

Repositori de Dades de recerca (RDR)

Context

En els darrers anys, les dades recollides, generades o utilitzades en el desenvolupament dels projectes de recerca han rebut l'atenció de la comunitat científica i dels òrgans gestors de la recerca. A nivell europeu, des del programa marc Horizon 2020 i el seu pilot Open Research Data Pilot s'ha promogut la "Gestió de dades de recerca" (Research Data Management, RDM). Aquest concepte és un paraigua que engloba diferents activitats relacionades amb la creació, organització, estructuració, emmagatzematge, preservació i compartició de les dades. És per això que els projectes finançats per la Comissió Europea requereixen que s'elabori un pla de gestió de dades (Data Management Plan, DMP) i es dipositin les dades seguint els principis FAIR (Findability, Accessibility, Interoperability, and Reuse) amb l'objectiu d'augmentar l'eficiència i la transparència de la recerca a través d'una ràpida difusió dels resultats i de facilitar-ne la seva reutilització.

Per aquest motiu, els vicerectors de Recerca que formen la Comissió Funcional de l'Àrea de Ciència Oberta del CSUC, van acordar encarregar un informe que determinés els requeriments funcionals raonables que ha de tenir un repositori de dades per a què compleixi amb els requisits FAIR i les directrius establertes per l'European Open Science Cloud (EOSC). Aquest informe, publicat al 2019, "[FAIR x FAIR. Requeriments factibles, assolibles i implementables per a un repositori de dades de recerca FAIR](#)" va establir les bases d'aquest repositori de recerca que, després d'una etapa en accés restringit, es va obrir al públic el març del 2021.

Què és l'RDR?

El CORA. Repositori de Dades de Recerca (RDR) és un repositori de dades federat i multidisciplinar que permet a les universitats catalanes, centres de recerca CERCA i altres entitats que fan recerca, la publicació de conjunts de dades de recerca en mode FAIR i seguint les directrius de l'EOSC.

Els objectius de l'RDR són augmentar l'eficiència i la transparència de la recerca a través d'una ràpida difusió dels conjunts de dades i facilitar-ne la seva reutilització.

Quines característiques té l'RDR?

L'RDR té unes característiques que beneficien a qui hi participa de la forma següent:

- Compliment dels requisits de les agències de finançament



- Publicar les dades de recerca al repositori ajuda a complir els requisits dels ajuts europeus o del Plan Nacional, entre d'altres.

- Personalització



- Permet personalitzar l'espai de cada institució participant.

- Augment de la visibilitat



- L'ús d'identificadors persistents (DOIs) i de metadades genèriques o específiques per disciplina afavoreixen la visibilitat de les dades.

- Gran capacitat



- Es permet el dipòsit fins a 10GB per defecte. Per a conjunts més grans, cal una petició especial.

- Gestió de dades i control d'accés



- Permet gestionar les dades i limitar l'accés sobre qui pot accedir a elles, així com oferir una URL privada per mostrar el dataset abans de ser públic.

- Preservació a llarg termini de les dades



- Les dades seran preservades durant 10 anys, com a mínim.

- Llicències



- Per defecte s'assigna una llicència CC0, però es permet l'ús d'altres llicències personalitzades.

- Transformació dels formats de fitxers



- Alguns formats de fitxers, com els tabulars, es transformen automàticament a formats oberts de preservació.

- Curació de les dades



- La curació a diferents nivells assegura la qualitat dels datasets.

Què és DataVerse?

L'RDR funciona amb el programari de codi lliure Dataverse, desenvolupat pel Harvard's Institute for Quantitative Social Science, que permet emmagatzemar datasets (conjunts de dades d'una mateixa recerca) tot mostrant les metadades descriptives i permetent la descàrrega dels fitxers corresponents. El programari Dataverse és àmpliament usat arreu del món i entre els seus usuaris a Europa en destaquen

- [DataverseNO](#) és el repositori Dataverse que s'està utilitzant a Noruega. Per contactar amb ells researchdata@hjelp.uit.no.
- [DataverseNL](#) és el repositori Dataverse que s'està utilitzant als Països Baixos. Per contactar amb ells info@dataverse.nl.
- [Jülich DATA](#) és un dels repositoris Dataverse que s'està utilitzant a Alemanya. Per contactar amb ells accedeix a l'apartat Contact del seu Dataverse.

CSUC i l'RDR

La coordinació del projecte l'assumeix el CSUC, que fomenta l'intercanvi d'experiències i ajut mutu entre la comunitat formada per les universitats catalanes i els centres de recerca participants de cara a desenvolupar expertesa i bones pràctiques en curació de dades.

Col·laboracions



L'RDR està basat en el programari de codi lliure Dataverse creat pel Harvard's Institute for Quantitative Social Science. L'objectiu d'aquest programari és facilitar la gestió de dades en accés obert.



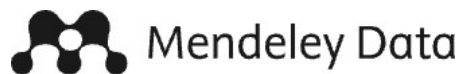
Cadascun dels datasets de l'RDR compten amb un identificador permanent DOI que es gestionen a través de DataCite. DataCite és una organització sense ànim de lucre que proporciona DOIs a la comunitat científica per a les dades de recerca i recol·lecta les metadades dels datasets incorporats a l'RDR. D'aquesta manera els datasets de l'RDR són més fàcilment descobribles per la comunitat investigadora internacional. Alhora DataCite proporciona un sistema per cercar i trobar dades de recerca.



L'RDR està registrat a Re3data.org. Re3data.org és un registre de repositoris de dades d'investigació i una eina per a la ciència oberta. Proporciona informació sobre diversos repositoris de dades d'investigació de totes les disciplines acadèmiques.



L'RDR està registrat a FAIRsharing.org. FAIRsharing.org és un recurs sobre estàndards de dades i metadades, relacionat amb bases de dades i polítiques de dades.



Els datasets de l'RDR es troben accessibles a través de Mendeley Data. Mendeley Data és un repositori a través del qual es poden cercar més de 20 milions de conjunts de dades i també permet l'emmagatzematge directe.