

Pla de treball 2022 per CORA. Repositori de Dades de Recerca

(Doc.CO22/01) (4 RDM\Repositori de dades\Objectius - PlaTreball MemEco\2201 PlaTreball2022.docx,31.02.22)

1. Antecedents

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, 24 institucions (11 universitats¹ i 13 centres de recerca CERCA²) l'usen de manera desigual. També des de llavors ha estat en funcionament un Grup de treball pilot per a la publicació de dades de recerca que van valorar de manera positiva CORA.RDR, tant la seva existència, com les facilitats que proporciona (*Doc.CO21/22; CORA. Repositori de Dades de Recerca Opinions sobre el seu ús i funcionament per part de les institucions participants en el grup pilot març-juny 2021*).

En el *Doc.CO21/04; Pla de treball 2021 pel repositori de dades de recerca* es marcava el full de ruta dels treballs a realitzar a CORA. RDR. Això va permetre progressar paulatinament en algunes qüestions que eren importants per al correcte desenvolupament del repositori, així com, per a les institucions que l'usaven. Per aquest motiu, s'ha considerat elaborar aquest document amb un pla de treball per al 2022 que es desenvolupa i detalla en aquest document.

2. Pla de treball per al 2022

A continuació es detallen les principals accions a dur a terme al 2022 per quadrimestres.

¹ UB, UAB, UPC, UPF, UdG, UdL, URV, UOC, UVic-UCC, URL, UIC

² IRTA, CRM, IDIBELL, VHIR, IFAE, CRAG, IrsiCAIXA, ICAC, IJC, IBEL, IdibGI, ICFO, IRB Barcelona

Primer quadrimestre (gener-abril)

- **Elaborar i signar un conveni amb cada institució**

El 2021 es va elaborar un conveni que es va signar per diferents institucions, totes elles universitats, que determina els drets i deures de cadascuna de les parts. A desembre de 2021 es disposava de 5 convenis signats.

Cal finalitzar aquesta tasca amb les universitats restants i els centres de recerca que s'han incorporat.

- **Tenir un reglament general del repositori i uns procediments**

Associat al conveni, cal disposar d'un conjunt de regles a complir al CORA. RDR. Aquestes normes generals haurien d'anar acompanyades dels procediments que expliquin com cal complir-les.

Iniciar un reglament en base els acords de funcionament que es vagi prenent

- **Traduir al català el repositori**

Durant el 2021 s'ha estat treballant en la traducció del repositori (tant a la part d'usuari, com la d'administració) i, actualment, està en procés de revisió.

Cal finalitzar aquesta tasca per oferir una plataforma multilingüe.

- **Procedimentar accions ja experimentades**

Durant el 2021 es van fer proves amb l'exportació i importació en el CORA.RDR. Pel que fa a l'exportació es va provar en metadades Dublin Core per a que puguin ser ingestades en els sistemes d'informació i/o repositoris institucionals de les universitats i centres de recerca a través del protocol OAI-PMH. D'altra banda, es va importar un conjunt de 48 datasets del repositori institucional de la Universitat de Barcelona.

Cal procedimentar:

- Els processos d'exportació i importació per a que la resta d'institucions puguin dur-los a terme.
- El dipòsit de datasets de mida gran. Aquest procés ja ha estat provant durant el 2021 però cal finalitzar aquesta acció per a que les institucions puguin fer-ho.
- Les accions que permetin poder importar datasets dipositats, a hores d'ara, a Zenodo cap a CORA. RDR.

- **Autenticació**

Les institucions participants al pilot del repositori han manifestat reiterades vegades la necessitat de disposar d'un mètode que permeti als usuaris procedir a una única autenticació (single sign-on).

Atès que el sistema UNIFICAT està estès entre les universitats i els centres de recerca i s'ofereix des del CSUC cal prioritzar aquesta funcionalitat.

- **Definir metadades mínimes**

Durant el 2021 s'ha treballat en la definició d'unes metadades mínimes obligatòries, no només per estandarditzar les metadades entre instàncies sinó per permetre una major exportabilitat d'aquestes.

Cal acordar un conjunt de metadades mínimes.

- **Assignar identificadors persistents (DOIs)**

En l'actualitat CORA. RDR assigna automàticament un DOI a cada arxiu que hi ha dintre d'un dataset però hi ha datasets amb molts arxius que consumeixen molts DOI, que són de pagament.

Cal definir el procés d'un únic DOI per dataset i, en el cas que se'n necessitin més, procedimentar com dur a terme una petició especial.

- **Preservació amb Madroño**

Durant el 2021 s'han mantingut converses conjuntament amb el Consorci Madroño per a disposar còpies creuades dels datasets d'ambdós repositoris a fi i efecte de preservar el seu contingut. Actualment, es disposa d'un esborrany de conveni que caldria signar per ambdues parts per iniciar els treballs tècnics.

A la vegada, un cop signat el conveni amb el Consorci Madroño, cal iniciar els treballs tècnics per preservar els datasets. Caldrà definir els processos per dur a terme aquesta tasca i la periodicitat en que es farà.

- **Definir com s'embarquen/restringeixen els datasets**

L'accés a alguns data sets ha de quedar embargat fins a una data determinada, per exemple, fins que l'article al qual s'associa el data set es publiqui. Avui en dia, hi ha diverses maneres d'embargar però no s'ha establert amb el grup pilot quina és la més adequada.

Cal consensuar els supòsits i els processos per a tothom l'abans possible, per tal d'unificar criteris.

Segon quadrimestre (maig-agost)

- **Revisar el procés de curació de dades (amb accions mínimes a fer)**

El 2020 es va presentar el document "REVISAT: Criteris orientadors per curar conjunts de dades de manera FAIR" que donava una sèrie de checklists amb els processos que cal que un curador de dades porti a terme per tal de millorar la qualitat d'un dataset.

Caldrà tornar a revisar el document per veure si s'adequa a les accions que s'estan duent a terme a les institucions o cal modificar-lo. Cal recordar que curar les dades serà bàsic a fi i

efecte que el repositori assoleixi uns mínims de qualitat i pugui optar a rebre un segell de qualitat.

- **Interoperabilitat: OpenAIRE**

Una vegada s'hagin acordat les metadades mínimes al primer quadrimestre, cal començar a exportar les dades a d'altres eines de descoberta; preferentment, a OpenAIRE, el projecte europeu que té la missió de donar suport a la ciència oberta a Europa.

Per això cal iniciar els treballs per mapejar les metadades de CORA. RDR amb les que es requereixen en aquest portal.

- **Repositoris disciplinaris**

També caldrà estudiar i valorar la relació que CORA. RDR pugui tenir amb d'altres repositoris de dades; per exemple, recollir tota la producció d'investigadors catalans que es troben dipositats en repositoris disciplinaris.

Tècnicament és possible importar registres de metadades d'altres repositoris però cal provar-ho i, decidir, a nivell polític quin serà el reglament d'aquests casos.

- **Certificació del segell CoreTrustSeal**

Per tal d'oferir confiança i seguretat als investigadors, cal disposar d'un segell de certificació del repositori com el de CoreTrustSeal.

Cal iniciar els treballs per a certificar el CORA. RDR com a repositori de confiança.

- **Contacte amb nous usuaris**

Una vegada s'hagin iniciat els tràmits per a que totes les institucions que ja utilitzen el CORA. RDR tinguin el conveni signat, caldria començar a contactar amb noves institucions.

A més a més de mantenir converses inicials, caldrà preparar sessions informatives/demostratives del repositori, així com de formació a l'hora d'usar-lo.

Tercer quadrimestre (setembre-desembre)

- **Interoperabilitat: PRC (i EOSC, si fos el cas)**

A mesura que es vagin exportant metadades a diferents sistemes d'informació, és necessari que el Portal de la Recerca de Catalunya també mostri els datasets de les institucions participants a CORA. RDR. L'escenari òptim seria que les pròpies institucions enviessin aquesta informació des dels seus CRIS però, ara per ara, la majoria no estan preparades per fer-ho, així com el PRC tampoc està preparat per incloure-hi aquesta informació.

Cal iniciar els treballs des del PRC per a que pugui incloure aquesta nova metadada en el seu portal. I, de la mateixa manera, dissenyar un procediment que permeti exportar les metadades de CORA.RDR i, incloure-les, al PRC. Paral·lelament a tot això, el CSUC forma part com a

membre observador a l'EOSC. Per aquest motiu, cal tenir en compte tot el que es dicti des d'aquesta plataforma per a implantar-ho també en el CORA. RDR.

- **Fer tasques de preservació**

En el cas que es disposi del personal de curació de dades, aquest pot iniciar els estudis i treballs sobre quines tècniques es poden aplicar per garantir una correcta preservació de les dades.

Durant tot l'any

- **Continuar les reunions i el contacte amb el grup**

Des de l'inici del projecte i durant el 2021 s'han mantingut diferents reunions per anar progressant en el desenvolupament i regulació del repositori. Durant aquest 2022 caldrà programar reunions més o menys periòdiques, o en el seu defecte, a través de correus electrònics, per seguir treballant.

- **Assessorament amb casos pràctics**

Durant el 2021 s'han mantingut diferents sessions d'assessorament per permetien als usuaris poder millorar la curació les dades que anaven dipositant al repositori. A més a més, en l'últim trimestre de l'any, s'hi va sumar la presentació per part dels usuaris mateixos, de com curaven i dipositaven els seus datasets. Se seguiran fent aquest tipus d'accions.

- **Formació**

De formació se n'ha fet a diferents nivells i per diferents grups d'usuaris. Caldrà seguir fent aquesta tasca.

A més a més, seria necessari crear uns materials de formació introductoris per aquelles noves institucions que vulguin iniciar-se en els serveis de gestió de dades de recerca (CORA. RDR, eina DMP, etc.)

- **Revisar i unificar les eines de suport**

Actualment es disposa de diferents eines de suport que els investigadors i administradors poden usar a l'hora de conèixer com cal fer cada acció dins el repositori: des de la pàgina d'"About" al CORA. RDR, el Confluence de Gestió de Dades de Recerca o a la Web del CSUC.

Cal iniciar l'anàlisi i treballar per a que tot estigui unificat i segueixi el mateix criteri.

Calendarització de les tasques

Tasca	Gener	Febrer	Març	Abril	Maig	Juny	Juliol	Agost	Setembre	Octubre	Novembre	Desembre
Q1: Elaborar i signar un conveni amb cada institució												
Q1: Tenir un reglament general del repositori i uns procediments												
Q1: Traduir al català el repositori												
Q1: Procedimentar accions ja experimentades												
Q1: Autenticació												
Q1: Definir metadades mínimes												
Q1: Assignar identificadors persistents (DOIs)												
Q1: Preservació amb Madroño												
Q1: Definir com s'embarguen/restringeixen els datasets												
Q2: Revisar el procés de curació de dades												
Q2: Interoperabilitat: OpenAIRE												
Q2: Certificació del segell CoreTrustSeal												
Q2: Repositoris disciplinaris												
Q2: Contacte amb nous usuaris												
Q3: Interoperabilitat: PRC (i EOSC, si fos el cas)												
Q3. Fer tasques de preservació												
TL: Continuar les reunions i el contacte amb el grup												
TL: Assessorament amb casos pràctics												
TL: Formació												
TL: Revisar i unificar les eines de suport												

Q1: Quadrimestre 1

Q2: Quadrimestre 2

Q3: Quadrimestre 3

TL: Tot l'any