

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

(Versió 2, Novembre 2016)

(Doc. 16/57) (B6SR\GDR\1611RecomanacionsSeleccionarRepositoriDades.docx, 15.11.16)

Aquest document va adreçat al recolzament dels investigadors en el moment de seleccionar un repositori per al dipòsit de les seves dades de recerca atès que a partir de l'any 2017 l'*Open Research Data Pilot* (ORD Pilot) de l'Horizon 2020 requerirà a tots els projectes que rebin finançament la publicació de les dades de recerca en obert.

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 Jaume I.

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/2072/268020>

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 comportarà certa planificació a l'hora de generar les dades, ja que caldrà complir amb els requisits que allí es demanin (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 (vegeu figura 1):

1. Comprovar si es disposa d'un repositori temàtic consolidat a l'àmbit [vegeu punt 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 punt 3].

En el cas que les dades es dipositin en un repositori temàtic o multidisciplinar, informar la universitat sobre les dades dipositades i on s'han dipositat.

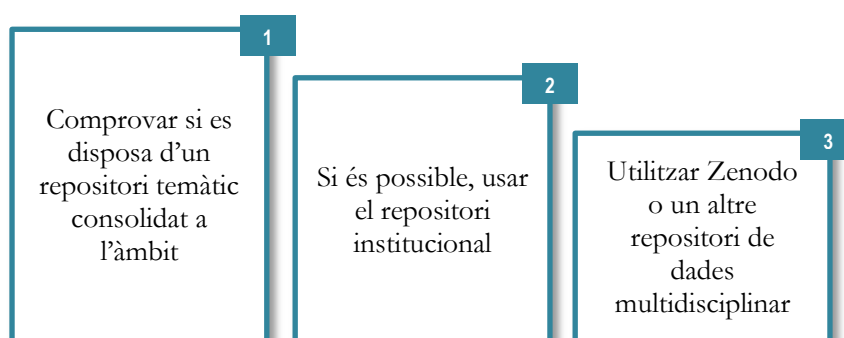


Figura 1. Proposta basada en les recomanacions d'OpenAIRE (vegeu bibliografia)

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 en el teu format?
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 són adients 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](http://bit.ly/Re3Data) (bit.ly/Re3Data)
- [Data repositories OAD](http://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](http://bit.ly/ElsevierDataRepositories) (bit.ly/ElsevierDataRepositories)
- [Nature](http://go.nature.com/DataRepositories) (go.nature.com/DataRepositories)
- [Plos](http://bit.ly/PlosDataRepositories) (bit.ly/PlosDataRepositories)
- [Springer](http://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 a partir d'un document de treball del Grup de Suport a la Recerca del CSUC.

	 CERN	 EUDAT	 Digital Science (Macmillan)	 NCSU	 IOSS (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: publicacions, literatura grisa, imatges i audiovisuals, text pla, etc.	Tot tipus de formats	Tot tipus de formats : xifres, conjunts de dades, imatges i audiovisuals, literatura grisa, conjunts d'arxius, etc.	Tot tipus de formats (amb exclusió de software) i amb preferència per formats de codi obert	Tot tipus de formats amb funcions específiques per alguns (Stata, SPSS, R, Excel, CSV, etc.)	Tot tipus de formats : text sense format, imatges i audiovisuals, PDF, HTML, etc.
3. Mida aproximada dels fitxers a incloure	Fins a 50 GB Per a mides superiors es pot contactar amb ells	Fins a 2 GB. Per a mides superiors, contactar amb ells	Fins a 5 GB	20 GB per un conjunt de dades (cost addicional per sobre de 20 GB)	Fins a 2 GB. Per fitxers més grans contactar amb ells	10 GB
4. Dades obertes, embargades, restringides o tancades	Obertes Tancades Restringides Embargades (Polícies)	Obertes Restringides Embargades. La descripció del conjunt de dades és pública.	Obertes Tancades (per defecte fins que no es publiquen) Restringides Embargades	Obertes Embargades (per defecte fins el moment de la publicació)	Obertes Restringides La descripció (títol, etc.) d'un conjunt de dades és pública però els arxius poden ser públics o restringits. Sense informació sobre embargaments	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	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	Handle	DOI	DOI	DOI	DOI
7. Condicions per dipositar o retirar contingut	Es permet retirar les dades justificant-ne el perquè. El DOI i la URL quedaran retinguts.	Per retirar contingut contactar amb: EUDAT	Sense informació	Cessió de dades irrevocable	Sense informació	Es permet retirar les dades abans de publicar-les. Una vegada publicades cal contactar amb Mendeley.

8. Cost	Gratuït. Però hi ha la possibilitat de més espai pagant (FAQS)	Gratuït	Gratuït fins a una certa mida: -5 GB pels arxius individuals -20 GB d'espai privat -Espai públic il·limitat -10 espais col·laboratius Per més espai i altres prestacions, ofereixen versió de pagament. Contacte	110 € per un conjunt de dades (de fins a 20GB) (Payment plans Data publishing charges)	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 de recerca publicat o en procés de publicació - Integració amb les publicacions (Submission integration)	Permet afegir enllaços a documents relacionats	Sense informació
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	Dades emmagatzemades a Amazon S3 , a Alemanya
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 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
9. Més informació (URL)	FAQS Policies Features 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
10. Altres	- Integració amb ORCID, GitHub i DropBox - 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 afegir la cita d'un arxiu a algunes 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	Permet afegir enllaços a documents relacionats

4. Bibliografia

OpenAIRE. *Research Data Management. Briefing paper : Understanding Research Data Management* [en línia]. European Commission, February 2016. Disponible a: <bit.ly/2016OpenAIRE> [Consulta: 15.11.2016]

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: 15.11.2016]

Selecting a Repository for Data Deposit [en línia]. Maryland: Johns Hopkins University Data Management Services, 2016. Disponible a: <bit.ly/2016Hopkins> [Consulta: 15.11.2016]

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: 15.11.2016]