

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

(Versió 4, Setembre 2019)

(Doc. CO19/09) (Ciència Oberta\4 GDR\Repositori dades\1909_Recomanacions.docx, 30.09.19)

Aquest document té per objectiu recolzar els investigadors en el moment de seleccionar un repositori per al dipòsit de les seves dades de recerca atès que els projectes finançats en el marc del programa Horitzó 2020 de la Comissió Europea cal que dipositin en obert les dades de recerca seguint els principis FAIR (Findable, Accessible, Interoperable i Reusable).

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, Universitat Jaume I 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/2072/362618>

1. Selecció d'un repositori de dades de recerca

Avui en dia, la decisió d'escollir un repositori de dades 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 de dades temàtics o específics per una disciplina** ofereixen el coneixement i l'experiència en la gestió de les dades en el seu context.

Els **repositoris multidisciplinars o de dades generals** permeten el dipòsit de dades de qualsevol àrea temàtica. Normalment ofereixen funcionalitats de navegació i visualització més amigables.

1.1 Procés per seleccionar un repositori de dades

En el procés de selecció d'un repositori de dades, cal seguir els següents passos (figura 1):

1. Comprovar si es disposa d'un repositori temàtic consolidat a l'àmbit [vegeu secció 2].
2. Si és possible, sol·licitar el dipòsit de dades al repositori institucional de la universitat.
3. Utilitzar Zenodo (repositori proposat per la Comissió Europea) o un altre repositori de dades multidisciplinars [vegeu secció 3].

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

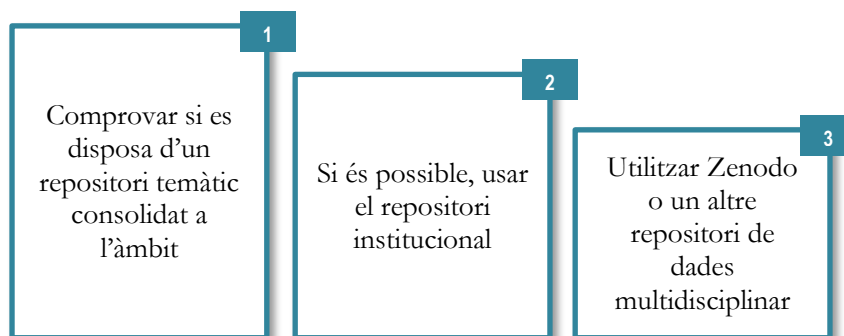


Figura 1. Proposta basada en les recomanacions d'OpenAIRE (vegeu bibliografia)

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 (vegeu bibliografia) que cal tenir en compte a l'hora de seleccionar un repositori de dades.

1. Existeix algun repositori representatiu i consolidat en el teu àmbit?
2. El repositori permet dipositar les dades en el format utilitzat?
3. La mida dels fitxers que pots dipositar s'ajusta a les teves necessitats?
4. Es permet la restricció d'accés a les dades (tancades, restringides o embargades)?
5. Els drets i llicències que t'ofereix el repositori s'ajusten amb els teus interessos?
6. El repositori atorga un identificador permanent?
7. Les condicions per dipositar o retirar contingut del repositori s'ajusten a les teves necessitats?
8. Hi ha costos associats a l'ús del repositori?
9. Es poden enllaçar les dades a les publicacions associades?
10. El repositori té una política de preservació (còpies, terminis de conservació)?
11. Permet dipositar diferents versions d'un mateix arxiu?
12. Compleix els requisits dels ajuts europeus?

2. Fonts per seleccionar un repositori de dades temàtic

A continuació, es mostren directoris de repositoris de dades de recerca i recomanacions d'editorials a l'hora de seleccionar un repositori de dades de recerca.

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](https://bit.ly/Re3Data) (bit.ly/Re3Data)
- [Data repositories OAD](https://bit.ly/OADDataRepositories) (bit.ly/OADDataRepositories)

2.2 Recomanacions d'editorials

Algunes editorials recomanen repositoris temàtics per dipositar les dades suplementàries a les publicacions. Vegeu com a exemple:

- [Elsevier](https://bit.ly/ElsevierSupportedDataRepositories) (bit.ly/ElsevierSupportedDataRepositories)
- [Nature](https://bit.ly/NatureDataRepositories) (bit.ly/NatureDataRepositories)
- [Plos](https://bit.ly/PlosDataRepositories) (bit.ly/PlosDataRepositories)
- [Springer](https://bit.ly/SpringerDataRepositories) (bit.ly/SpringerDataRepositories)

3. Fonts per seleccionar un repositori de dades multidisciplinar

A continuació es mostra una taula comparativa dels diferents repositoris multidisciplinars realitzada per la Universitat Autònoma de Barcelona i ampliada i revisada pel Grup de Suport a la Recerca del CSUC.

	 CERN	 EU DAT	 Digital Science	 NCSU	 Harvard University	 Elsevier
1. Àmbit temàtic i geogràfic	Multidisciplinar Internacional	Multidisciplinar Internacional	Multidisciplinar Internacional	Ciència i medicina Internacional	Multidisciplinar Internacional	Multidisciplinar Internacional
2. Tipologia de dades	Tot tipus de formats i també documents publicats	Tot tipus de formats	Tot tipus de formats	Tot tipus de formats incloent arxius comprimits de diversos fitxers amb preferència per formats de codi obert	Tot tipus de formats amb funcionalitats addicionals per alguns	Tot tipus de formats
3. Mida aproximada dels fitxers a incloure	Fins a 50 GB per fitxer. Per a mides superiors, contactar amb ells	Fins a 10 GB per fitxer (màxim de 20 GB per registre). Per a mides superiors, contactar amb ells	Fins a 5 GB per fitxer.	300 GB per un conjunt de dades. Per a mides superiors, contactar amb ells	Fins a 2 GB	Fins 10 GB per conjunt de dades
4. Dades obertes, embargades, restringides o tancades	<ul style="list-style-type: none"> - Obertes - Embargades - Restringides - Tancades 	<ul style="list-style-type: none"> - Obertes - Embargades - Restringides 	<ul style="list-style-type: none"> - Obertes - Embargades - Restringides - Tancades (per defecte fins que no es publiquen) 	<ul style="list-style-type: none"> - Obertes - Embargades (segons editors i sempre que l'autor ho requereixi mentre no es publica l'article) - Restringides (sempre que l'autor ho requereixi mentre es revisa per parells) 	<ul style="list-style-type: none"> - Obertes - Restringides 	<ul style="list-style-type: none"> - Obertes - Tancades (per defecte fins que es validen els requisits) - Restringides (Per institucions) - 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.	<ul style="list-style-type: none"> - CC0 per a dades i metadades - CC-BY per a no dades - MIT per a codi i software - Altres llicències: GNU, Apache 2.0, BSD 3 clause. 	CC0	Per defecte CC0 però es poden especificar altres termes d'ús .	Permet varietat de llicències per a dades (CC0, CC-BY, CC-BY-NC), software (MIT, Apache, BSD, GPL, LGPL, MPL, CeCILL) i maquinari (CERN, TAPR).
6. Identificador permanent	DOI	Handle i DOI	DOI	DOI	Handle i DOI	DOI
7. Condicions per retirar contingut	<ul style="list-style-type: none"> - Es permet retirar les dades justificant-ne el perquè. - El DOI i la URL quedaran retinguts. 	Sense informació.	Es pot sol·licitar la retirada de contingut públic justificant-ne el motiu.	<ul style="list-style-type: none"> - Cessió de dades irrevocable. - Es reserva el dret a eliminar el contingut justificant-ne el motiu. 	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 per investigadors i científics europeus.	Gratuït (fins 20 GB). Per més espai i altres prestacions, ofereixen versió de pagament .	<ul style="list-style-type: none"> - 120\$ per un conjunt de dades. - Tarifes i càrrecs addicionals. 	Gratuït.	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 publicat o en procés de publicació o dades associades a documents no revisats per parells de fonts acadèmiques. - Integració amb les publicacions.	Permet afegir enllaços a documents relacionats.	Permet afegir enllaços a documents i programaris relacionats a les publicacions associades.
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).	Les dades públiques es troben emmagatzemades a Amazon S3 i DuraCloud / Chronopolis .	- Dades emmagatzemades i permanentment disponibles a California Digital Library's Merritt Repository . - Rèplica a DataONE.	- Harvard Dataverse Preservation Policy . - Conveni amb Data Preservation Alliance per preservar les dades a llarg termini - Integració amb Archivematica, DuraCloud/Chronopolis.	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 7PM de la UE).	Recomanat pels projectes europeus. Projecte finançat per H2020 amb participació del BSC , DANS , LIBER i la Univ. d'Edimburg entre d'altres.	Sense informació.	Acord de col·laboració amb Zenodo (signat el juliol de 2019).	- Es poden fer exportacions compatibles amb OpenAIRE (a partir de la versió 4.14) - Compleix amb els principis FAIR.	Sense informació.
13. Més informació (URL)	- FAQS - Policies - Features - About	- FAQS - DARUP	- FAQS - Features - Blog - Tools	- FAQS - Terms of service - Features and technology - Blog	- Dataverse Guides - Demo (proves Harvard) - About (general) - Blog	- FAQS - Terms of use
14. Altres	- Integració amb ORCID, GitHub - Espais col·laboratius - Permet exportar cites a gestors bibliogràfics i xarxes socials.	- B2FIND permet cercar dades d'aquest 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 - 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 - Usa Make Data Count - Permet tenir la citació en format RIS i BibTex - Permet exportar cites a Mendeley - Permet compartir enllaços a continguts web a través de Reddit.	- Integració amb ORCID, Dropbox, Rspace, OSF, DataCite, OJS - Aplicació de codi obert (qualsevol pot crear el seu propi Dataverse) - Usa Make Data Count per recopilar i mostrar mètriques - Permet tenir la citació en format XML o RIS - Permet compartir un conjunt de dades a les xarxes socials.	- Integració amb ORCID, Dropbox, Drive - Certificació "Data Seal of Approval" de la indústria dels repositoris de dades. - Ofereix una API oberta - Estadístiques de visualitzacions i descàrregues per dataset.

4. Bibliografia

OpenAIRE. *Research Data Management. Briefing paper : Understanding Research Data Management* [en línia]. European Commission, February 2016. Disponible a: <bit.ly/2016OpenAIRE> [Consulta: 15.05.2017]

Repositoris multidisciplinars per a dades en Accés Obert [en línia]. Barcelona: Universitat Autònoma de Barcelona. Servei de Biblioteques, 2016. Disponible a: <bit.ly/2016UAB> [Consulta: 15.05.2017]

Selecting a Repository for Data Deposit [en línia]. Maryland: Johns Hopkins University Data Management Services, 2016. Disponible a: <bit.ly/2019Hopkins> [Consulta: 23.08.2019]

Whyte, A. *Where to keep research data: DCC checklist for evaluating data repositories* [en línia]. Edinburgh: Digital Curation Centre, 2015. Disponible a: <bit.ly/2015Whyte> [Consulta: 15.05.2017]