

Grup Operatiu de Ciència Oberta

Gestió de les dades de recerca

Natàlia Plancheria Roca
22 de juny de 2023

Índex

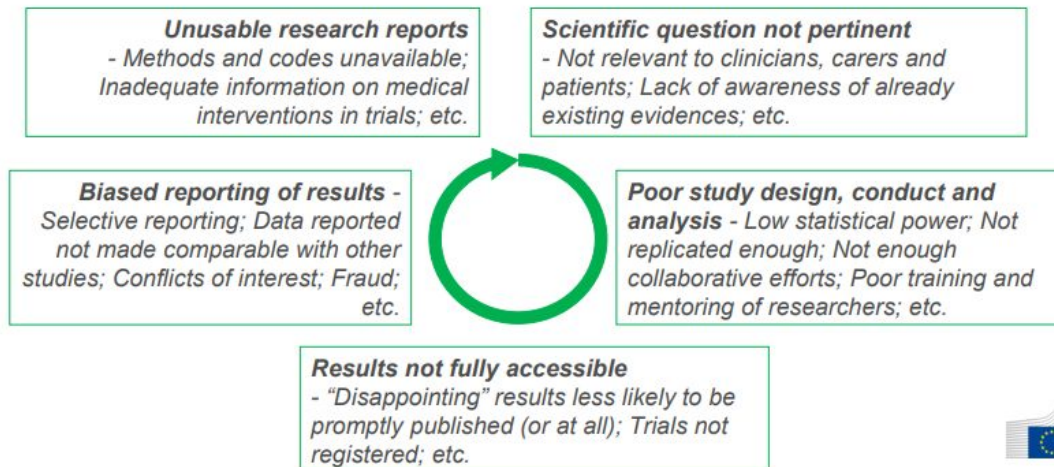
1. **Ciència Oberta**
2. **Dades de recerca**
 - a. Què són dades de recerca
 - b. Dades FAIR
3. **Pla de Gestió de Dades**
 - a. eiNaDMP
4. **Publicar dades de recerca**
 - a. repositoris de dades
 - b. CORA.RDR
5. **Serveis i recursos a la UOC**

Ciència Oberta

Crisi a la ciència?

Reproducibility: The 'crisis' (zoom in *health R&I*)

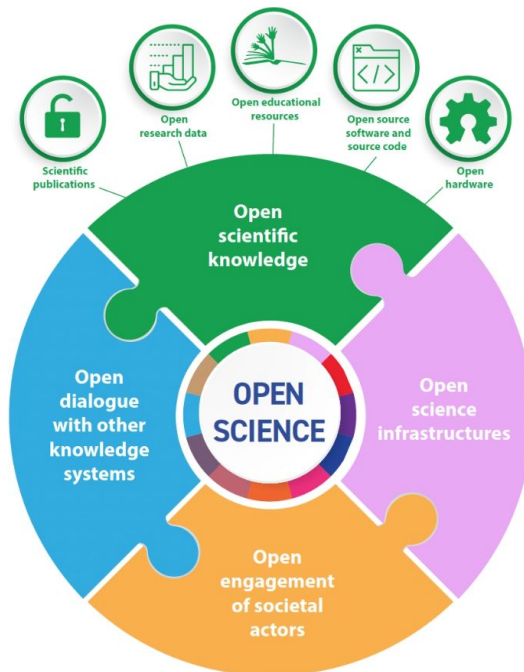
- Close to €300 billion/year for Health R&I (worldwide)
- A large share of the research investment may be wasted: potentially as much as 85%, according to Chalmers & Glasziou 2009, Lancet; Macleod 2014, Lancet



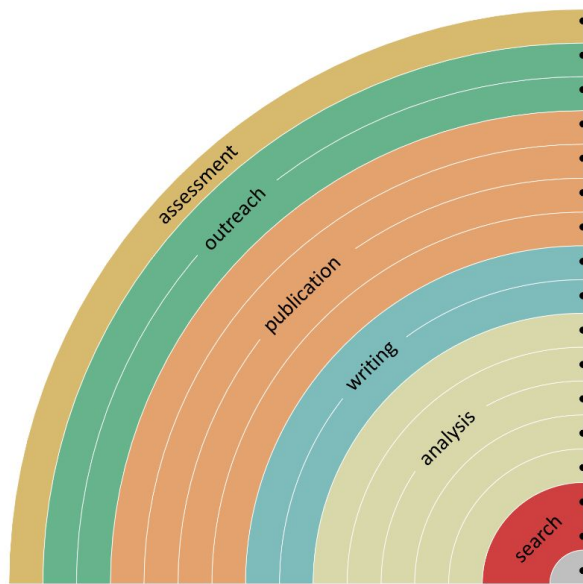
Què és la ciència oberta?

- Open Science is the practice of science in such a way that others can **collaborate** and **contribute**, where research data, lab notes and other research **processes** are **freely available**, under terms that enable **reuse**, **redistribution** and **reproduction** of the research and its **underlying data** and **methods**. ([FOSTER](#))
- Open science is a set of principles and **practices** that aim to make scientific research from all fields **accessible** to everyone for the benefits of scientists and society as a whole. ([UNESCO Recommendation on Open Science](#))
- Nou enfocament del **procés de recerca** basat en el treball **cooperatiu** i les noves formes de **difondre** el coneixement utilitzant les tecnologies digitals i les noves eines de **col·laboració**. Els resultats de la recerca han de ser **accessibles** a la societat de manera **lliure** i **gratuïta**, per tal d'aconseguir estendre els principis d'obertura a tot el cicle de vida de recerca i fomentar que es contribueixi a la difusió i **reutilització** d'aquests. ([CSUC](#))

Què inclou la ciència oberta?



You can make your workflow more open by ...



- adding alternative evaluation, e.g. with altmetrics
- communicating through social media, e.g. Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. CC0 or CC-BY
- publishing open access, 'green' or 'gold'
- using open peer review, e.g. at journals or PubPeer
- sharing preprints, e.g. at OSF, arXiv or bioRxiv
- using actionable formats, e.g. with Jupyter or CoCalc
- open XML-drafting, e.g. at Overleaf or Authorea
- sharing protocols & workfl., e.g. at Protocols.io
- sharing notebooks, e.g. at OpenNotebookScience
- sharing code, e.g. at GitHub with GNU/MIT license
- sharing data, e.g. at Dryad, Zenodo or Dataverse
- pre-registering, e.g. at OSF or AsPredicted
- commenting openly, e.g. with Hypothes.is
- using shared reference libraries, e.g. with Zotero
- sharing (grant) proposals, e.g. at RIO



Marc legal

UOC:

- [Pla d'acció Coneixement Obert](#)
- [Política institucional de coneixement obert](#)

Autonòmic:

- [Llei 9/2022, del 21 de desembre, de la ciència](#)
- [Pacte Nacional per la Societat del Coneixement](#)

Estatat:

- [Llei 17/2022 de la Ciència, la Tecnologia i la Innovació](#) (Article 37. Ciència oberta)
- [Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario](#)
- [Estrategia Nacional de Ciencia Abierta \(ENCA\)](#)

6 eixos principals



Publicacions en obert

Dades FAIR

Aprenentatge en obert

Innovació oberta

Oberta a la societat

Models d'avaluació de la recerca

Dades de recerca

Què són dades de recerca?

- Una **dada** és qualsevol informació recollida, observada, generada o creada per validar un procés de recerca.
- Les **dades en obert** són dades accessibles en línia, sense cost d'ús, que es poden utilitzar, reutilitzar i distribuir sempre que se'n citi la font de dades.
- Un **dataset** és un conjunt de fitxers (que inclou dades, codi i documentació relacionada) i les seves metadades descriptives associades.

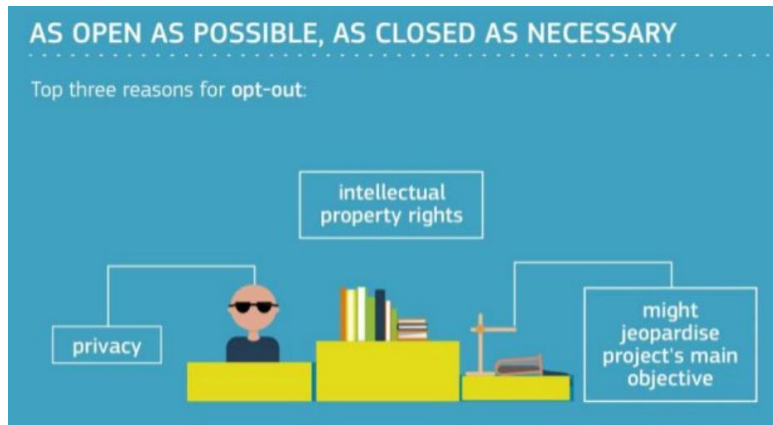
Font: [FOSTER](#)



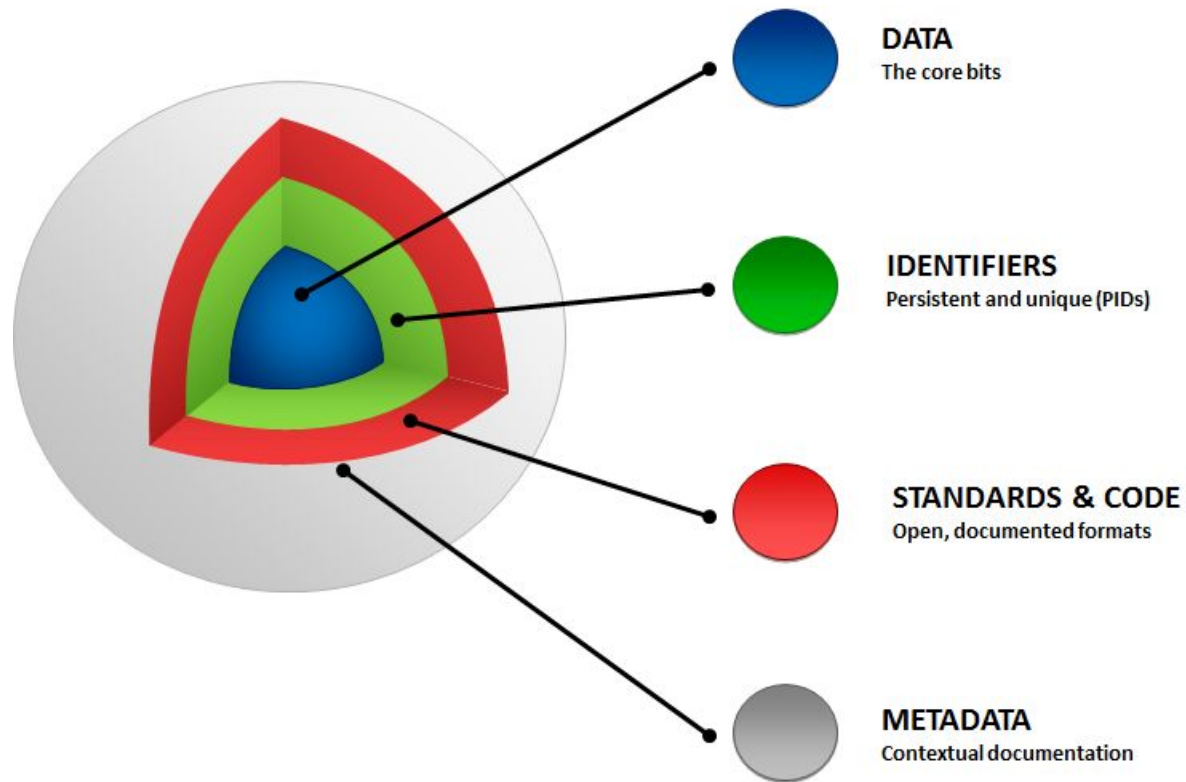
- **Tipologies** diverses: observacionals, experimentals, compilacions, simulacions, ...
- **Formats** múltiples: textual, fulls de càlcul, imatges, àudios, vídeos, bases de dades, ...
- **Mida**: a partir d'1 MB

En què beneficia publicar les dades de recerca?

- Per complir amb els **requisits** de les entitats finançadores.
- Permet **validar** els resultats publicats als articles científics.
- Fomenta la **col·laboració** entre personal investigador i projectes.
- Millora la **transparència** del procés científic.
- Incrementa la **visibilitat** del personal investigador.
- Augment de les **citacions**.
- Millora l'**impacte** de la recerca.



Les dades FAIR



Les dades FAIR

To be Findable:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.

F2. data are described with rich metadata.

F3. (meta)data are registered or indexed in a searchable resource.

F4. metadata specify the data identifier.

Com fer-les trobables:

- Ús d'**identificadors persistents** (DOI) del repositori per a publicar les dades
- Les **metadades** utilitzades per a descriure les dades han de permetre que les dades es trobin de manera fàcil, en línia.
- Comprovar que el repositori on publiquem segueixi esquemes estàndards de metadades.

Les dades FAIR

To be Accessible:

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.

A1.1 the protocol is open, free, and universally implementable.

A1.2 the protocol allows for an authentication and authorization procedure, where necessary.

A2 metadata are accessible, even when the data are no longer available.

Com fer-les accessibles:

- Recuperable (dades i metadades) mitjançant protocols estàndards
- Revisar si les dades poden ser publicables i obertes
- Protegir les **dades sensibles** (anonimització)
- Escollir un repositori de dades que permeti restriccions d'accés a les dades (en cas de ser necessari)

Les dades FAIR

To be Interoperable:

11. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
12. (meta)data use vocabularies that follow FAIR principles.
13. (meta)data include qualified references to other (meta)data.

Com fer-les interoperables :

- Ús de **vocabularis controlats**, tesaurus o ontologies si és possible per a descriure les dades.
- Utilitzar **formats** de fitxers oberts. [Bones pràctiques en gestió de dades de recerca: recomanacions de format](#) (CSUC)
- Tenir cura de l'organització dels **fitxers**. [Bones pràctiques en gestió de dades de recerca: organitzar, anomenar i versionar fitxers](#) (CSUC)

Les dades FAIR

To be Re-usable:

R1. (meta)data have a plurality of accurate and relevant attributes.

R1.1. (meta)data are released with a clear and accessible data usage license.

R1.2. (meta)data are associated with their provenance.

R1.3. (meta)data meet domain-relevant community standards.

Com fer-les reutilitzables:

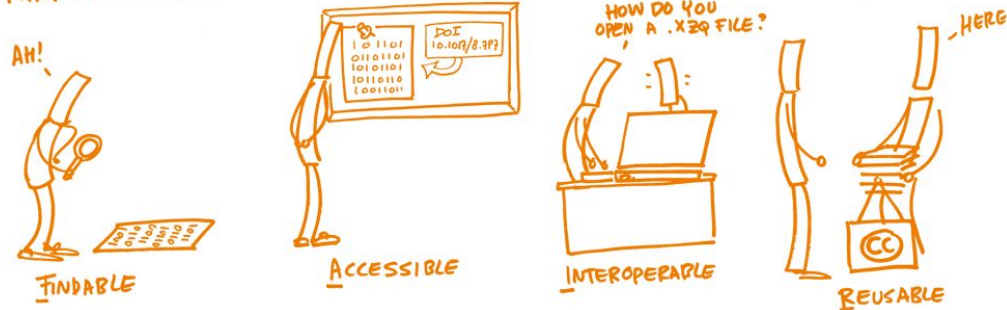
- Escollir **llicència** que en permeti la reutilització. [Bones pràctiques en gestió de dades de recerca: drets i llicències de les dades de recerca](#) (CSUC)
- Descriure les dades de recerca utilitzant un fitxer **readme.txt**. [Plantilles](#)

Com sabem si les dades que tenim son FAIR?

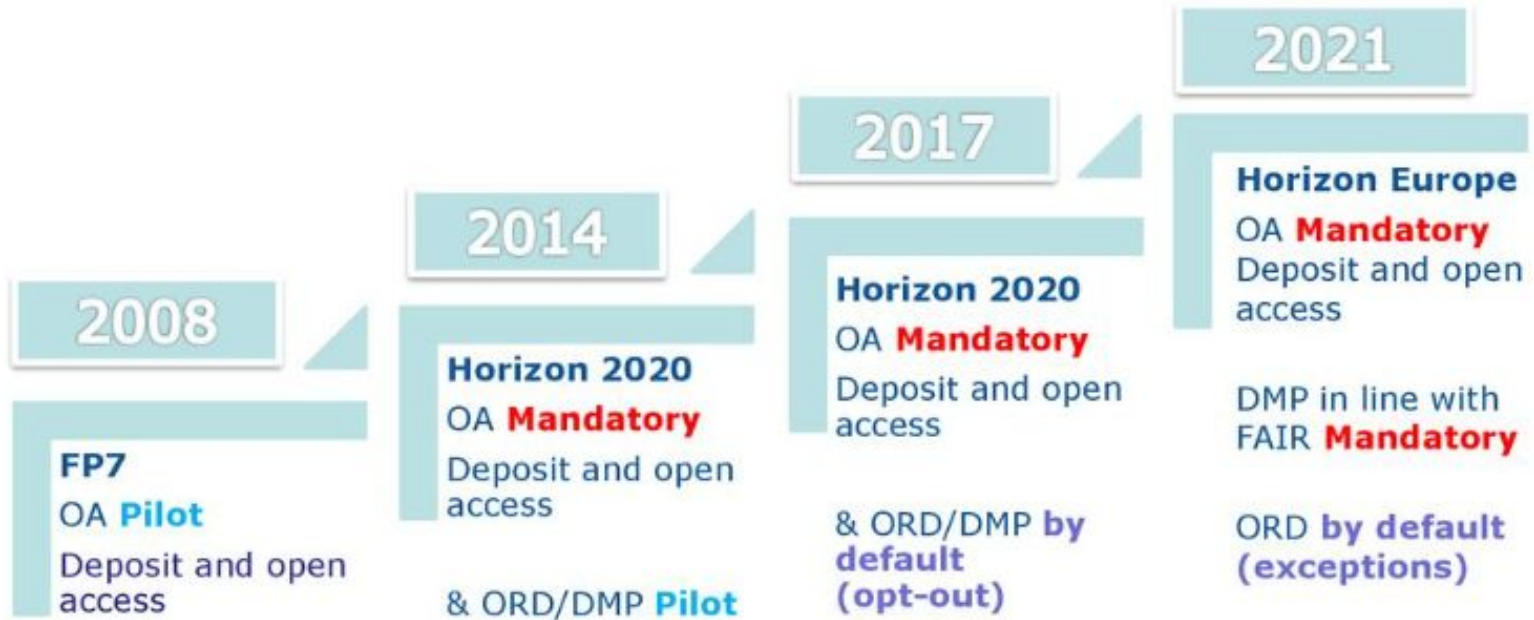
Existeixen diferents llistes de comprovació per assegurar que les dades que es generen compleixen els requisits FAIR:

- [FAIR data self-assessment tool](#)
- [FAIR Aware](#)
- [ARDC FAIR Data self-assessment tool](#)
- [EUDAT Fair Data Checklist](#)

FAIR DATA PRINCIPLES



Què demanen els finançadors: Comissió Europea



Què demanen els finançadors: Plan Nacional



Plan Nacional

INSTRUCCIONES PARA RELLENAR LA MEMORIA CIENTÍFICO TÉCNICA
PROYECTOS INDIVIDUALES



Convocatoria 2020 «Proyectos de I+D+i» – Proyectos individuales

2. IMPACTO ESPERADO DE LOS RESULTADOS

- Impacto científico-técnico
 - a) El plan de difusión
 - b) Conocimientos/resultados potencialmente transferibles
 - c) Previsión del plan de gestión de datos de investigación en el que se indique qué datos se van a recoger o generar (tipologías y formatos), cómo será el acceso (quién, cómo y cuándo se podrá acceder a los datos) y en qué repositorio está previsto que se depositen. En el caso de datos que estén sometidos a la reglamentación de protección de los datos personales o de aspectos éticos, indicar cómo se gestionarán. En el caso de los proyectos que resulten financiados, durante la ejecución del proyecto y junto al informe final se podrá solicitar un plan de gestión de datos formal completo.

Pla de Gestió de Dades

Què és un Pla de Gestió de Dades?

- Pla de Gestió de Dades = Data Management Plan
- Requeriment dels Projectes H2020, Horizon Europe i Plan Estatal
- Document viu i formal on es detalla com es gestionaran les dades durant el projecte de recerca i una vegada aquest estigui finalitzat.
- Us ajudarà a:
 - estalviar temps
 - assegurar la reproducció de les dades
 - preservar i difondre de les dades
 - facilitar la comunicació entre els socis i la comprensió de les dades del projecte
 - evitar duplicitat o pèrdua de dades
 - definir les responsabilitats



10 passos per elaborar un Pla de Gestió de Dades

Un **Pla de Gestió de Dades** (PGD) o Data Management Plan (DMP) és un **document formal**, **que es presenta a l'inici de la investigació**, en el qual es

descriu què es farà amb les dades al llarg i després de finalitzar la investigació i que es pot modificar si es produeixen canvis en la mateixa.

Per què?

És una **bona pràctica**, és un **element clau de l'Open Science** i és **obligatori** als nous projectes H2020.

Eines gratuïtes per elaborar un PGD



PGDonline
(Consorcio Madroño)
<http://dmp.consorcioTRODRO.es/>



DMPonline (Digital Curation
Centre, UK)
<https://dmponline.dcc.ac.uk/>



01

Revisu els **requeriments** de l'ens finançador (H2020).



02

Identifiqueu les dades: tipologia, procedència, volum, formats i fitxers.



03

Definiu com s'organitzaran i gestionaran les dades: nom dels fitxers, control de versions, programari necessari...



04

Expliqueu com es documentaran les dades: identifiqueu la informació a processar, consulteu si hi ha estàndards o esquemes de metadades, identifiqueu eines per a gestionar-los.



05

Descriviu els processos que asseguruen una **bona qualitat de les dades**.



06

Prepareu una estratègia d'emmagatzemament (durant el procés) i de preservació de dades (repositori).



07

Definiu les polítiques de dades del projecte: qüestions sobre propietat intel·lectual i com es tractaran les dades sensibles i personals.



08

Descriviu com es difondran les dades: on, quines, quan es difondran. Si publicareu les dades en un repositori, com a informació suplementària de l'article o com un "data paper".



09

Assigneu rols i responsabilitats a les persones i organitzacions participants en el projecte.



10

Prepareu un pressupost realista: la gestió de dades costa temps i diners en termes de programari, maquinari, serveis i personal



CORA - eiNa DMP

- Eina en línia gratuïta, senzilla i col·laborativa
- Conté plantilles per Horizon Europe, Horizon 2020, Plan Estatal, Doctorands, Software, etc.
- Permet col·laborar amb d'altres investigadors atorgant-los permisos de lectura, edició o co-autoria
- Exportació: converteix el teu document a DOC, PDF, XML, etc.

SIGN IN
CREATE ACCOUNT

* Email

Remember email

* Password

[Forgot password?](#)

This tool helps you to create, review, and share data management plans that meet institutional and funder requirements. Join the growing number of researchers that have adopted eiNa DMP:



1331

USERS



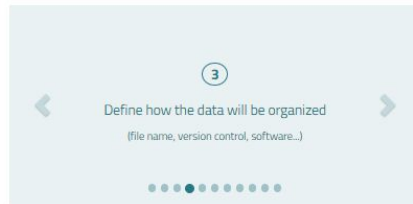
560

PLANS



33

ORGANISATIONS



You can download funder templates without logging in, but from this tool you will find tailored guidance from these organizations. Why not sign up for an account and try it out?



Què inclou un Pla de gestió de dades?

Project Details	Plan overview	Write Plan	Share	Download
expand all collapse all				
1. Data summary (0 / 6)				
2. FAIR data (0 / 4)				
3. Other research outputs (0 / 2)				
4. Allocation of resources (0 / 4)				
5. Data security (0 / 2)				
6. Ethics (0 / 1)				
7. Other issues (0 / 1)				
8. Further support in developing your DMP (0 / 1)				

Guidance

Comments

Horizon Europe

expand all | collapse all

2.1.A Will data be identified by a persistent identifier? +

Explain how the data and metadata are assigned to a globally unique and eternally persistent identifier (DOI, Handle...).

Real example

1. The institutional repository provides a unique URL to access the document with the format `https://repository/record/1234`.
2. The repository assigns Handle/DOIs for persistent identification and citability of the dataset.

2.1.B Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? +

2.1.C Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use? +

2.1.D Will metadata be offered in such a way that it can be harvested and indexed? +

Publicar dades de recerca

On publicar les dades de recerca?

- Disposeu d'algun repositori temàtic consolidat en el vostre àmbit?
 - [Re3data.org](https://re3data.org)
 - [FAIRsharing.org](https://fairsharing.org)
- [CORA.RDR](https://cora.rdr.org)
- La revista us recomana algun repositori?
- Altres repositoris generalistes: Zenodo, Dataverse, Mendeley Data, ...

CORA.RDR

- Complexos requeriments agències de finançament
- Publicació en mode FAIR
- Segueix directrius de l'[EOSC](#)
- Característiques:
 - Repositori federat per a dades de recerca (datasets)
 - Multidisciplinar
 - Obert
 - Per a PDI i doctorands/des
 - 10GB per fitxer i 100GB màxim per dataset
 - Sense límit de datasets

The screenshot shows the CORA.RDR website interface. At the top, there is a navigation bar with links for Home, About, Search, User Guide, Metrics, Contact, English, and Log In. Below the navigation bar is a search bar with the text "Search this dataverse..." and an "Advanced Search" button. The main content area displays statistics: 33 UNIVERSITIES & CERCA CENTERS, 496 DATASETS, 13,550 FILES, 52,823 DOWNLOADS, and 1 TB STORAGE. Below the statistics is a carousel of logos for various institutions: IBEC, ICRAC, Universitat de les Illes Balears, IFAE, and UOC. The search results section shows "1 to 10 of 624 Results" and a "Sort" button. The results list includes:

- Dades comparatives de formació presencial i virtual asincrònica de la SGPC** (Jun 12, 2023 - Universitat Rovira i Virgili)
- Replication Data for: Sign language translation for instructional videos** (Jun 8, 2023 - IDEAI-UPC - Intelligent Data Science and Artificial Intelligence Research Center)
- Biological effects of cytotoxic inflammatory proteins formulated as self-assembling, protein-only nanoparticles** (Jun 8, 2023 - IBB - Institut de Biotecnologia i de Biomedicina)

Serveis i recursos a la UOC

Serveis

Publicació de les dades de recerca a [CORA.RDR](#)

Suport en la creació de Pla de Gestió de Dades. [eiNa DMP](#)

[Ètica i protecció de les dades](#)

Suport amb l'[elecció de les llicències](#)

Tecnologia per a la seguretat i anonimització de les dades. [Data Science Lab](#)

Catàleg de serveis: [Intrauoc > Recerca i Innovació > Catàleg de serveis > Gestió de dades de recerca](#)

Envieu-nos les consultes a ciencia_oberta@uoc.edu

Recursos

[Gestiona les dades de la teva recerca](#)

[Eines per a la gestió de dades de recerca](#) (biblioguia)

[Gestió de dades de recerca](#) (CSUC)

Envieu-nos les consultes a ciencia_oberta@uoc.edu

ciencia_oberta@uoc.edu