

Servei de suport a la gestió de dades de recerca: estat a desembre de 2022

(Doc. CO23/01) (4 RDM\2212RDMsituacio2022.docx, 20.01.23)

1. ANTECEDENTS	2
1.1 L'IMPULS DE REACT	3
2. PLANS DE GESTIÓ DE DADES	4
CORA.EINA DMP	4
Línies d'actuació pel 2023	5
3. REPOSITORIS DE DADES DE RECERCA	5
CORA.REPOSITORI DE DADES DE RECERCA.....	6
Situació.....	6
Pla de treball.....	9
Línies d'actuació pel 2023.....	9
4. POLÍTICA D'ACCÉS OBERT A LES DADES	10
5. PRESERVACIÓ	11
6. FORMACIÓ	12
Línies d'actuació pel 2023.....	13

1. Antecedents

L'any 2014 es va creure convenient ampliar les línies estratègiques d'actuacions de l'Àrea de Biblioteques del CSUC amb una dedicada al suport a la recerca (Doc. 14/51) i es va constituir un grup de treball específicament dedicat a això (acord de la CT de 25.04.14). El Grup de treball de Suport a la Recerca va establir un pla de treball (Doc. 15/54) que tenia com objectius:

- Elaborar i acordar procediments que permetessin crear Plans de Gestió de Dades i recomanar un repositori propi on dipositar les dades.
- Si es creies oportú, recomanar la creació d'un repositori consorciat de dades de recerca en el marc del CSUC.
- Elaborar i proposar l'aprovació d'uns principis generals de política a seguir per a la gestió de les dades de recerca per part de les universitats de Catalunya.

Fruit de les activitats del grup de treball es va considerar que totes les universitats membres del CSUC des de les seves biblioteques i/o oficines de recerca podien iniciar un servei de suport a la gestió de dades de la recerca (*Research Data Management*, RDM) a partir de setembre de 2016 (Doc. 16/45) que consistia en: suport per la creació de plans de gestió de dades i recolzament a l'obertura i dipòsit de les dades de recerca a través de recomanacions per seleccionar un repositori i de l'ampliació de les prestacions dels repositoris institucionals.

L'any 2017 la SUR va suggerir que el CSUC dediqués una atenció especial a la Ciència oberta, creant una nova Àrea de Ciència Oberta (CE 30.01.17). En el *Pla de Treball per al període 2017-2019* (Doc. CO17/01) s'indicava que els objectius en RDM eren: assessorar els investigadors/es en la confecció de plans de gestió de dades i garantir l'existència de repositoris per a dades de recerca.

Així doncs, des de l'inici del servei, les activitats de l'Àrea respecte RDM han girat entorn aquests grans eixos:

- els plans de gestió de dades (*Data Management Plans*, DMP)
- els repositoris de dades de recerca
- la preservació a llarg termini de les dades de recerca
- la política d'accés obert a les dades
- la formació en gestió de dades

1.1 L'impuls de REACT

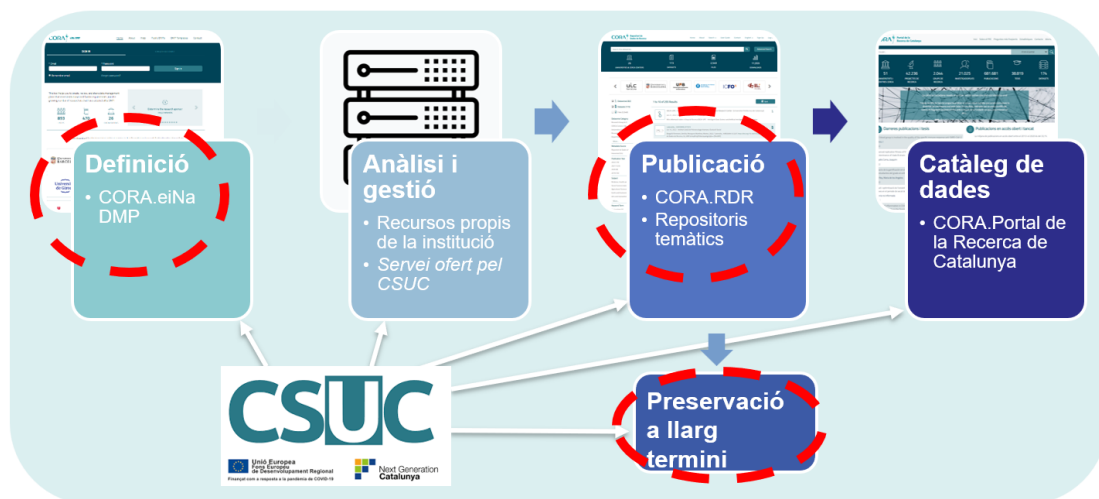
El CSUC va rebre el desembre de 2021 un ajut europeu de 12,73 M€ procedents del fons REACT-EU (Ajuda a la Recuperació per a la Cohesió i els Territoris d'Europa), un dels mecanismes principals del Next Generation, el Pla per a la reactivació de l'economia europea per superar els efectes de la pandèmia.

El Projecte REACT-EU, dins l'actuació “Ampliació i millora dels serveis de dades” vol posar en marxa serveis relacionats amb les dades de recerca produïdes pel sistema català de coneixement. L'objectiu és tenir un servei que faciliti la gestió de dades de recerca, la seva publicació en obert seguint les directrius dels principis FAIR i la seva preservació. Per això, es van encarregar dos informes que fessin un escaneig dels diferents serveis de dades que s'estan oferint en el context europeu¹ i americà².

El servei de publicació de dades té per objectiu posar en marxa un servei que en faciliti la publicació en obert seguint les directrius dels principis FAIR. Aquest servei es compondrà de diferents elements orientats, fonamentalment, a facilitar: la creació de plans de gestió de dades (DMPs), la curació de les dades i la seva publicació al repositori CORA.RDR.

A més a més, per tal de complir amb les exigències dels plans de l'EOSC (European Open Science Cloud) respecte els principis FAIR, els conjunts de dades fets públics han de ser reutilitzables en els propers anys, és a dir, se n'ha de garantir la seva preservació. Per aquest motiu, també cal dissenyar aquest servei.

L'esquema següent il·lustra la complementarietat de les diferents accions respecte les dades de recerca que avui està duent a terme el CSUC:



¹ López-Borrull, Alexandre (2022). *Serveis i nodes de dades de recerca operatius a Europa*. Consorci de Serveis Universitaris de Catalunya. <<http://hdl.handle.net/2072/521726>>. [Consulta: 14-01-23].

² Llebot Lorente, Clara (2022). *Serveis per als investigadors en dades de recerca als USA*. Consorci de Serveis Universitaris de Catalunya. <<http://hdl.handle.net/2072/521699>>. [Consulta: 14-01-23].

2. Plans de gestió de dades

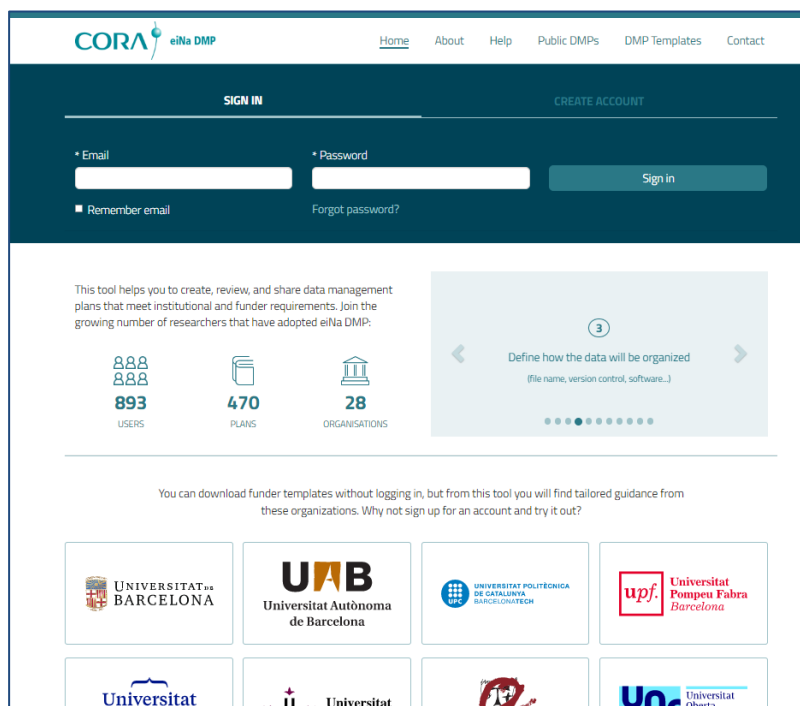
Per ajudar els investigadors/es a dur a terme una millor recerca és necessari disposar de plans de gestió de dades (DMP), un concepte que abasta tasques com ara la recopilació de les dades utilitzades, la seva organització, documentació, emmagatzematge i preservació. Es tracta, per tant, d'elaborar documents que descriuen el cicle de vida de les dades de recerca recopilades i generades en el transcurs d'un projecte.

Diferents agències de finançament, entre els quals en destaca la Comissió Europea, obliguen a que els projectes de recerca tinguin tots un DMP.

Per donar recolzament a la creació dels DMPs s'han elaborat unes guies per a diferents agències i programes de finançament, que s'actualitzen periòdicament) on es detalla la informació que ha de contenir el pla, amb descripcions i exemples reals. A més a més, aquestes guies es troben instrumentalitzades a través de l'aplicació **CORA.eiNa DMP** (dmp.csuc.cat).

CORA.eiNa DMP

CORA.eiNa DMP és un aplicatiu que permet treballar de forma col·laborativa i en línia en la creació d'aquests plans pels diferents organismes de finançament o col·lectius, consultar guies i exemples, compartir els plans amb d'altres investigadors/es atorgant permisos de lectura, escriptura o co-administració, sol·licitar la revisió dels plans pel servei de suport a la gestió de dades de la universitat o centre de recerca i exportar-los en diferents formats per a presentar-los als organismes de finançament.



L'eina és gratuïta i representa un estalvi important de temps i esforç a la comunitat investigadora. L'any 2021, l'eina era usada per 20 institucions (13 universitats i 7 centres CERCA) i s'havien creat fins al moment 406 DMPs. En canvi, un any després, el desembre del 2022 l'eina era usada per 29 institucions (12 universitats i 17 centres CERCA) i amb ella s'havien creat 534 DMPs.

Línies d'actuació pel 2023

Durant el 2023 s'adequarà el servei i l'aplicatiu per fer plans de gestió de dades que permetin la vinculació entre aquests plans i els seus resultats de la recerca finals (datasets, programari, imatges, etc.).

3. Repositoris de dades de recerca

Als inicis del servei d'RDM es va tenir com objectiu el de poder assessorar als grups de recerca sobre quin repositori complia millor els seus requeriments per a dipositar-hi les dades. A tals efectes es va fer el document "[Recomanacions per seleccionar un repositori per al dipòsit de dades](#)" que mostra les millors opcions disponibles de repositoris de dades temàtics i multidisciplinaris i guia als investigadors en el moment de seleccionar un repositori per al dipòsit de les seves dades. Alhora també es van fer modificacions menors als repositoris institucionals per permetre-hi el dipòsit de dades.

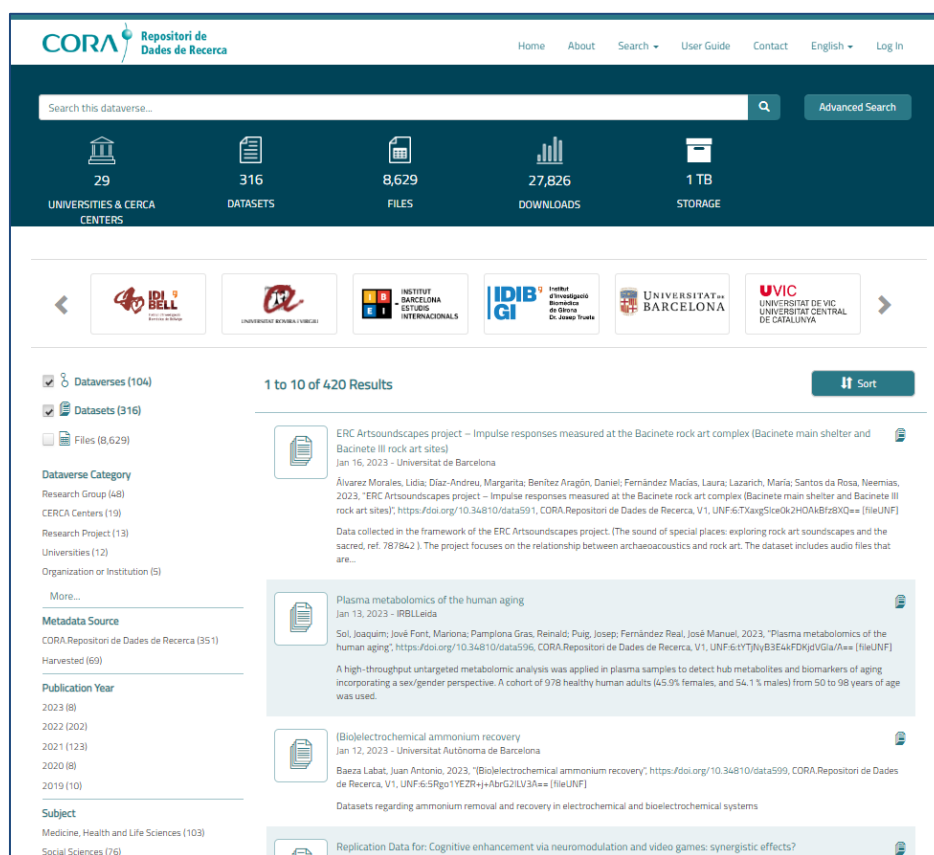
El Pla de Treball de l'àrea recollia com a objectiu per finals del 2018 tenir un repositori de dades del sistema de recerca de Catalunya, tal com feien llavors els països i institucions més avançats en aquesta matèria. Es deia que 'per economia d'escala, el repositori hauria de ser cooperatiu però no necessàriament centralitzat. La Comissió Funcional de l'àrea va aprovar el novembre del 2017 iniciar els treballs per elaborar una proposta de requeriments funcionals mínims raonables per a la creació d'un repositori de dades de recerca. L'elaboració d'aquesta proposta de requeriments va fer-la un grup de treball i va tenir en compte les millors experiències estrangeres.

El document de requeriments "[FAIR x FAIR: Requeriments factibles, assolibles i implementables per a un repositori de dades de recerca FAIR](#)" (Doc. CO19/03) va presentar-se a la **Comissió Funcional de l'Àrea de Ciència Oberta** del 27.03.19 que el va aprovar i que va demanar un pla de treball que tingués per objectiu fer un repositori amb les característiques indicades i que anés acompanyat de la memòria econòmica corresponent. El pla i la memòria van ser presentats a la Comissió Funcional del 20.05.20 que va acordar "fer un repositori de dades de recerca (RDR) per a les universitats de Catalunya, que cobreixi una part de les necessitats dels centres CERCA i que permeti publicar les dades de manera FAIR", així com que "inicialment fer el repositori de dades de recerca amb els recursos actualment disponibles al CSUC".

CORA.Repositori de dades de recerca

Entre el setembre de 2020 i febrer de 2021, es va posar en funcionament de forma pilot **CORA.Repositori de Dades de Recerca** (a partir d'ara, **CORA.RDR**). Posteriorment, al març de 2021, es va obrir el seu ús a les institucions que el volguessin utilitzar. Des de llavors, 30 institucions (11 universitats i 19 centres de recerca CERCA) l'usen de manera desigual. També des de llavors ha estat en funcionament una Comissió Tècnica de CORA.RDR que valida el pla de treball, així com fa avançar les funcionalitats i millores del repositori.

CORA.RDR és un repositori de dades federat i multidisciplinari 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.



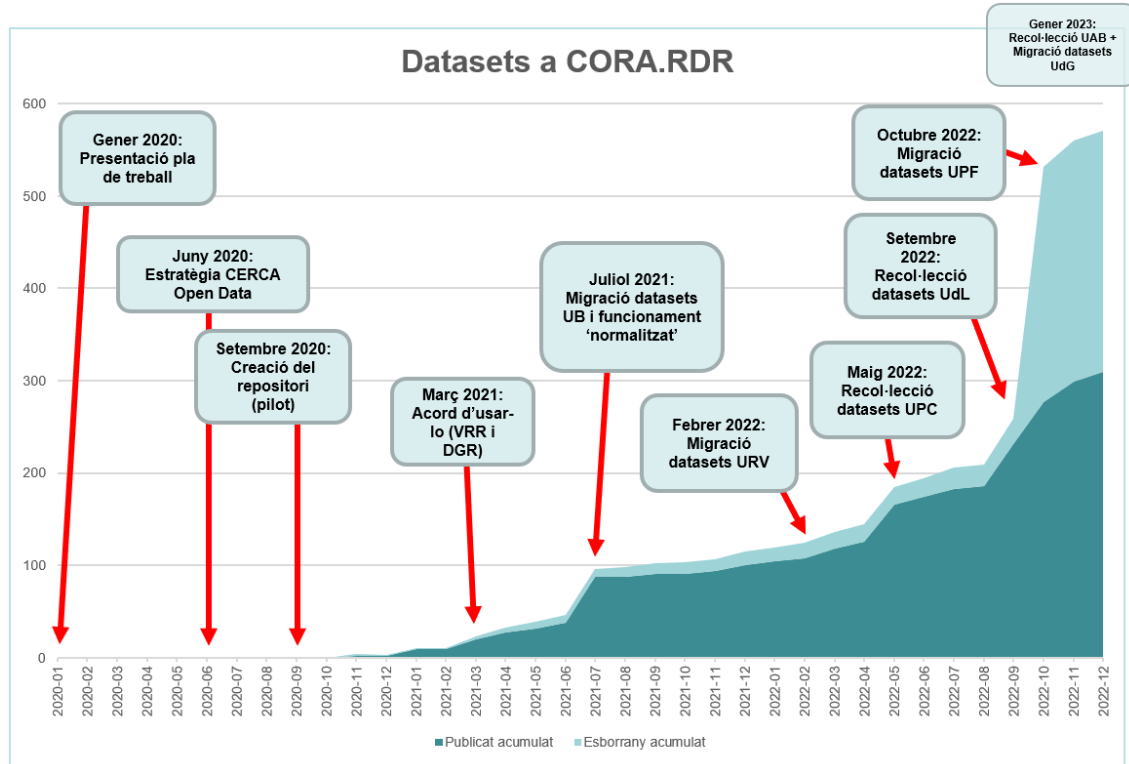
The screenshot shows the CORA.RDR website interface. At the top, there is a navigation bar with links for Home, About, Search, User Guide, Contact, English, and Log In. Below this is a search bar with the text "Search this dataverse..." and an "Advanced Search" button. A dashboard section displays key statistics: 29 UNIVERSITIES & CERCA CENTERS, 316 DATASETS, 8,629 FILES, 27,826 DOWNLOADS, and 1 TB STORAGE. Below the dashboard, there are logos for partner institutions: IDIBELL, UNIVERSITAT DE VALÈNCIA, INSTITUT BARCELONA ESTUDIS INTERNACIONALS, IDIBGI, UNIVERSITAT DE BARCELONA, and UVIC. The main content area shows search results for "1 to 10 of 420 Results". On the left, there are filters for "Dataverses (104)", "Datasets (316)", and "Files (8,629)". Below these are filters for "Dataverse Category" (Research Group (48), CERCA Centers (19), Research Project (13), Universities (12), Organization or Institution (5)), "Metadata Source" (CORA.Repositori de Dades de Recerca (351), Harvested (69)), "Publication Year" (2023 (8), 2022 (202), 2021 (123), 2020 (8), 2019 (10)), and "Subject" (Medicine, Health and Life Sciences (103), Social Sciences (76)). The search results list several datasets, including "ERC Artsoundscapes project", "Plasma metabolomics of the human aging", and "(Bio)electrochemical ammonium recovery".

Situació

Les dades d'ús de CORA.RDR són les que segueixen a sota:

- 30 institucions participants (11 universitats, 19 centres CERCA)
- 302 *datasets* publicats i 260 en procés de publicació
- 887GB d'ocupació en una màquina de 5TB

La gràfica del seu creixement mostra les diferents fases que ha viscut el repositori: creació, fase pilot, període inicial i un funcionament més generalitzat a partir del primer trimestre de 2021 quan el DGR i els vicerectors de recerca de les universitats van acordar usar-lo.



El 2022 conclou de manera favorable degut al creixement de la participació activa de les diferents institucions (tant d'universitats, com de centres de recerca). A més a més, s'ha contactat a 13 noves institucions per a que s'incorporessin a CORA.RDR, així com s'ha fet un seguiment estret a 10 institucions més per formar-los en usar el repositori.

A continuació es mostra la diferència dels centres entre el 2022 i l'any anterior. Entre parèntesi es dona el nombre de dats i l'asterisc significa que disposen del conveni signat:

UNIVERSITATS	2021	AMB INSTÀNCIA (ACTIUS)	AMB INSTÀNCIA (INACTIUS)	EN PROCÉS	CONTACTATS PERÒ SENSE PROCÉS	NO CONTACTATS
		UB (67) *	UAB *		UIB	UAO
		UPF (6) *	UPC			
		UdG (2) *	UVic-UCC *			
		UdL (3) *	URL			
		URV (2)	UIC			
		UOC (3)				
	2022					
	UB (93) *	UVic-UCC *	UIB		UAO	
	UAB (1) *	UIC				
UPC (46) *						
UPF (38) *						
UdG (11) *						
UdL (47) *						
URV (6) *						
UOC (9)						
URL (1)						

CENTRES DE RECERCA CERCA	2021	AMB INSTÀNCIA (ACTIUS)	AMB INSTÀNCIA (INACTIUS)	EN PROCÉS	CONTACTATS PERÒ SENSE PROCÉS	NO CONTACTATS
		Agrotecnio (6)	CRM	IDIBAPS	IRSICAIXA	CIMNE
		ICAC (3)	CRAG	IDIBGI	ISGLOBAL	CREAF
		IRTA (6)	CTFC	ICFO		CREI
			IBEI			CRG
			IDIBELL			CTTC
			IFAE			ICP
			IJC			IGTP
			IRB BCN			IISPV
			VHIR			IREC
			IRB Lleida			VHIO
						ICRPC
						ICN2
				IMIM		
				IIEC		
				CVC		
				I2CAT		
				IR ST PAU		
				ICIQ		
				IBEC		
			IR SJD			
			ICRA			
			CED			
			IPHES			
2022						
Agrotecnio (25)	CRM	IDIBAPS	ICRPC	CIMNE		
CRAG (1) *	IBEI	I2CAT	ICN2	CREAF		
CTFC (6)	IDIBGI	IR ST PAU	IMIM	CREI		
ICAC (16) *	IFAE	ICIQ	IIEC	CRG		
ICFO (1)	IJC	IR SJD	CVC	CTTC		
IPHES (4)	IRB BCN	ICRA		ICP		
IRSICAIXA (1) *	VHIR	CED		IGTP		
IRTA (7)	IRB Lleida			IISPV		
ISGLOBAL (10)				IREC		
IBEC (1) *				VHIO		
IDIBELL (1)						

Pla de treball

Durant el 2022 s'han dut a terme diferents activitats proposades i validades en el pla de treball (Doc. CO22/01, Doc. CO22/22):

- Difusió. S'han fet reunions de la Comissió Tècnica i contactat amb nous usuaris.
- Formació. S'han fet 4 sessions d'autoaprenentatge per a institucions actives i iniciat converses per fer alguns "Carpentry" per a les institucions actives.
- Organització. S'han signat els convenis amb totes les universitats i alguns amb els centres de recerca; s'han començat els treballs per desenvolupar el reglament i procediments pel repositori.
- Qualitat. S'ha revisat el grau de compliment de metadades mínimes obligatòries, els fitxers no transformats en la ingesta dels datasets i comprovat el grau de FAIRness de l'RDR mitjançant l'eina F-UJI.
- Interoperabilitat. Les metadades s'exporten als principals recollidors (OpenAIRE, Google Dataset Search i RECOLECTA) i s'han iniciat els treballs per exportar metadades dels datasets als CRIS.
- Preservació. S'han iniciat les accions que han de permetre aconseguir el CoreTrustSeal.
- Funcionament. S'ha actualitzat el software a la versió 5.11, cosa que permet diverses millores, s'ha fet la traducció al català del repositori i activat UNIFICAT per a utilitzar les credencials institucionals.

Línies d'actuació pel 2023

En el Pla General d'Actuacions del CSUC per al 2023 s'hi estableixen dos objectius principals per al repositori:

- Completar la posada en marxa del servei de publicació de dades FAIR.
- Definir les condicions de preservació digital del repositori CORA.RDR per obtenir el CoreTrustSeal

4. Política d'accés obert a les dades

De cara a donar un suport institucional a la gestió de dades de recerca el 2016, els vicerectors de recerca van acordar, com a presa de posició conjunta, una “[Proposta per establir una política d'accés obert a les dades de recerca a les Universitats de Catalunya](#)” (Doc. 16/30). Aquesta era una guia per facilitar la confecció de polítiques institucionals a cada universitat.

En aquesta reunió també es va acordar que el grup de treball de Suport a la Recerca elaborés un model de política d'accés obert a les dades de recerca amb la intenció de facilitar que cada universitat aprovés una política pròpia. Aquest document es va aprovar en forma de recomanació (Doc. CO18/13, “[Model de política de gestió de dades de recerca per a una universitat](#)”) amb l'objectiu que servis de pauta per a la confecció de plans específics per cada universitat.

Algunes universitats han aprovat documents de política d'accés obert a les dades de recerca, concretament les que es citen a sota:

- [Política de gestió de dades de recerca de la Universitat de Barcelona](#) (2019)
- [Política institucional d'accés obert per a les dades de recerca de la Universitat Autònoma de Barcelona](#) (2020, act. 2022)
- [Política institucional d'accés obert per a les dades de recerca de la Universitat Rovira i Virgili](#) (2021)
- [Política institucional de coneixement obert de la Universitat Oberta de Catalunya](#) (2021)
- [Política de gestió de dades de recerca de la Universitat de Girona](#) (2022)

El 2018 el Patronat de la Institució CERCA aprovà encarregar al director la implementació d'una iniciativa d'Open Data per als centres CERCA per facilitar i racionalitzar les infraestructures de dades, treballar amb els criteris europeus FAIR i disposar d'un Pla de Gestió de Dades adequat a cada institució i projecte. El Novembre de 2018 es va aprovar el Codi de Conducta dels Centres CERCA que estableix que cada Centre ha de tenir una estratègia d'Open Science i adaptar-se als principis europeus FAIR, aplicar plans de gestió de dades quan escaigui, i tenir cura i emmagatzemar les dades de recerca durant un període recomanat de 10 anys.

El 2020, el Patronat de la Institució CERCA va aprovar l’[“Estratègia Open Data de CERCA”](#). Aquesta té vocació de coordinació amb altres agents del sistema així com amb altres iniciatives transversals que puguin tenir lloc a Catalunya com el Pacte Nacional per a la Societat del Coneixement, un cop sigui aprovat i vigent.

5. Preservació

El gener del 2021 es va signar un conveni amb el Consorcio Madroño de col·laboració per al suport remot de repositoris.

El conveni té per objectiu regular la prestació recíproca d'un servei d'allotjament de còpies corresponents als conjunts de dades dels datasets de CORA.RDR i e-cienciaDatos. Aquestes còpies s'allotgen en els sistemes de l'altra institució en qualitat de còpia de suport.

El mes de febrer del 2022 ja estaven disponibles als dos consorcis les còpies replicades dels datasets dels respectius repositoris. Aquestes rèpliques a distància es fan amb la política estàndard: diàries incrementals, full setmanal, full mensual, bi-mensuals i anuals amb el sistema de backups que fa el WBS.

L'ajut REACT ha permès potenciar les accions adreçades a millorar les prestacions de preservació del repositori. Concretament, el novembre de 2022, es va disposar de l'informe fet pel Dr. Miquel Tèrmens: "Avaluació NDSA dels nivells de preservació del repositori CORA.RDR del CSUC" (vegeu annex 1). Aquest informe avalua el grau amb que els serveis analitzats apliquen la metodologia establerta per la National Digital Stewardship Alliance (NDSA) dels Estats Units, que és una metodologia d'autoavaluació, no preparada per ser auditada ni certificada per tercers.

Es conclou que:

- Els punts forts del repositori són:
 - Tenir una bona infraestructura d'emmagatzematge, que no arriba a òptima per la manca d'una tercera còpia de les dades;
 - que el control d'accessos disposa de tots els punts requerits;
 - les metadades estan reglamentades i segueixen els estàndards.
- Els principals aspectes a millorar són:
 - comprovar la integritat dels continguts, així com refer els possibles continguts corromputs.
 - fer control de virus en el moment de la ingesta de continguts.
 - disposar de l'Esquema Nacional de Seguretat (ENS) o la ISO27001.

6. Formació

El pla de treball de l'àrea de Ciència Oberta 2017-2019 (Doc. CO17/01) tenia com objectiu per al 2018 el d'“Elaborar un programa de formació en Ciència Oberta per a les persones que a les universitats i centres de recerca donen suport als investigadors/es” on s'inclouïa la gestió de dades de recerca.

L'informe “[FAIRxFAIR. Requeriments factibles, assolibles i implementables per a un repositori de dades de recerca FAIR](#)” (Doc. CO19/03) recomanava “Fer formació sobre els conceptes de Ciència Oberta, en general, i, concretament, sobre gestió de dades de recerca. Seguint les directrius de l'*European Commission Expert Group on FAIR data*, aquesta formació ha d'incloure la totalitat de membres de la comunitat universitària però, per ser eficaç, hauria de distingir entre usuaris avançats, investigadors/es joves i personal de suport de les universitats.”

Les activitats de formació en gestió de dades per al personal de suport a la gestió de dades de recerca d'universitats i centres CERCA han estat:

- Entre 2015 i 2019, es va iniciar diferents cursos sobre gestió de dades de recerca, impartits per Paul Ayrís (University College London Library Services), Fernanda Peset (Universitat Politècnica de València), Xavier García-Massó (Universitat de València), Eloy Rodrigues (Universidade do Minho) i Heather Andrews (Technische Universiteit Delft).
- Entre el 2019 i el 2021, s'han realitzat 10 cursos de formació per al personal de suport a la gestió de dades de recerca d'universitats i centres de recerca CERCA. Entre els diferents formadors se'n destaca: Heather Andrews (Technische Universiteit Delft), Isabel Bernal (CSIC), Philipp Conzett (The Arctic University of Norway), Ignasi Labastida (Universitat de Barcelona), Clara Llebot (Oregon State University), Eloy Rodrigues (Universidade do Minho) i Nadia Tonello (Barcelona Supercomputing Centre).
- A més a més, durant el 2020, es va dur a terme un cicle de conferències sobre la gestió de dades de recerca a càrrec de Mercè Crosas, responsable de la gestió de dades de recerca a Harvard University.
- L'any 2021 i 2022, les institucions participants a CORA.RDR també han participat en 11 tallers (7 al 2021 i 4 al 2022) autoformatius de publicació de dades. L'objectiu d'aquestes sessions era agrupar al personal que donava suport en la publicació de datasets i dotar-los d'eines comunes per tal de curar els datasets publicats a CORA.RDR. En un primer moment, aquestes sessions eren conduïdes per Heather Andrews, data curator de la TUDelft, i, posteriorment, pel propi personal del CSUC.
- Paral·lelament, l'any 2022 també s'han fet formacions a mida per diferents institucions. Per exemple, universitats com la Universitat Politècnica de Catalunya, la Universitat Pompeu Fabra o la Universitat Autònoma de Barcelona han fet formacions per al personal implicat en el dipòsit i publicació de dades. En canvi, a centres de recerca CERCA com l'ICAC, l'IDIBELL o l'IPHES (entre d'altres), s'han fet nombrosos actes més destinats al suport i la promoció de la gestió de dades de recerca.

Línies d'actuació pel 2023

L'any 2021 es va encarregar un informe que fes una proposta de formació en ciència oberta als doctorands del sistema de recerca de Catalunya³. L'ajut REACT ha permès valorar la realització d'aquest pla de formació que es faria de forma no presencial i que es preveu tenir acabat el 2023 de cara a iniciar-lo aquest mateix any.

³ López-Borrull, Alexandre (2021). *Ciència oberta a Catalunya 2021. Proposta de formació transversal per al sistema universitari català*. Consorci de Serveis Universitaris de Catalunya.<<http://hdl.handle.net/2072/520605>>. [Consulta: 15-01-23].