedir-hi de manera restringida.
- Establir les llicències utilitzables al centre per compartir dades.
- Oferir un servei d'anonimització de dades.

Informació addicional rellevant:

³⁴ https://en.wikipedia.org/wiki/Differential_privacy

³⁵ <https://www.technologyreview.es/s/11974/tr10-privacidad-diferencial>

Guia *Orientaciones y garantías en los procedimientos de anonimización de datos personales* de l'Agència Espanyola de Protecció de Dades: <https://datos.gob.es/es/documentacion/orientaciones-y-garantias-en-los-procedimientos-de-anonimizacion-de-datos-personales>.

5.2 Repositoris

A l'hora de preservar i compartir les dades es pot optar per **diferents alternatives de repositoris**:

- La primera opció consisteix a **seleccionar un repositori dedicat a una determinada disciplina o àmbit de coneixement**, quan existeixi i estigui consolidat.
- La segona opció es basa a decidir entre **un repositori institucional o un repositori consorciat**.
- Finalment, també es pot valorar la possibilitat de disposar d'un **repositori multidisciplinari**.

El volum dels fitxers, el format i les condicions de preservació també poden influir en l'elecció del repositori més adequat. Tot seguit, s'analitza cadascuna d'aquestes opcions.

5.2.1 Repositori temàtic

En algunes disciplines hi ha **repositoris dedicats a compartir dades**. Podem citar exemples fàcilment recognoscibles com el GenBank,³⁶ un repositori de seqüències d'ADN, o el GBIF,³⁷ un repositori de dades de

biodiversitat, però altres repositoris no són tan coneguts, ni tan sols per als investigadors mateixos. A l'hora d'identificar repositoris, pot ser d'ajuda el registre re3data,³⁸ que ofereix una llista de repositoris de dades de recerca i permet la consulta en funció de diferents paràmetres com ara l'àmbit científic, el tipus d'accés o les llicències d'ús, entre d'altres. Una altra alternativa pot consistir a fer servir el portal FAIRsharing,³⁹ que facilita l'aplicació de diversos filtres per localitzar el repositori més adequat.

La primera opció és sempre dipositar les dades en un repositori temàtic consolidat entre la comunitat científica d'un àmbit o disciplina determinats. El centre de recerca ha d'analitzar les alternatives i oferir una llista d'aquests repositoris als investigadors, en funció de les àrees en què treballin.

Dipositar les dades en repositoris temàtics **no vol dir que el centre se'n desvinculi, sinó que pot crear un catàleg de dades amb la descripció i l'enllaç corresponent**, de la mateixa manera que es recullen altres resultats de recerca, com les publicacions. En aquest catàleg s'hi han de poder trobar tant les dades dipositades en repositoris externs com en repositoris interns.

Informació addicional rellevant:

Recomanacions per seleccionar un repositori per al dipòsit de dades de recerca: Versió 5, Octubre 2020:

<http://hdl.handle.net/2072/377586>

36 <https://www.ncbi.nlm.nih.gov/genbank/> **37** <https://www.gbif.org/>

37 <https://www.gbif.org/>

38 <https://www.re3data.org/>

39 <https://fairsharing.org/databases/>

5.2.2 Repositori institucional

Una altra alternativa consisteix a utilitzar un **repositori propi del centre o bé un repositori d'alguna institució vinculada**, com una universitat. Dins d'aquesta mateixa alternativa també es té en compte la creació d'un **repositori institucional consorciat**, en què participen diferents institucions. **El seu principal avantatge és optimitzar recursos i serveis.**

En qualsevol cas, **el centre ha de designar una persona que coordini el dipòsit de les dades al repositori.** A més, si s'opta per crear un repositori propi, **cal dotar-lo d'un suport tècnic i fins i tot destinar una persona per al manteniment i l'evolució d'aquesta infraestructura.**

Als repositoris consorciats normalment s'estableixen unes pautes de treball comunes, però cada centre o institució s'ha de responsabilitzar del seu espai i del contacte amb els seus investigadors per dipositar les dades, i no cal el manteniment tècnic individual.

Moltes institucions que ja tenen un repositori institucional consolidat i inicialment destinat a documents han optat per utilitzar-lo també com a repositori de dades. Així, han creat col·leccions específiques per a dades i fins i tot per a un altre tipus de resultats com el programari. És el cas de Digital CSIC, el repositori del Consell Superior d'Investigacions Científiques (CSIC), que inclou una col·lecció específica de conjunts de dades per als seus centres a cadascuna de les seves seccions.⁴⁰

Altres institucions han escollit crear un repositori dedicat exclusivament a dades. És el cas de la Universitat d'Edimburg amb l'Edinburgh DataShare,⁴¹ o de la Universitat d'Amsterdam⁴². En el primer cas s'utilitza un repositori instal·lat a la universitat mateixa, mentre que en el segon s'ha contractat la infraestructura a un proveïdor extern.

També hi ha múltiples exemples de repositoris consorciats, com el madrileny e-cienciaDatos,⁴³ que agrupa les universitats que formen part del Consorci Madroño, o els repositoris nacionals dels Països Baixos (DataverseNL⁴⁴) o de Noruega (DataverseNO⁴⁵).

En cas d'optar per un repositori institucional, ja sigui propi o consorciat, s'ha de definir una política de preservació. D'una banda, cal pensar en una estratègia de còpies de seguretat distribuïdes i, de l'altra, establir un període de temps durant el qual s'asseguri la preservació de les dades. En molts casos aquest període es fixa en 10 anys, encara que s'ha de valorar si hi ha alguna obligació de mantenir les dades durant períodes de temps superiors.

5.2.3 Repositori multidisciplinari extern

Si no hi ha la possibilitat de disposar d'un repositori temàtic o d'un repositori institucional, les dades es poden dipositar en un repositori multidisciplinari extern. Cal tenir en compte que **aquesta alternativa**, de vegades utilitzada individualment pel personal investigador, **sol ser atractiva per la seva gratuïtat inicial, però generalment l'espai disponible i la mida dels conjunts de dades estan limitats.** En cas que

Informació addicional rellevant:

40 <https://digital.csic.es/handle/10261/166574> 41 <https://datashare.ed.ac.uk/>

41 <https://datashare.ed.ac.uk/>

42 <https://uva.uva.figshare.com/uva>

43 <https://edatos.consorciomadrono.es/>

44 <https://dataverse.nl/>

45 <https://dataverse.no/>

es vulguin reduir aquestes limitacions, **es pot optar per contractar un servei de pagament**, encara que aquesta decisió requereix valorar conjuntament tant les necessitats dels investigadors del centre com les prestacions que ofereix el repositori i altres alternatives.

Alguns dels proveïdors d'aquests repositoris ofereixen solucions globals a les institucions per convertir-se en un recurs estable i acollir la col·lecció de dades com a repositori institu-

cional. És el cas de la Universitat d'Amsterdam, citat a l'apartat anterior.

El grup de treball de suport a la recerca del CSUC manté una llista amb les diferents característiques d'aquests repositoris multidisciplinaris. L'última versió data d'octubre del 2020.⁴⁶ A la taula següent s'ofereix una mostra d'alguns d'aquests repositoris i les seves característiques.

Repositori	Figshare	Mendeley Data	Zenodo
Proveïdor	Digital Science	Elsevier	CERN
URL	https://figshare.com	https://data.mendeley.com/	https://zenodo.org/
Mida dels fitxers	Fins a 5 GB per fitxer. Fins a 20 GB d'espai privat	Fins a 10 GB per a un conjunt de dades. Ampliable si la institució se subscriu a Mendeley Data	Fins a 50 GB per a un conjunt de dades. Es pot contactar per a mides superiors
Llicències d'ús	Per defecte. CCO per a dades i CC BY per a altres tipus de material	Permet varietat de llicències per a dades: CCO, CC BY, CC BY-NC	Permet varietat de llicències, encara que recomana llicències obertes
Identificador permanent per a les dades	DOI	DOI	DOI
Cost	Gratuït. Hi ha la possibilitat d'obtenir més espai pagant	Gratuït	Gratuït, però no descarten que en el futur s'hagi de pagar si se supera un cert límit
Versionat	Permet tenir diferents versions d'un fitxer ja publicat	Permet comparar diferents versions d'un fitxer	Permet tenir diferents versions d'un fitxer

Font: Consorci de Serveis Universitaris de Catalunya (CSUC), extret de <http://hdl.handle.net/2072/377586>.

Informació addicional rellevant:

⁴⁶ <http://hdl.handle.net/2072/377586>



Capítol 6

EL COST
DE GESTIONAR
DADES

6

EL COST DE GESTIONAR DADES

A l'hora d'estimar els costos de gestió de les dades de recerca **cal analitzar tots els processos que hi estan relacionats**. Per tant, cal estimar les **hores dedicades pel personal investigador i pel personal expert** encarregat de guiar i supervisar els treballs, però també cal tenir en compte els **costos de les infraestructures i els serveis associats**.

En el procés de planificació cal estimar **les hores dedicades a elaborar el PGD** per part del personal investigador i del personal assessor expert. Els primers plans requereixen més temps, però sempre és útil tenir un protocol que permeti acollir-se a unes pràctiques preestablertes.

El cost estimat del procés de recollida de dades varia en funció de diferents condicionants. Així, de vegades s'ha de considerar el **cost d'accés a bases de dades externes**, mentre que en altres ocasions cal estimar el **cost de generar les dades**, bé amb els mitjans de què disposa el centre o bé mitjançant la contractació de recursos externs o amb fórmules mixtes. En aquest procés també hi pot haver **costos associats a la transcripció, normalització, anonimització o gestió de consentiments informats**. A més, cal avaluar el **cost de les infraestructures relacionades amb la recollida de dades**.

La documentació de les dades requereix personal que descriu els conjunts de dades i afegeixi les metadades necessàries per a la publicació o l'intercanvi posteriors amb

altre personal investigador. **Un cop documentades, les dades s'han d'emmagatzemar en alguna infraestructura el cost de la qual pot variar** si calen determinats nivells de seguretat i preservació.

El procés de publicació de dades també necessita una estimació de costos. Hi pot haver costos associats a la publicació en un dipòsit extern, al manteniment d'un dipòsit propi o consorciat o simplement costos vinculats a l'atorgament d'identificadors permanents als conjunts de dades. També hi pot haver **costos associats a la gestió dels drets**.

A tots aquests costos sempre s'hi han d'afegir, com hem dit, els costos propis del personal especialitzat que ajudi el personal investigador en tot el procés de gestió de les dades.

ACCIONS RECOMANADES

- Identificar tots els processos relacionats amb la gestió de dades del centre.
- Estimar el cost de la gestió de dades del centre.
- Imputar, quan sigui possible, les despeses de gestió de dades als projectes de recerca.

Informació addicional rellevant:

Eina del UK Data Service per estimar el cost de gestió de dades: <https://www.ukdataservice.ac.uk/media/622368/costingtool.pdf>

- Estimació econòmica del cost de gestió de dades realitzada per la Universitat d'Utrecht: <https://www.uu.nl/en/research/research-data-management/guides/costs-of-data-management>
- Estimació econòmica del cost de gestió de dades realitzada per un grup de recerca britànic per justificar els costos d'un projecte europeu.



Capítol 7

UN MODEL DE
POLÍTICA DE GESTIÓ
DE DADES

7

UN MODEL DE POLÍTICA DE GESTIÓ DE DADES

La política de gestió de dades d'un centre és el conjunt de **principis fonamentals i genèrics** que aquest es compromet a seguir en aquest àmbit. Aquesta política es materialitza en **activitats i eleccions** concretes **segons el projecte**, la **disciplina científica** i el **tipus de dades** que s'han de tractar. Ha de tenir per objectiu **millorar la gestió de les dades**, i protegir-ne la preservació perquè estiguin disponibles per reutilitzar-les tant internament com externament.

La política de gestió de dades **ha d'incloure la creació i el manteniment d'un catàleg de dades del centre**, per tal que se sàpiga on es troben les dades fins i tot si no són accessibles públicament. No obstant això, la política no ha d'incloure l'obligatorietat de publicar les dades en obert en determinats repositoris, per diverses raons. Primer, perquè, com ja s'ha vist, no totes les dades es poden o s'han de publicar en obert. Segon, perquè ja hi ha repositoris temàtics consolidats que allotgen dades de determinades disciplines o àrees de coneixement i que poden no atendre aquesta obligatorietat. Com a part de la política s'aconsella incorporar-hi la revisió periòdica pròpia.

Un bon model a seguir per crear una política de gestió de dades és el que suggereix el projecte LEARN,⁴⁷ finançat per la Comissió

Europea. El projecte va posar a disposició del públic diferents materials relacionats amb la gestió de dades:

- Una enquesta per saber si una institució està preparada per a la gestió de les dades.
- Un catàleg de bones pràctiques.
- Un model de política de gestió de dades.
- Una sèrie d'indicadors⁴⁸ per al seguiment de la política de gestió de dades.

El model suggerit segueix l'estructura següent:

- Un preàmbul on s'exposen els motius de la política.
- Establir la jurisdicció que ha de regir la política.
- La normativa de propietat intel·lectual aplicable, així com la titularitat dels drets que se'n derivin.
- La gestió de les dades.
- La responsabilitat, els deures i els drets de les parts implicades.
- La validesa de la política, així com les futures revisions i actualitzacions.

Així mateix, s'han de **crear i utilitzar indicadors de seguiment** per conèixer l'avenç de l'organització cap als objectius que s'han establert en implantar la política i, en cas que no s'assoleixin els objectius plantejats, posar

Informació addicional rellevant:

⁴⁷ <http://learn-rdm.eu/>

⁴⁸ http://learn-rdm.eu/wp-content/uploads/KPIs_MD.pdf

en marxa canvis en la política o en les accions que s'estiguin duent a terme.

El projecte LEARN suggereix una llista d'indicadors que hauria d'ajudar les institucions a mesurar l'èxit en la implementació de la política de gestió de dades. El primer conjunt d'indicadors té com a objecte preparar la institució per a la gestió de dades. Això comporta disposar d'un grup que lideri la gestió de dades al centre, i disposar de serveis, perfils adequats i un programa de formació. El segon conjunt d'indicadors s'ha de fer servir per mesurar la implementació de les polítiques de gestió de dades. Estem parlant, per exemple, de nombre d'investigadors que utilitzen els serveis i la possible infraestructura que pot oferir el centre, el nombre de conjunts de

dades emmagatzemades i publicades o el nombre de sessions de formació realitzades i el nivell de participació.

7.1 Aspectes tècnics

Cada centre ha d'informar el personal investigador i de suport sobre els principis que cal seguir en la presa de decisions sobre aspectes tècnics. Generalment, aquesta informació s'incorpora a un reglament de gestió de dades que desenvolupa la política de gestió de dades de la institució. En cas que no hi hagi una posició definida per part del centre, s'ha de delegar la presa de decisions en el personal qualificat o el personal investigador, sobre els quals ha de recaure la responsabilitat.

Les qüestions tècniques sobre les quals les institucions han d'expressar la seva posició són les següents:

- Mètode de publicació de dades produïdes i llicències que cal escollir.
- Grau de reutilització de dades existents i processos de formalització de peticions per accedir a repositoris de dades restringides.
- Formats preferits i organització interna de les dades en la infraestructura local, si és rellevant.
- Eines disponibles o preferides per emmagatzemar les dades i accedir-hi.
- Política de preservació de les dades a mitjà i llarg termini i responsabilitats sobre les dades.
- Grau d'obligatorietat sobre l'accés, la utilització i la deposició de noves dades que es produeixin en repositoris de dades temàtics, multidisciplinaris i institucionals.
- Incentius o limitacions de la institució per a la publicació de dades en obert o mecanismes per limitar l'accés a les dades produïdes per grups vinculats amb la institució.
- Cost de la gestió de les dades i possibilitat que els grups de recerca o el centre mateix repercuteixin aquest cost en els projectes.

Informació addicional rellevant:

- Política de gestió de dades de recerca de la Universitat de Barcelona: <http://hdl.handle.net/2445/142043>.
- University College London (UCL) Research Data Policy: https://www.ucl.ac.uk/isd/sites/isd/files/ucl_research_data_policy_v6.pdf
- Centre National de la Recherche Scientifique (CNRS) Research Data Plan: https://www.science-ouverte.cnrs.fr/wp-content/uploads/2021/04/Cnrs_Research-Data-Plan_mars21.pdf



Glossari 8

8

GLOSSARI

- **Accessibilitat de les dades:** es refereix al mecanisme pel qual les dades poden ser visualitzades o descarregades. L'accés a les dades pot ser totalment obert o restringit a algun tipus de registre o permís. Per exemple, per compartir dades personals cal tenir en compte el tipus d'autorització concedida pels propietaris de les dades, així com regular-ne i monitoritzar-ne acuradament l'accés, respectant la legislació europea i nacional sobre la protecció de dades personals.
- **Anonimització de dades personals:** procés a través del qual es trenca la connexió entre la dada personal i la persona física a què es refereix, per evitar-ne la identificació. Les dades resultants es denominen *anonimitzades* i no entren dins els casos protegits per l'RGPD o altres normatives d'aplicació. Quan l'element (programari, taula de valors, etc.) que permet la reidentificació de la persona a què es refereixen les dades està simplement aïllat i allunyat de les dades, i no destruït, es parla de *pseudonimització*, i les dades segueixen considerant-se dades personals.
- **Base de dades:** conjunt estructurat de dades i metadades emmagatzemades de manera digital. Permet la consulta de les dades de manera manual, mitjançant una interfície gràfica d'usuari (o GUI, de l'anglès *graphical user interface*), i automatitzada, a través d'interfícies de programació d'aplicacions (o API, de l'anglès *application programming interface*). Habitualment es distingeix entre bases de dades relacionals i no relacionals, segons les característiques de les metadades que desen i les operacions que caldrà fer. Les bases de dades relacionals utilitzen com a llenguatge de programació i comunicació *el llenguatge SQL (structured query language)*, mentre que les bases de dades no relacionals (*NoSQL*) permeten utilitzar un sistema més flexible per desar la informació.
- **Catàleg:** llista d'objectes ordenada segons les seves característiques o metadades. Els objectes poden ser fitxers, conjunts de dades, articles, documents, etc., caracteritzats per metadades o paraules clau que han de ser estàndards en la mesura del possible i seguir el vocabulari propi de la disciplina. Els objectes d'un catàleg es poden cercar tant manualment com amb cercadors automàtics.
- **Cicle de vida de les dades de recerca:** és el conjunt de fases i processos per on passen les dades de cada projecte durant i després de l'activitat de recerca. La creació de noves dades alimenta el cicle, que es desplega amb el processament, l'anàlisi, la curació i la publicació de les dades. La reutilització de dades creades en altres projectes tanca i reactiva el cicle.
- **Citació de les dades:** referència unívoca als conjunts de dades publicades que s'han reutilitzat. La citació de les dades es fa, com a mínim, mitjançant identificadors únics com els PID i els DOI, encara que sol incloure més informació (autor, títol, data, etc.).

- **Confidencialitat de les dades:** limitació d'ús o distribució de les dades definida pel seu propietari i que implica posar en acció mecanismes de seguretat adequats.
- **Dades obertes:** dades que no requereixen autorització d'accés, exploració, anàlisi i reutilització. També es pot considerar que les dades són obertes quan les condicions d'accés són mínimes (per exemple, mitjançant un registre previ o l'acceptació d'uns termes d'ús molt laxos). El tipus de control o autorització necessaris per accedir a les dades determina quan aquestes es poden considerar obertes. En cas contrari, les dades es consideren no obertes o tancades. Aquest concepte no implica necessàriament garanties sobre la qualitat o el valor científic de les dades, ni sobre la seva possible utilitat futura.
- **Dades curades:** són dades acompanyades de metadades i que han passat per un procés de documentació i arxiu. El procés de curació pot seguir els principis FAIR, íntegrament o parcialment.
- **Dades de recerca:** conjunt d'informació, digitalitzada en fitxers o en qualsevol suport físic, que entra en el procés de recerca i contribueix a la producció d'un resultat científic. Al llarg del document també es parla de *conjunt de dades*.
- **Dades dinàmiques:** són les dades de recerca que s'actualitzen periòdicament. És important informar de cada actualització mitjançant la identificació de versions.
- **Dades personals:** segons l'RGPD, són dades personals totes les dades que permeten la identificació d'una persona. Es consideren dades personals: noms, dates, característiques físiques, imatges personals, adreces de correu físiques o electròniques, adreces IP de dispositius, números d'identificació, etc., quan aquests estan associats a una base de dades a través de la qual es poden trobar dades que identifiquen una persona.
- **Dades primàries:** són la base de partida d'una recerca, sense cap modificació (mesuraments, enquestes, observacions, etc.). S'han d'emmagatzemar juntament amb la informació de la font de procedència.
- **Dades secundàries:** són dades que han patit algun tipus de processament (dades intermèdies o resultats).
- **Gestió de dades de recerca:** totes les activitats que tenen com a objectiu la recollida, organització, preservació, accés i publicació de les dades.
- **Gestor de dades (data steward):** perfil especialitzat en la gestió de dades de recerca i en el compliment dels principis FAIR. Les competències específiques són principalment de caràcter científic, tècnic i administratiu. Segons el nivell d'especialització del personal de l'organització, es poden crear equips especialitzats en la gestió de dades de recerca.

- **Identificador permanent (*PID, persistent identifier*):** referència que identifica unívocament un objecte digital (com articles, llibres, fitxers, conjunts de dades, etc.) i que alhora permet arribar-hi a través d'un *URL (uniform resource locator)*, independent de la ubicació física, gràcies a una organització que el manté funcional. Són exemples d'identificadors permanents el DOI i l'identificador del sistema Handle, utilitzat generalment en repositoris per a publicacions.

- **Indexar:** procés que afegeix un camp amb un índex (un número) en una taula perquè les cerques siguin més ràpides. S'utilitza especialment per a bases de dades que contenen un gran nombre d'entrades.

- **Interoperabilitat:** característica de les dades que es poden llegir i utilitzar amb eines diferents, per grups diferents o per disciplines científiques diferents de les que les van generar. La interoperabilitat s'aconsegueix mitjançant la utilització d'estàndards àmpliament acceptats i utilitzats o degudament descrits.

- **Llicència d'utilització:** és una eina legal que defineix els drets de les persones a utilitzar un determinat producte subjecte a propietat intel·lectual (dades, programari, articles, llibres, etc.). El titular dels drets d'explotació decideix sobre la llicència que vol assignar als seus productes, segons la normativa vigent i els acords que ha adoptat (per exemple, amb la revista on publica, amb la institució per a la qual treballa, etc.).

- **Metadades:** dades que descriuen o caracteritzen les dades de recerca a què es refereixen. Les metadades, degudament organitzades i indexades, contribueixen a una bona gestió de les dades, perquè aquestes siguin localitzables i reutilitzables.

- **Organització de dades:** manera com s'emmagatzemen les dades en l'espai de treball, de manera estructurada i consistent.

- **Pla de gestió de dades (*PGD o DMP, per les sigles en anglès de data management plan*):** document viu, actualitzable i versionable, que recull la descripció de les dades i les eines i les decisions de gestió triades abans i durant l'execució d'un projecte de recerca. El responsable del PGD d'un projecte pot ser l'investigador principal o qualsevol altre responsable de reunir la informació necessària per editar-lo i revisar-lo.

- **Política de gestió de dades:** conjunt de principis fonamentals i genèrics que una organització es compromet a seguir pel que fa a la gestió de les dades. Les polítiques de dades es materialitzen en activitats i eleccions concretes segons el projecte, la disciplina científica i el tipus de dades que cal tractar.

- **Principis FAIR (*Findable, Accessible, Interoperable, Reusable*):** conjunt de principis que cal seguir perquè les dades siguin accessibles, es puguin entendre i es puguin reutilitzar, també en estudis interdisciplinaris. Aquests principis es poden estendre al programari i, en general, a qualsevol altre producte de recerca. Els conceptes de dades FAIR i dades obertes no són excloents ni equivalents, sinó complementaris. Unes dades FAIR podrien no ser obertes, si es necessita algun tipus d'autorització per accedir-hi i reutilitzar-les.

- **Propietat intel·lectual:** es refereix als drets sobre qualsevol producció (obres artístiques i científiques o bases de dades, per exemple), que corresponen als autors, fruit de la seva intel·ligència i creativitat, i a altres titulars. El terme *propietat intel·lectual* sovint s'estén al concepte de propietat industrial, que inclou els models d'utilitat, les patents i les marques. A Espanya aquests dos conceptes estan regulats per lleis diferents que atorguen condicions i durades diferents als dretahavents.
- **Protecció de dades personals:** la reglamentació general de protecció de dades europea (Reglament general de protecció de dades, o RGPD) conté la normativa a què s'ha d'atendir cada legislació específica nacional sobre l'ús i el tractament de les dades personals.
- **Repositori de dades:** es caracteritza per la possibilitat d'emmagatzemar dades de manera segura i exposar metadades de manera que les dades es puguin cercar i trobar. Els dipòsits de dades es poden classificar genèricament com a: institucionals, si recullen les dades d'una mateixa institució (per exemple, d'una universitat o d'un centre de recerca); consorciats, si recullen les dades d'institucions que pertanyen a un consorci; temàtics, si recullen dades d'una mateixa disciplina o branca de recerca. També hi ha repositoris genèrics o multidisciplinaris per a dades de disciplines que no tenen un repositori temàtic específic.

- **Reutilització de dades:** operació que consisteix a utilitzar dades generades i publicades per altres processos (de recerca, de mesurament, etc.) per incorporar-les a un projecte i, a partir d'aquestes dades, generar nous resultats científics. La reutilització de dades evita la duplicat d'esforços en la generació de dades ja existents. Això és possible si les dades estan curades d'acord amb els principis FAIR.
- **Seguretat de les dades:** mecanismes tècnics, administratius i físics que assegurin el compliment de les limitacions d'accés, ús i distribució de les dades, en compliment de la legislació i dels acords de confidencialitat vigents i aplicables.
- **Vocabulari:** conjunt de termes que s'utilitzen per a la descripció d'objectes en determinats contextos. L'elecció d'un vocabulari adequat és indispensable per evitar que les dades siguin mal interpretades o que s'utilitzin de manera incorrecta. Hi ha vocabularis que reuneixen els termes i les definicions estàndard per a molts contextos.

