

# Reglament de CORA.Repositori de Dades de Recerca

Doc.CO23/30 - v1

## 1. Criteris per dipositar

---

### 1.1 Qui pot dipositar?

- Per dipositar i publicar un conjunt de dades, almenys un/a dels autors/es del conjunt de dades ha d'estar afiliat/da a la universitat o centre de recerca en el moment de l'obtenció i/o generació de les dades. En el cas que cap autor/a estigui afiliat/da, també s'acceptaran aquells conjunts de dades d'autors/es que tinguin l'autorització de la institució.
- Es demana que almenys un/a dels/les autors/es disposi d'un identificador únic, com és l'ORCID, i que s'inclogui aquesta informació a les metadades.
- La publicació al repositori requerirà de la validació de la interlocució tècnica de la institució segons les necessitats o funcionament de l'entitat responsable de les dades (universitat o centre de recerca).
- Recau en l'autor/a informar a l'investigador/a principal i/o director/a del projecte de recerca i tenir-ne el seu vistiplau previ. De la mateixa manera, els/les doctorands/es seguiran els procediments de la seva institució.
- Les institucions podran aprovar casos excepcionals on hi hagi un lliurament col·lectiu de la pròpia institució o d'una de les seves unitats on no s'hagi de consignar autories individuals. Tot i així caldrà indicar a les metadades el nom d'una persona com a dipositant i contacte del conjunt de dades.

### 1.2 Què es permet dipositar?

- Es permet el dipòsit i publicació de dades de recerca. S'entén per dades de recerca, l'evidència no teòrica que fonamenta les conclusions derivades d'un treball d'investigació i que es poden utilitzar per validar aquestes conclusions. Es pot tractar d'informació quantitativa o qualitativa recollida durant el desenvolupament d'una investigació mitjançant l'experimentació, l'observació, el modelatge, l'entrevista, informacions extretes d'altres fonts (de les quals es poden retenir els drets) o d'altres mètodes. Les dades poden ser primàries o derivades de dades primàries.
- Qualsevol format de fitxer és acceptat. No obstant això, s'aconsella proporcionar dades en formats no propietaris per tal d'assegurar un ús més ampli i per raons de sostenibilitat

a llarg termini. Es recomanen i preserven els formats indicats a la taula: “[Formats recomanats i acceptats](#)”.

### 1.3 Què no es permet dipositar?

Els sistemes de gestió de la recerca (CRIS) són els encarregats d'incloure la informació dels conjunts de dades fets a la institució, independentment d'on s'hagin dipositat. El Portal de la Recerca de Catalunya es nodreix de la informació proporcionada pels sistemes CRIS de les diferents institucions que hi participen i és, per tant, l'instrument que fa de catàleg dels conjunts de dades del sistema de recerca de Catalunya.

- Conjunt de dades que no compleixen amb els requisits del 1.4 d'aquest reglament i dels quals només es vol disposar de la informació de les metadades per tal de tenir-ne constància.

### 1.4 Què ha de contenir un conjunt de dades?

- Un conjunt de dades ha d'incloure els fitxers que contenen les dades, les metadades associades, la documentació descriptiva, [els drets i la llicència d'ús](#).
  - La tipologia de dades i formats que han de contenir els fitxers es troba descrita en l'apartat 1.2 del [Reglament](#).
  - S'han d'introduir les [metadades mínimes obligatòries](#) seguint les directrius establertes.
  - La [documentació descriptiva](#) mínima obligatòria és un fitxer de text pla (README) que descriu el contingut del conjunt de dades.
  - La llicència ha de seguir els requeriments descrits a l'apartat 2 del [Reglament](#).
- Cada conjunt de dades ha de tenir, com a mínim, un contacte designat per tal de garantir, almenys durant el període establert de preservació, qualsevol petició relacionada amb el mateix.

### 1.5 Mida dels conjunt de dades

- Es permet el dipòsit de conjunt de dades fins a 10 GB per defecte.
- Per a conjunt de dades entre els 10GB i fins a 100GB, cal fer una petició seguint el “[Procediment per dipositar datasets de més de 10GB i fins els 100GB](#)”.
- Malgrat el límit de 100GB, en aquests estadis inicials i, de cara a guanyar experiència, s'admetran conjunt de dades de mida més gran. Aquests casos hauran tenir una petició expressa signada per part de la universitat pel vicerector de recerca de la universitat, o qui delegui, o per part dels centres de recerca pel director científic, o qui faci les seves

funcions; així com, el vistiplau de la Direcció General del CSUC seguint el “[Procediment per depositar datasets de més de 100GB](#)”.

- En relació al punt 2 i 3, en el cas que algun dels fitxers que contingui el conjunt de dades superi els 10 GB, cal comunicar-ho al CSUC per a que es modifiqui el límit durant un període acordat entre institucions. Un cop transcorregut aquest temps, l’RDR tornarà al límit estàndard per carregar fitxers fins a 10GB.

## 2. Llicències

---

Quan es publica un conjunt de dades a CORA.RDR, cal definir una [llicència](#) per tal d’informar als usuaris de les condicions de reutilització de les dades. Es recomana l’ús de les llicències [Creative Commons](#) (CC) CC0 i CC-BY seguint el lema de la Unió Europea "Tan obert com sigui possible i tan tancat com sigui necessari" per millorar el descobriment, la comprensió i la reutilització de les dades de recerca.

- La persona que fa el dipòsit ha de seleccionar una llicència d’una llista predefinida. CORA.RDR ofereix una gamma completa de llicències Creative Commons, adoptant la CC0 (Creative Commons Zero) com a llicència definida per defecte.
- La persona que fa el dipòsit pot definir una llicència diferent a les Creative Commons, seleccionant “Custom Dataset Terms”. En aquest cas, cal que descrigui les condicions d’ús de les dades, segons la llicència escollida.
  - Es poden utilitzar llicències específiques per dades de recerca i bases de dades, com les que ofereix [Open Data Commons](#): ODC-BY, ODbL i PDDL.
  - També es poden aplicar llicències específiques per a programari, aprovades per la [Open Source Initiative](#) (OSI): [GNU General Public Licenses](#), [European Union Public License](#) i [MIT license](#), entre altres.
- Les metadades dels conjunt de dades publicats a CORA.RDR estan accessibles mitjançant el protocol OAI-PMH i subjectes a la [llicència CC0](#).

## 3. Nivell d’obertura dels conjunt de dades

---

- *Dades en obert*. No hi ha restriccions en el seu accés, cosa que permet que qualsevol persona pugui veure-les i descarregar-les. Aquesta és la opció més adequada per publicar dades que no són confidencials, sensibles, comercials o amb altres restriccions legals.
- *Dades embargades*. Hi ha restriccions temporals en el seu accés. Aquest embargament podrà aplicar-se a la totalitat o part del conjunt de dades, perquè no siguin accessibles

durant un termini màxim de dos anys. Tot i així, les metadades del conjunt de dades sempre seran visibles.

- *Dades restringides.* Hi ha restriccions en el seu accés. La persona investigadora i/o institució pot considerar restringir les dades de recerca sempre que sigui per raons de copyright, temes ètics o de confidencialitat. Cal subratllar que el principi bàsic de les dades de la investigació finançada amb fons públics és que les dades han de ser tan accessibles com sigui possible i estar tan protegides com sigui necessari. Tot i així, les metadades del conjunt de dades sempre seran visibles.

## 4. Donar de baixa un conjunt de dades

---

Per retirar un conjunt de dades publicat al repositori cal donar-lo de baixa. Això suposarà que deixarà de ser trobable, però es podrà consultar, per mitjà del DOI, la citació del conjunt de dades i el motiu pel qual s'ha donat de baixa. En cap cas, es podrà eliminar el DOI, ja que és un identificador registrat a [DataCite](#) i ha de ser permanent en el temps.

- No es permet eliminar un conjunt de dades publicat però es pot donar de baixa. Aquesta acció només ha de produir-se si hi ha una raó legal i/o vàlida.
- Per donar de baixa un conjunt de dades cal posar-se en [contacte](#) amb la persona responsable de la instància indicant-ne els motius. La institució, seguint les seves pròpies consideracions, farà efectiva o no la baixa.
- Els conjunt de dades en esborrany que faci més de 12 mesos que es troben dipositats al repositori són susceptibles a eliminar-se. Abans d'aquesta data límit, la persona responsable de la instància contactarà amb qui consti a l'apartat de contacte del registre per valorar l'estat del conjunt de dades i dur a terme l'acció pertinent.

## 5. Dades personals

---

CORA.RDR no accepta actualment:

- *Dades personals.* Les dades de caràcter personal que identifiquen o poden identificar persones concretes de manera directa o indirecta no es poden introduir al CORA.RDR. Aquestes dades necessiten una especial protecció i normes específiques per evitar els riscos derivats del seu tractament. Les dades de caràcter personal que estiguin anonimitzades sí que s'hi poden dipositar.

## **6. Definicions**

---

*(en desenvolupament)*