

Recomanacions per seleccionar un repositori per al dipòsit de dades de recerca

(Versió 6, Gener 2025)

(Doc. CO25/02) (Ciència Oberta\RDM\RDR\0Fases anteriors\F1 Recomanacions RepDades\2501_RecomanacionsSeleccionarRepositoriDades.docx, 16.01.25)

Aquest document té per objectiu aconsellar els investigadors en el moment de seleccionar un repositori per al dipòsit de les seves dades de recerca que els hi permeti seguir els principis FAIR (Findable, Accessible, Interoperable i Reusable) atès que els projectes finançats per diverses agències, com ara la Comissió Europea o la Agencia Estatal de Investigación (entre d'altres), requereixen que es dipositin en obert les dades de recerca.

Aquest document ha estat elaborat pel Grup de Treball de Suport a la Recerca del CSUC que està format per representants de les universitats següents: Universitat de Barcelona, Universitat Autònoma de Barcelona, Universitat Politècnica de Catalunya, Universitat Pompeu Fabra, Universitat de Girona, Universitat de Lleida, Universitat Rovira i Virgili, Universitat Oberta de Catalunya, Universitat de Vic-Universitat Central de Catalunya, Universitat Ramon Llull i Universitat Illes Balears.

Aquest document està subjecte a la llicència de Reconeixement de Creative Commons (<http://creativecommons.org/licenses/by/4.0/>).

Versió digital: <http://hdl.handle.net/2072/480150>

Sumari

1. SELECCIÓ D'UN REPOSITORI DE DADES DE RECERCA	3
1.1 PROCÉS PER SELECCIONAR UN REPOSITORI DE DADES	4
1.2 CRITERIS A TENIR EN COMPTE A L'HORA D'ESCOLLIR UN REPOSITORI DE DADES	5
2. FONTS PER SELECCIONAR UN REPOSITORI DE DADES TEMÀTIC	6
2.1 DIRECTORIS	6
2.2 RECOMANACIONS D'EDITORIALS	6
3. FONTS PER SELECCIONAR UN REPOSITORI DE DADES MULTIDISCIPLINARI	7
REFERÈNCIES.....	11

1. Selecció d'un repositori de dades de recerca

Avui en dia, la decisió d'escollir un repositori de dades de recerca pot arribar a ser tot un repte. Algunes disciplines acadèmiques han establert i consolidat certs repositoris de dades de recerca. En canvi, d'altres àmbits d'investigació no tenen aquestes possibilitats.

La selecció d'un o altre repositori comporta certa planificació a l'hora de generar les dades, ja que cal complir amb els requisits que aquests demanen (estàndards de metadades, documentació, drets d'autor, preservació...).

Els repositoris es poden classificar en:

1. **Repositoris institucionals** que poden ser temàtics o multidisciplinaris. Si no es disposa d'un repositori temàtic de l'àmbit és possible que la teva institució et recomani fer el dipòsit en el seu repositori de dades institucional, com el [CORA.Repositori de Dades de Recerca](#). Per a conèixer les funcionalitats del repositori, podeu consultar l'apartat "[Sobre](#)".
2. **Repositoris externs** que poden ser **temàtics o específics per una disciplina** i **multidisciplinaris o generalistes**. Els repositoris temàtics ofereixen el coneixement i l'experiència en la gestió de les dades en el seu context, esquemes de metadades més específics i permeten que altres investigadors de l'àmbit puguin trobar les dades de recerca amb més facilitat. En canvi, els repositoris multidisciplinaris permeten el dipòsit de dades de qualsevol àrea temàtica que no disposen d'un repositori disciplinari de confiança.



Figura 1. Tipologies de repositoris per a dades de recerca

1.1 Procés per seleccionar un repositori de dades

Per seleccionar un repositori de dades cal seguir els següents passos (figura 2):

1. Comprovar si es disposa d'una [política o recomanació sobre gestió de dades de recerca a la teva institució](#) i seguir les indicacions establertes. En el cas de les universitats i centres de recerca CERCA, el repositori institucional és l'RDRC. Donat que sigui un projecte finançat, revisar la convocatòria de l'agència de finançament.
2. Consultar si es disposa d'un repositori temàtic consolidat de l'àmbit [vegeu secció 2].
3. En el cas que no hi hagi cap repositori temàtic consolidat, ni tampoc es pugui dipositar al repositori institucional, utilitzar un repositori de dades multidisciplinari [vegeu secció 3].

En el cas que les dades es dipositin en un repositori temàtic o multidisciplinari, informar a la institució (unitats o serveis de suport a la recerca) sobre les dades dipositades i on s'han dipositat.

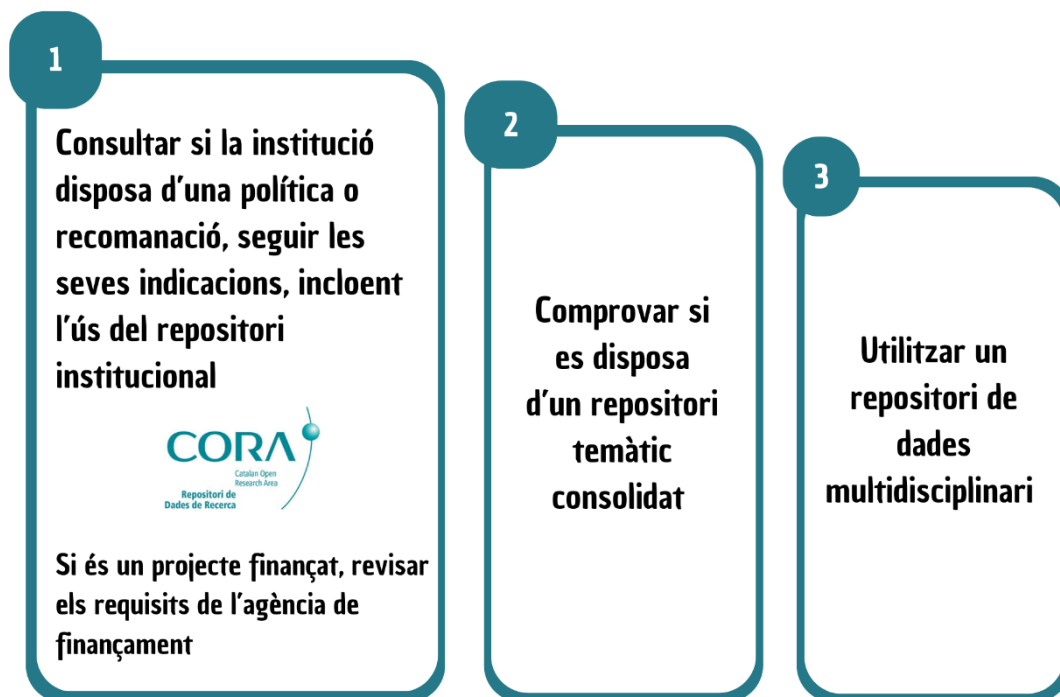


Figura 2. Procés per seleccionar un repositori de dades

1.2 Criteris a tenir en compte a l'hora d'escollir un repositori de dades

A continuació, es mostren una sèrie de consideracions basades en la proposta de la John Hopkins University, el Digital Curation Centre i la UC Santa Barbara Library (vegeu bibliografia) que cal tenir en compte a l'hora de seleccionar un repositori de dades (preferiblement durant l'elaboració del pla de gestió de dades).

1. Existeix algun repositori representatiu i consolidat en el teu àmbit?
2. És de confiança i disposa o està en procés d'assolir alguna certificació, com per exemple el CoreTrustSeal?
3. El repositori permet dipositar les dades en el format utilitzat?
4. La mida dels fitxers que pots dipositar s'ajusta a les teves necessitats?
5. Es permet la restricció d'accés a les dades (tancades, restringides o embargades)?
6. Els drets i llicències que t'ofereix el repositori s'ajusten amb els teus interessos?
7. Els termes i condicions d'ús del repositori estan explicats clarament? (o disposa d'un reglament públic)
8. Disposa de guies per publicar datasets, informa de característiques tècniques, etc.?
9. El repositori atorga un identificador permanent?
10. Les condicions per dipositar o retirar contingut del repositori s'ajusten a les teves necessitats?
11. Hi ha costos associats a l'ús del repositori?
12. Es poden enllaçar les dades a les publicacions associades?
13. El repositori té una política de preservació (còpies, terminis de conservació) per tal de garantir la integritat, l'accessibilitat i la usabilitat de les dades a llarg termini?
14. Permet dipositar diferents versions d'un mateix arxiu?
15. El repositori ofereix estadístiques d'accés i descàrrega sobre els conjunt de dades per poder fer un seguiment de com s'estan utilitzant les dades?
16. S'ofereix servei de curació de dades?
17. Compleix els requisits de la teva agència de finançament, en cas de tenir-ne?
18. Permet la integració amb altres infraestructures (p.e. Zenodo i Github)?
19. Proporciona visibilitat als datasets?

2. Fonts per seleccionar un repositori de dades temàtic

2.1 Directoris

Els següents directoris de repositoris de dades de recerca cobreixen diferents àrees de coneixement. Es presenten els resultats ordenats alfabèticament per disciplina.

- [Re3data](#)
- [FAIRsharing.org](#)

2.2 Recomanacions d'editorials

Sovint, algunes editorials suggereixen, sense obligar-hi, l'ús de repositoris temàtics per dipositar les dades suplementàries a les publicacions. Vegeu com a exemple:

- [Elsevier](#)
- [Nature](#)
- [Plos](#)
- [Springer](#)

3. Fonts per seleccionar un repositori de dades multidisciplinari

A continuació es mostra una taula comparativa dels diferents repositoris multidisciplinaris que actualitza les dades del document CO17/02.

L'actualització s'ha fet en base a una taula comparativa de repositoris de dades elaborada a partir de la NIH Workshop on the Role of Generalist Repositories to Enhance Data Discoverability and Reuse del 2020 traduïda i adaptada pel Consorci de Serveis Universitaris de Catalunya.

	 CERN	 EUDAT	 Digital Science (Macmillan)	 NCSU	 IQSS (Harvard University)	 Elsevier
1. Àmbit temàtic i geogràfic	Multidisciplinar Internacional	Multidisciplinar Internacional	Multidisciplinar Internacional	Amb èmfasi en ciència i medicina Internacional	Multidisciplinar Internacional	Multidisciplinar Internacional
2. Tipologia de dades (programari, imatges, raw data, etc.)	Tot tipus de formats i també documents publicats	Tot tipus de formats	Tot tipus de formats	Tot tipus de formats (amb exclusió de software) i amb preferència per formats de codi obert	Tot tipus de formats amb funcionalitats addicionals per alguns (RData, SPSS, STATA, CSV, xlsx, FITS)	Tot tipus de formats
3. Mida aproximada dels fitxers a incloure	Fins a 50 GB Per a mides superiors es pot contactar amb ells	Fins a 10 GB. Per a mides superiors, contactar amb ells	Fins a 20 GB	10 GB per un conjunt de dades. Per a mides superiors revisar les condicions	Fins a 100 GB	10 GB
4. Dades obertes, embargades, restringides o tancades	Obertes Tancades Restringides Embargades (Policies)	Obertes Restringides Embargades	Obertes Tancades (per defecte fins que no es publiquen) Restringides Embargades	Obertes Embargades (per defecte fins el moment de la publicació)	Obertes Restringides	Obertes Tancades (per defecte fins que no es publiquen) Restringides Embargades
5. Llicències amb les que es volen difondre les dades	Permet varietat de llicències encara que recomanen llicències obertes. La descripció de les dades està sota CC0	Permet varietat de llicències per a dades i per a software	- CC0 per a dades (<i>metadata & datasets</i>) - CC-BY per a no dades (audiovisuals, <i>papers</i> , etc.) - MIT per a codi i software - Altres llicències : GNU, Apache 2.0, BSD 3 clause	CC0 (FAQ)	Per defecte CC0 però es pot canviar	Permet varietat de llicències: CC-BY 4.0, CC-BY-NC, CC0 1.0, MIT, Apache 2.0, BSD 3 clause, BSD 2 clause, gplv3
6. Identificador permanent	DOI o URL	Handle	DOI	DOI	DOI	DOI

7. Condicions per retirar contingut	Es permet retirar les dades justificant-ne el perquè. El DOI i la URL quedaran retinguts.	Per retirar contingut contactar amb EUDAT	Es pot sol·licitar la retirada de contingut públic justificant-ne el motiu	Cessió de dades irrevocable	Es permet retirar les dades, tot i que les metadades quedaran sempre visibles	Per retirar el contingut contactar amb Mendeley .
8. Cost	Gratuït. Però hi ha la possibilitat de més espai (FAQS)	Gratuït	Gratuït fins a una certa mida: -20 GB pels arxius individuals -20 GB d'espai privat -100 espais col·laboratius Per més espai i altres prestacions, ofereixen versió de pagament. Contacte	150 \$ per un conjunt de dades de fins a 50GB. Cost addicional per sobre de 20 GB (Payment plans)	Gratuït fins a 1TB per investigador	Gratuït
9. Dades enllaçades a les publicacions	Permet afegir enllaços a documents relacionats	Permet afegir enllaços a documents relacionats	Permet afegir enllaços al text complet o a documents relacionats	-Les dades han d'estar associades amb un document de recerca publicat o en procés de publicació - Integració amb les publicacions (Submission integration)	Permet afegir enllaços a documents relacionats	Permet afegir enllaços a documents i programaris relacionats
10. Preservació	Dades emmagatzemades durant mínim 20 anys al CERN Data Center	Com a mínim 2 anys i amb garanties de migració a altres repositoris (el centre de dades d'emmagatzematge està certificat per la ISO)	Preservació (Preservation policies)	Dades emmagatzemades a DataONE	Harvard Dataverse Preservation Policy . Conveni amb Data Preservation Alliance per preservar les dades a llarg termini	Dades emmagatzemades a Amazon S3 , a Alemanya. Conveni amb DANS per preservar les dades a llarg termini
11. Versions dels documents	Permet tenir diferents versions d'un arxiu	Sense informació	Permet tenir diferents versions d'un arxiu ja públic	Permet tenir diferents versions d'un arxiu	Permet tenir diferents versions d'un arxiu	Permet tenir diferents versions d'un arxiu

12. Compleix els requisits dels ajuts europeus	Recomanat per projectes europeus. Desenvolupat pel CERN (projecte OpenAIRE de la UE)	Recomanat per als projectes europeus. Projecte finançat per la CE amb participació del BSC , DANS , LIBER i la Univ. d'Edimburg entre d'altres	Sense informació	Sense informació	Sense informació	Sense informació
13. Més informació (URL)	FAQS Policies About	FAQS DARUP	FAQS Features Blog Tools About	FAQS Terms of service Wiki Blog	User guide (Dataverse general) Demo (proves Harvard) About (general) Blog	FAQS Terms of use
14. Altres	- Integració amb ORCID i GitHub - Espais col·laboratius (<i>Community collections</i>) - Permet exportar cites a gestors bibliogràfics i xarxes socials	- El cercador de dades B2FIND permet cercar dades de B2SHARE i d'altres repositoris - Es pot utilitzar de manera independent B2DROP com una infraestructura per actualitzar i intercanviar dades	- Integració amb ORCID - Espais col·laboratius - Estadístiques de descàrregues - Les dades públiques es veuen a Google Scholar - Permet afegir al perfil personal enllaços a xarxes socials i a les nostres publicacions - Permet exportar cites a gestors bibliogràfics i xarxes socials	- Integració amb ORCID - Estadístiques de descàrregues - Permet exportar cites a gestors bibliogràfics i xarxes socials	- Aplicació de codi obert (una institució o investigador pot crear el seu propi Dataverse) - Integració amb OJS i DropBox - Permet obtenir la citació en format XML o RIS - Permet compartir un conjunt de dades a les xarxes socials	- Estadístiques de descàrregues

Referències

Stall, Shelley; Martone, Maryann E.; Chandramouliswaran, Ishwar; Federer, Lisa; Gautier, Julian; Gibson, Jennifer; Hahnel, Mark; Larkin, Jennie; Pfeiffer, Nicole; Sedora, Brian; Sim, Ida; Smith, Tim; Van Gulick, Ana E.; Walker, Erin; Wood, Julie; Zaringhalam, Maryam; Zigoni, Alberto. Generalist Repository Comparison Chart (version 3.0). 2023 [10.5281/zenodo.3946719](https://doi.org/10.5281/zenodo.3946719).