

CERCA OPENDATA

ESTRATÈGIA DE GESTIÓ DE DADES

Antecedents

- En el marc del programa Horizon 2020 (2014-2020) la Comissió Europea va iniciar un pilot sobre dades de recerca anomenat ORD Pilot, que l'any 2017 es va estendre a tots els projectes. Continuant amb aquesta estratègia, en el proper programa marc Horizon Europe serà obligatori publicar les dades de forma oberta i fer un pla de gestió de dades.
- El 27 de març de 2018 la I-CERCA i el Barcelona Supercomputing Center (BSC-CNS) van signar un acord marc de col·laboració.
- El 4 de maig de 2018 el Patronat de la I-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 en el seu punt 2 estableix "Open Access to Research Data" és a dir, 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 maig de 2019 es va crear el grup de treball CERCA sobre Gestió de dades de recerca, el qual es va presentar posteriorment a la Conferència CERCA i s'ha anat reunint periòdicament.
- El 6 de juny de 2019 la I-CERCA va organitzar una jornada sobre Open Data i gestió de dades amb ponents d'EUDAT, del Dutch National Data Service Infrastructure, i del Govern de Flandes, entre d'altres.
- L'any 2020, s'aprova a Catalunya el Pacte Nacional per la Societat del Coneixement el qual conté un annex específicament dedicat a la implementació d'una estratègia catalana de ciència oberta.

Pla d'acció

Aquest Pla neix amb la vocació de coordinació amb altres agents del sistema, com ara les Universitats, el Consorci de Serveis Universitaris de Catalunya (CSUC) el BSC i altres. 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.

1. Reunir periòdicament el grup de treball CERCA sobre Gestió de dades de recerca, de caràcter consultiu, encarregat del seguiment d'aquesta estratègia cada 3 mesos per discutir, proposar solucions i facilitar la implementació de les diferents accions a emprendre i els paràmetres o indicadors de seguiment de les accions. En paral·lel, reunir un grup d'experts tècnics (GETEC) dels centres CERCA per contribuir a refinar adequadament les accions referents a infraestructura i repositoris.

Termini: cada 3 mesos aproximadament

Responsable: I-CERCA

2. Proposar als centres CERCA un model o models de Pla de Gestió de Dades ponderat en funció de cada disciplina científica, per a les dades científiques relacionades amb la seva producció científica. En una primera fase es cercaran els models adequats i les ontologies estàndard per a les metadades. En una segona fase es procedirà a fer una adaptació per als centres CERCA.

Termini: Caràcter urgent

Responsable: ICERCA en coordinació amb els centres CERCA, el CSUC i les universitats catalanes

3. Adherir-se, si s'escau, al Repositori d'Open Data del CSUC i determinar quin és el programari i format necessari per facilitar la interconnexió entre el repositori i els ordinadors on s'emmagatzemen les dades. En el seu defecte, i si no és possible la primera opció crear un repositori de recerca per als centres CERCA.

Termini: 3 mesos

Responsable: CSUC, ICERCA (GETEC)

4. Fer formació als investigadors per promoure projectes de recerca amb dades compartides de diferents institucions d'R+D. Els investigadors han de ser capaços de treballar amb dades d'altres investigadors i poder facilitar les seves en obert des d'un repositori. Caldrà fer un focus especial en investigadors predoctorals i postdoctorals (propera generació de recercaires). També caldrà fer formació per ajudar a omplir els plans de gestió de dades (DMP)

Termini: 6 mesos

Responsable: ICERCA, Centres CERCA, CSUC

5. Fer formació per formar tècnics en gestió de dades que puguin donar servei als centres CERCA (formació de formadors), i així contribuir a crear la posició de Chief data officer on s'escaigui. Incloure en aquesta formació la utilització dels DMPs proposats per la ICERCA (per formar formadors a cada centres CERCA).

Termini: 6 mesos

Responsable: ICERCA

6. Elaborar una Guia CERCA de gestió de dades en els processos de recerca científica, proposar tipus d'arxius, assignar DOI o equivalent (com a identificador persistent) a les dades, definir unes metadades generals (incloent ORCID per identificar als investigadors), i inferir anualment el volum de dades previst. Analitzar l'adaptació i incorporació de dades generades anteriorment a un repositori de dades. En el cas de disciplines amb repositoris disciplinaris existents, cal respectar les dinàmiques iniciades. Cal valorar la possibilitat que I-CERCA esdevingui un membre de la comunitat DataCite per poder generar i atorgar DOIs, o bé alternativament treballar amb alguna altra institució que pugui fer aquesta funció de sistema.

Termini: 6 mesos aprox.

Responsable: ICERCA amb el suport dels centres CERCA

7. Fer una iniciativa col·laborativa del sistema CERCA de gestió, xarxa, emmagatzematge físic, i preservació de dades de recerca (infraestructures), incloent les còpies remotes en cinta. La capacitat d'emmagatzemament ha de ser creixent anticipant els escenaris de generació de dades dels centres CERCA. La preservació hauria de ser mínim de 10 anys d'acord amb el compliment de les recomanacions del Codi de Conducta CERCA. Valorar per part del GETEC i els experts de CSUC i BSC, si escau, integrar en aquesta iniciativa les capacitats computacionals de cada centre CERCA, i definir nodes d'infraestructures de dades del sistema (interns i externs a CERCA) a Catalunya. Considerar la següent proposta de nodes:
 - a. Node I. Fitxers fins a aprox. 10 GB (CSUC Nexus)
 - b. Node II. Fitxers de més de 10 GB aprox. (IFAE-PIC)
 - c. Node III (BSC).
 - d. Node IV. Còpies en cinta geogràficament allunyades (CSUC GEPA-Lleida o CERCA CTFC Solsona).

Termini: 6 mesos

Responsable: ICERCA (GETEC), Centres CERCA, Generalitat de Catalunya

- 8.** Redactar un protocol d'actuació sobre la seguretat de les dades i de l'anella científica, amb el suport del CESICAT, sobre la protecció de les dades sensibles (dades personals, dades d'interès comercial, etc.) i sobre l'embargament temporal de les dades que ho requereixin.

Termini: 1 any

Responsable: CESICAT, ICERCA (GETEC)

- 9.** Determinar el cost dels serveis comuns associats a la gestió, emmagatzematge, consulta, i recuperació de dades de recerca dels centres CERCA segons el seu volum de dades. Analitzar la seva viabilitat econòmica en contraposició al manteniment de les infraestructures de dades individuals de cada Centre CERCA. Migrar progressivament a un sistema col·laboratiu d'emmagatzemament de dades, limitant l'adquisició de grans ordinadors fora dels nodes esmentats a l'acció 7.

Termini: 1 any

Responsable: ICERCA, Centres CERCA

- 10.** Definir un programa d'incentius perquè els centres CERCA migrin cap a un sistema majoritàriament FAIR d'acord amb el Codi de Conducta CERCA, facilitant la gestió i, si escau, accés compartit de les dades de recerca. Proposar a la Generalitat de Catalunya que consideri atorgar una assignació pressupostària addicional per facilitar la implementació d'aquesta estratègia pionera. Establir un protocol per determinar quines dades han de ser Open.

Termini: 1 any

Responsable: ICERCA, Generalitat de Catalunya

- 11.** Reforçar l'accés a l'anella científica de Catalunya per tenir l'ample de banda adequat a les necessitats dels centres CERCA. Planificar escenaris futurs de demanda de Xarxa.

Termini: 2 anys

Responsable: ICERCA, CSUC, Generalitat de Catalunya

- 12.** Mesurar sistemàticament els volums de dades generats, accessibles en accés obert, consultats, etc. per facilitar la planificació a mig/llarg termini en la gestió de dades

Termini: 2 anys

Responsable: ICERCA

- 13.** Curar les dades (millorar la qualitat i format). Aquesta restarà en mans de cada centre CERCA. Caldrà valorar eines com ara un robot de revisió i també un helpdesk intern o comú al sistema CERCA.

Termini: Contínuament

Responsable: Centres CERCA

- 14.** Proposar adaptar els organismes finançadors i acreditadors (I-CERCA, AGAUR, AQU, ACCIO, etc.) de la recerca perquè s'alineïn amb els principals objectius d'aquesta estratègia de Gestió de dades (Dades FAIR/Open). La I-CERCA valorarà l'evolució i la participació de cada centre CERCA en aquesta estratègia en l'avaluació periòdica CERCA.

Termini: Contínuament

Responsable: ICERCA

Pla d'acció

	1r any	2n any	3r any	A partir 3r any
1. Inici de l'estratègia i creació de dos grups de treball	■			
2. Pla de gestió de dades en funció de cada disciplina científica	■			
3. Construcció d'un repositori de dades de recerca	■			
4. Formació per al personal investigador	■	■	■	
5. Formació del personal tècnic	■	■	■	
6. Guia CERCA de gestió de dades	■			
7. Nodes i infraestructura. Gestió, xarxa, emmagatzematge i preservació.	■	■	■	
8. Protocol de seguretat de les dades i de l'anella científica		■		
9. Model de Gestió de costos		■		
10. Programa d'incentius		■	■	■
11. Reforç de l'accés a l'Anella Científica i planificació d'escenaris futurs			■	■
12. Mesura sistemàtica de les dades, trànsit, dades Open, FAIR i planificació			■	■
13. Curació de les dades			■	■
14. Alineació dels organismes finançadors i acreditadors.			■	■



Via Laietana, 2
08003 Barcelona
cerca.cat

Patrons

