

PROCEDIMENT DE PUBLICACIÓ DE DADES D'INVESTIGACIÓ AL DATAVERSE PER A INVESTIGADORS

1. OBJECTIU

En aquest document s'inclou el procediment que qualsevol investigador/a amb la necessitat de publicar les dades de recerca pot seguir com a pauta per a fer-ho. Aquest pot incloure diverses opcions possibles per a cada punt així com eines útils per a poder fer un una millor tasca.

2. ÀMBIT D'APLICACIÓ

Aquest procediment s'aplica a qualsevol investigador de la UB amb la necessitat de publicar les dades de la seva recerca.

3. ÀMBIT DE DIFUSIÓ

Aquest procediment s'ha de difondre a qualsevol persona que necessiti autoformació per a la publicació de dades d'investigació de la UB.

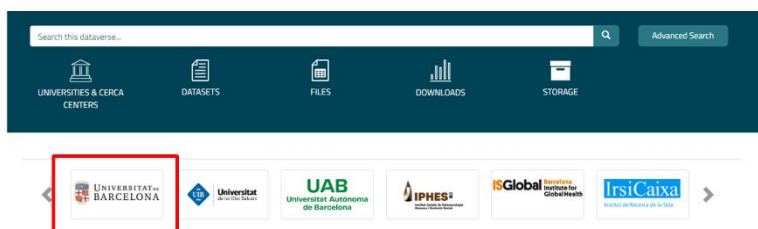
4. PROCESSOS RELACIONATS

Aquest procediment està relacionat amb el procés "Gestionar recursos d'informació" i amb el subprocés "Processar i Descriure".

5. PROCEDIMENT OPERATIU

5.1 Primer pas a l'hora de publicar dades al repositori CORA.Repositori de dades de Recerca

- a) Entrem a la web <https://dataverse.csuc.cat/> i ens creem un usuari amb el correu UB. Podeu seguir el següent procediment creat pel CSUC (<https://confluence.csuc.cat/pages/viewpage.action?pagelId=97231135>)
- b) Se'ns assignarà un rol per a poder crear datasets amb les dades de recerca, però no podrem publicar-les. Això és responsabilitat del tècnic en gestió de dades de la UB. Per tant, un cop haguem posat les dades i la informació necessària, haurem de contactar amb aquest. Per a crear el data set haurem d'anar a la instància on vulguem publicar. Si no tens una instància (o dataverse) específic pel teu projecte/grup de recerca hauràs d'anar a la instància de la Universitat de Barcelona:



5.2 Com poso de forma correcta les dades?

- a) En primer lloc sempre busquem un format adient. Al següent URL del CSUC trobarem una taula on veurem per a cada cas les opcions correctes.

URL: <https://confluence.csuc.cat/display/RDM/Recomanacions+de+formats>

- b) Amb les dades, sempre ha d'haver-hi una llegenda o explicació de què és cada dada. Per exemple, si utilitzem dades numèriques en un CSV, haurem d'explicar què és cada columna i fila, les magnituds que s'utilitzen... Això es pot fer en el mateix document o afegint al dataset un "Codebook" amb totes les explicacions necessàries. Aquest procés facilita la reutilització de les dades.

5.3 Com poso de forma correcta les metadades?

Per tal de fer que les dades siguin més fàcilment trobables afegirem una sèrie de metadades, és a dir, informació sobre les dades en qüestió. Quanta més informació proporcionem sobre les dades, millor.

Aquesta informació ha d'incloure obligatòriament els següents apartats:

- Títol senzill, que descrigui les dades (i no l'article per al que s'utilitzen) i que s'entengui.
- Descripció de les pròpies dades
- Autors (novament de la gent que ha participat en la presa de les dades, no de l'article, paper...) amb el seu ORCID. Sempre un d'ells haurà d'estar vinculat a la UB.
- Persona de contacte
- "Keywords" de les dades
- Descripció de com estan organitzades aquestes dades.

Per complementar la informació d'aquestes dades podem afegir altres metadades (Localització, d'on ve el pressupost o beca...).

El CRAI es fa càrrec d'omplir altres metadades com el "Topic Classification".

Per al "Topic Classification" utilitzarem la web del Thesaurus UB (<https://vocabulary.crai.ub.edu/ca/thub>). Allà trobarem el vocabulari general per a la definició de cada data set. Altre cop, per a fer aquesta tria, ens hem de fixar en les dades i no en l'article o paper al que van relacionades. S'han de triar d'entre 2 i 6 paraules.

5.4 Decisió de la llicència per a les dades d'investigació

S'ha de decidir el "Data Use Agreement" que preferiu per a les dades.

Des del CRAI s'aposta per a la ciència oberta, és a dir que es recomana l'ús de llicències que permetin adaptar i compartir les dades, sempre i que es doni crèdit a l'autor de les dades.

Una eina útil per decidir la llicència és la següent: <https://ufal.github.io/public-license-selector/>

5.5 Creació del Readme.txt

Un cop plenes les metadades i escollida la llicència, procedim a crear el README.txt. Aquest és un document amb informació general, amb la següent estructura:

```
Fitxer Edició Format Visualització Ajuda
DD-MM-YYYY
Contacte
Mail Contacte

GENERAL INFORMATION
-----
Dataset title:
    Títol

Authorship:
    Nom Investigador/a 1
    Departament Investigador/a 1
    ORCID: 0000-xxxx-xxxx-xxxx

    Nom Investigador/a 2
    Departament Investigador/a 2
    ORCID: 0000-xxxx-xxxx-xxxx

DESCRIPTION
-----
Descripció de les dades
Keywords: Keyword1, keyword2...
Data collected: Month day & day, Year

ACCESS INFORMATION
-----
Dataset licensed under a "Llicència escollida" (text available at https://creativecommons.org/licenses/Llicencia_escollida)
Dataset DOI: https://doi.org/xxxxxxx

FILE OVERVIEW
-----
Descripció de l'organització de les dades
Descripció dels paràmetres amb els que es classifiquen les dades, si escau
```

També pots trobar plantilles en diferents idiomes a:

<https://confluence.csuc.cat/pages/viewpage.action?pageId=104529933>

El CSUC disposa d'uns *scripts* que automatitzen el procés de creació del Readme:

https://colab.research.google.com/github/CSUC/RDR-scripts/blob/main/create_Readme/create_Readme_script.ipynb

5.6 Creació del link privat

Pots generar un enllaç per compartir aquesta versió no publicada del conjunt de dades amb altres persones, fins i tot si no tenen un compte al repositori. L'enllaç deixarà de funcionar quan el conjunt de dades es publiqui o si es desactiva manualment. Només es pot tenir una URL de previsualització activa per a cada conjunt de dades.

- Previsualització completa: Crea un enllaç perquè altres puguin revisar aquesta versió abans de la seva publicació. Tindran accés a tots els fitxers i podran veure totes les metadades, incloent-hi les que identifiquen l'autoria.
- Previsualització anònima: Crea un enllaç perquè altres accedeixin a una versió anonimitzada d'aquesta previsualització. Les metadades que podrien revelar l'autoria o el nom de l'espai no es mostraran. Els fitxers només seran accessibles si no tenen restriccions.

5.7 Publicació de les dades

Un cop estiguem conformes amb el nostre dataset, ens posarem en contacte amb l'oficina de suport a la recerca del CRAI per a que facin la revisió i procedeixin a publicar les dades.

5.8 Noves versions

Un cop publicat, si es volen fer canvis al data set, s'aniran creant noves versions cada cop.