



Consorti de
Serveis Universitaris
de Catalunya

Requisitos funcionales para la creación de repositorios consorciados de datos de investigación

Ricard de la Vega
Mireia Alcalá

XXX Jornadas Técnicas de RedIris
Sevilla, 29 de mayo de 2019

Índice

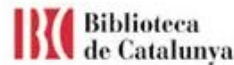
- El CSUC y la ciencia abierta
- El servicio de soporte a la gestión de datos de investigación (RDM)
- Requisitos para la creación de un repositorio consorciado de datos de investigación
 - Determinantes de contexto
 - Requisitos y buenas prácticas
- Conclusión

EL CSUC

Entidades consorciadas:



Entidades asociadas:



El CSUC

Cálculo
científico

Comunicaciones

Gestión
documentación
científica

Contrataciones

Infraestructuras
TIC

Compras
conjuntas

e-Administración

El CSUC y la Ciencia Abierta

- Desde **1999** trabaja para la promoción y gestión de diferentes actividades relacionadas con la investigación:
 - Creación de repositorios y portales cooperativos
 - Tesis Doctorales en Red (TDX, 30.000 tesis)
 - Revistas Catalanas en Acceso Abierto (RACO, 500 revistas)
 - ...
 - Cumplimiento de mandatos de Acceso Abierto
 - Portal de la Investigación de Catalunya
- En 2014 se crea una línea estratégica de actuaciones de apoyo a la investigación
- En 2017 se crea una nueva área organizativa: Ciencia Abierta

Área de Ciencia Abierta del CSUC

- **Misión:** Colaborar con las universidades para disminuir el esfuerzo de adaptación a los requerimientos de la Ciencia Abierta
- **Principios rectores:**
 - Explorar conjuntamente
 - Compartir buenas prácticas
 - Construir soluciones conjuntas (si es el caso)



Cerca per...

A tot el portal



Universitats i Centres de recerca	Departaments, Instituts i Àrees	Projectes de recerca	Grups de recerca	Investigadors	Publicacions	Tesis
21	399	35.882	1.552	11.879	539.430	30.725

9.700 projectes d'excel·lència finançats per l'ERC des de l'any 2007

L'European Research Council (ERC) ha presentat la publicació Annual Report on the ERC activities and achievements in 2018. La missió de l'ERC és impulsar les fronteres del coneixement generat per la recerca europea. Per això, es segueix l'estratègia de proporcionar finançament [...]

Darreres publicacions i tesis

Continuing professional education in engineering faculties: transversal integration of sustainable human development in basic engineering sciences courses

Lazzarini, B.
 2019

Multifunctional ZnO NPs-chitosan-gallic acid hybrid nanocoating to overcome contact lenses associated conditions and discomfort

Hoyo, J.
 2019

Impact and compression after impact response in thin laminates of spread-tow woven and non-crimp fabrics

Trias Mansilla, Daniel
 2019

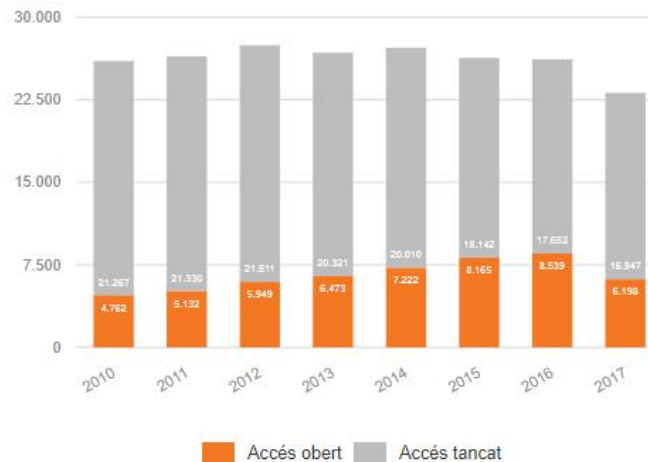
Case-base Maintenance of a Personalised and Adaptive CBR Bolus Insulin Recommender System for Type 1 Diabetes

Massana Raurich, Joaquim
 2019

Beyond music: Emotional skills and its development in young adults in choirs and bands

Publicacions en accés obert i tancat

La mitjana de publicacions en accés obert entre el 2010 i el 2017 és del 25%



Basat en les dades de setembre 2018

Índice

- El CSUC y la ciencia abierta
- El servicio de soporte a la gestión de datos de investigación (RDM)
- Requisitos para la creación de un repositorio consorciado de datos de investigación
 - Determinantes de contexto
 - Requisitos y buenas prácticas
- Conclusión

Servicio de soporte a RDM

- El grupo de trabajo de apoyo a la investigación se creó en 2015 y fruto de diferentes acciones cooperativas, las universidades catalanas **iniciaron el servicio de soporte a la gestión de datos en 2016.**
 - Ayuda para la confección de planes de gestión de datos
 - Recomendaciones para elaborar políticas de gestión de datos en las universidades
 - Acciones de difusión y formación
 - ...
- Realizadas 2 encuestas a investigadores de las universidades catalanas para conocer sus necesidades en gestión de datos
- Resumen de la situación
 - Tenemos más oferta de servicios que no demanda
 - A nivel mundial hay todavía pocas iniciativas, y las que hay son incipientes
 - Sin embargo, **es necesario y estratégico (pero aún no urgente) un buen servicio de gestión de datos, probablemente lleguemos por aproximaciones sucesivas**

Aproximación asimétrica a RDM

Tamaño y
complejidad

1. Planes de gestión de datos (DMP)
2. Recomendaciones para seleccionar un repositorio existente
3. Ampliación repositorios institucionales
4. Creación de un repositorio FAIR consorciado
5. Ampliación para la gestión de todo el ciclo de vida de los datos

Aproximación asimétrica a RDM

Tamaño y
complejidad

1. Planes de gestión de datos (DMP)
2. Recomendaciones para seleccionar un repositorio existente
3. Ampliación repositorios institucionales
4. Creación de un repositorio FAIR consorciado
5. Ampliación para la gestión de todo el ciclo de vida de los datos

Welcome

This tool helps you to create, review, and share data management plans that meet institutional and funder requirements.

Join the growing number of researchers that have adopted Research data management plan:



235 Users



73 Plans



11 Organisations

You can download funder templates without logging in, but from this tool you will find tailored guidance from universities. Why not sign up for an account and try it out?

Sign in

Create account

* Email

* Password

[Forgot password?](#)

Remember email

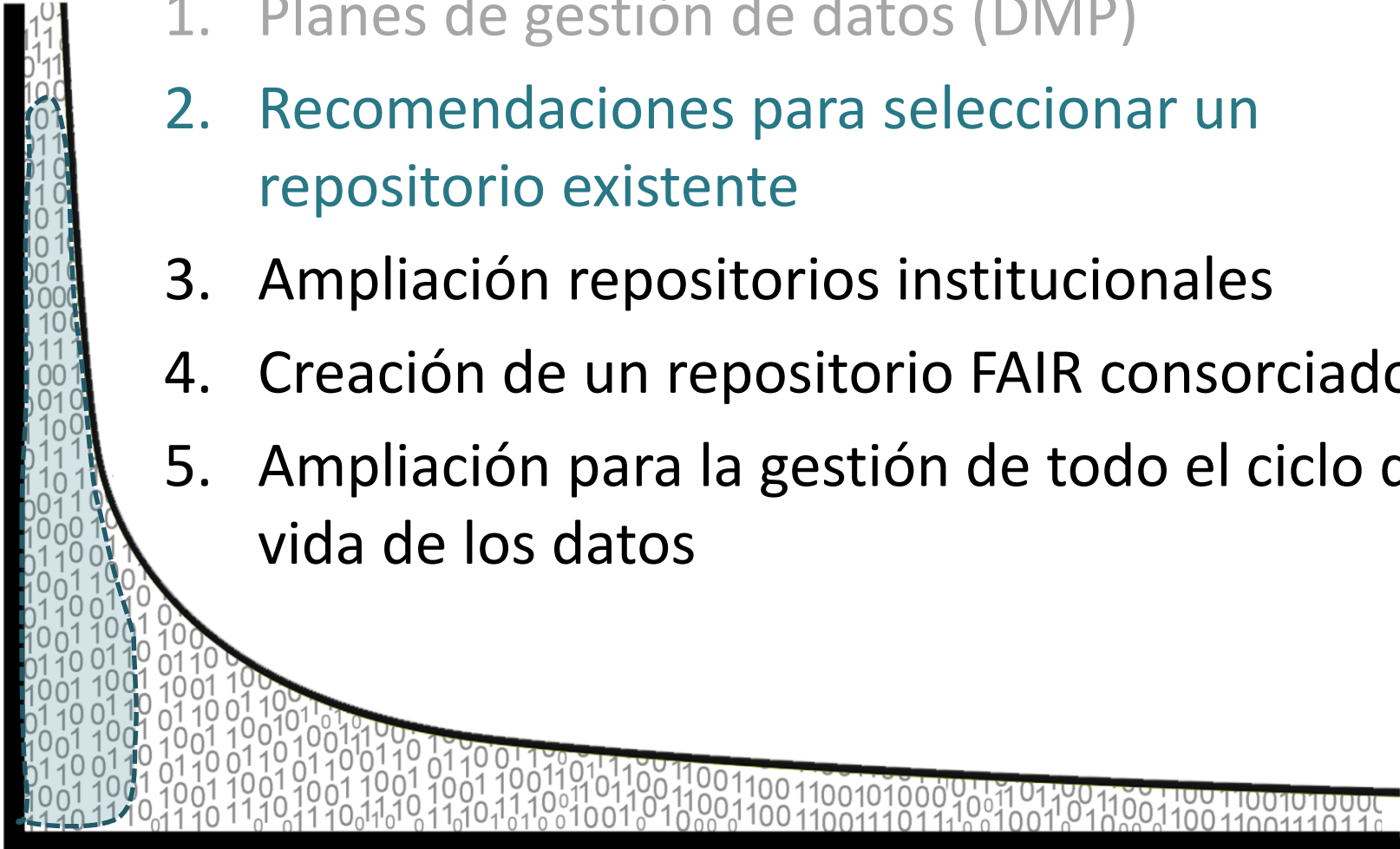
Sign in









10 simple rules for creating a good data management plan

Aproximación asimétrica a RDM

Tamaño y
complejidad



1. Planes de gestión de datos (DMP)
2. Recomendaciones para seleccionar un repositorio existente
3. Ampliación repositorios institucionales
4. Creación de un repositorio FAIR consorciado
5. Ampliación para la gestión de todo el ciclo de vida de los datos

	 CERN	 EUDAT	 Digital Science (Macmillan)	 NCSU	 IOSS (Harvard University)	 Elsevier
1. Àmbit temàtic i geogràfic	Multidisciplinar Internacional	Multidisciplinar Internacional	Multidisciplinar Internacional	Amb èmfasi en ciència i medicina Internacional	Multidisciplinar Internacional	Multidisciplinar Internacional
2. Tipologia de dades (programari, imatges, raw data, etc.)	Tot tipus de formats i també documents publicats	Tot tipus de formats	Tot tipus de formats	Tot tipus de formats (amb exclusió de software) i amb preferència per formats de codi obert	Tot tipus de formats amb funcionalitats addicionals per alguns (RData, SPSS, STATA, CSV, xls ^x , FITS)	Tot tipus de formats
3. Mida aproximada dels fitxers a incloure	Fins a 50 GB Per a mides superiors es pot contactar amb ells	Fins a 2 GB. Per a mides superiors, contactar amb ells	Fins a 5 GB	20 GB per un conjunt de dades. Per a mides superiors revisar les condicions	Fins a 2 GB.	10 GB
4. Dades obertes, embargades, restringides o tancades	Obertes Tancades Restringides Embargades (Polícies)	Obertes Restringides Embargades	Obertes Tancades (per defecte fins que no es publiquen) Restringides Embargades	Obertes Embargades (per defecte fins el moment de la publicació)	Obertes Restringides	Obertes Tancades (per defecte fins que no es publiquen) Restringides Embargades
5. Llicències amb les que es volen difondre les dades	Permet varietat de llicències encara que recomanen llicències obertes. La descripció de les dades està sota CC0	Permet varietat de llicències per a dades i per a software	- CC0 per a dades (<i>metadata & datasets</i>) - CC-BY per a no dades (audiovisuals, <i>papers</i> , etc.) - MIT per a codi i software - Altres llicències: GNU, Apache 2.0, BSD 3 clause	CC0 (FAQ)	Per defecte CC0 però es pot canviar	Permet varietat de llicències: CC-BY 4.0, CC-BY-NC, CC0 1.0, MIT, Apache 2.0, BSD 3 clause, BSD 2 clause, gplv3
6. Identificador permanent	DOI o URL	Handle	DOI	DOI	DOI	DOI
7. Condicions per retirar contingut	Es permet retirar les dades justificant-ne el perquè. El DOI i la URL quedaran retinguts.	Per retirar contingut contactar amb EUDAT	Es pot sol·licitar la retirada de contingut públic justificant-ne el motiu	Cessió de dades irrevocable	Es permet retirar les dades, tot i que les metadades quedaran sempre visibles	Per retirar el contingut contactar amb Mendeley .

Aproximación asimétrica a RDM

Tamaño y
complejidad

1. Planes de gestión de datos (DMP)
2. Recomendaciones para seleccionar un repositorio existente
3. Ampliación repositorios institucionales
4. Creación de un repositorio FAIR consorciado
5. Ampliación para la gestión de todo el ciclo de vida de los datos



Aproximación asimétrica a RDM

Tamaño y
complejidad

1. Planes de gestión de datos (DMP)
2. Recomendaciones para seleccionar un repositorio existente
3. Ampliación repositorios institucionales
4. Creación de un repositorio FAIR consorciado
5. Ampliación para la gestión de todo el ciclo de vida de los datos



Aproximación asimétrica a RDM

Tamaño y
complejidad

1. Planes de gestión de datos (DMP)
2. Recomendaciones para seleccionar un repositorio existente
3. Ampliación repositorios institucionales
4. Creación de un repositorio FAIR consorciado
5. Ampliación para la gestión de todo el ciclo de vida de los datos



Índice

- El CSUC y la ciencia abierta
- El servicio de soporte a la gestión de datos de investigación (RDM)
- **Requisitos para la creación de un repositorio consorciado de datos de investigación**
 - Determinantes de contexto
 - Requisitos y buenas prácticas
- **Conclusión**

Requisitos Factibles, Alcanzables e Implementables para un Repositorio FAIR

- Inputs
 - Análisis de referentes de otros países
 - Estudio de informes técnicos
 - 25 entrevistas con expertos
- Outputs
 - 25 requisitos mínimos
 - 4 buenas prácticas
 - 3 recomendaciones



<https://recercat.cat/handle/2072/356460>

Determinantes del contexto

- Ahora o esperamos
- Definitivo o infraestructura en evolución
- Datos finales o también todo el ciclo de vida
- Aquí o fuera
- Para todas las disciplinas o para aquellas sin repositorios de datos consolidados
- Repositorio o buenas prácticas

Determinantes del contexto

- **Ahora** o esperamos
- Definitivo o infraestructura en evolución
- Datos finales o también todo el ciclo de vida
- Aquí o fuera
- Para todas las disciplinas o para aquellas sin repositorios de datos consolidados
- Repositorio o buenas prácticas

Determinantes del contexto

- **Ahora** o esperamos
- Definitivo o **infraestructura en evolución**
- Datos finales o también todo el ciclo de vida
- Aquí o fuera
- Para todas las disciplinas o para aquellas sin repositorios de datos consolidados
- Repositorio o buenas prácticas

Determinantes del contexto

- **Ahora** o esperamos
- Definitivo o **infraestructura en evolución**
- **Datos finales** o también todo el ciclo de vida
- Aquí o fuera
- Para todas las disciplinas o para aquellas sin repositorios de datos consolidados
- Repositorio o buenas prácticas

Determinantes del contexto

- **Ahora** o esperamos
- Definitivo o **infraestructura en evolución**
- **Datos finales** o también todo el ciclo de vida
- **Aquí** o fuera
- Para todas las disciplinas o para aquellas sin repositorios de datos consolidados
- Repositorio o buenas prácticas

Determinantes del contexto

- Ahora o esperamos
- Definitivo o infraestructura en evolución
- Datos finales o también todo el ciclo de vida
- Aquí o fuera
- Para todas las disciplinas o para aquellas sin repositorios de datos consolidados
- Repositorio o buenas prácticas

Determinantes del contexto

- Ahora o esperamos
- Definitivo o infraestructura en evolución
- Datos finales o también todo el ciclo de vida
- Aquí o fuera
- Para todas las disciplinas o para aquellas sin repositorios de datos consolidados
- Repositorio i buenas prácticas

Requisitos funcionales mínimos y factibles



Requisitos funcionales mínimos y factibles

1. Identificadores persistentes
2. Capacidad de almacenaje alta
3. Prestaciones medio-altas de preservación
4. Interoperabilidad con otros sistemas
5. Gestión de características especiales

Requisitos funcionales mínimos y factibles

1. Identificadores persistentes

- Asignar el DOI como identificador
- Soporte de ORCID

2. Capacidad de almacenaje alta

3. Prestaciones medio-altas de preservación

4. Interoperabilidad con otros sistemas

5. Gestión de características especiales

Requisitos funcionales mínimos y factibles

1. Identificadores persistentes
- 2. Capacidad de almacenaje alta**
 - Ficheros de hasta 10 GB por defecto
 - Espacio elástico para crecimiento
3. Prestaciones medio-altas de preservación
4. Interoperabilidad con otros sistemas
5. Gestión de características especiales

Requisitos funcionales mínimos y factibles

1. Identificadores persistentes
2. Capacidad de almacenaje alta
- 3. Prestaciones medio-altas de preservación**
 - Disponibilidad del fichero un mínimo de 10 años
 - Dos copias geográficamente distribuidas como mínimo
 - Comprobar la integridad de los datos periódicamente
 - Seguir el modelo de preservación OAIS
4. Interoperabilidad con otros sistemas
5. Gestión de características especiales

Requisitos funcionales mínimos y factibles

1. Identificadores persistentes
2. Capacidad de almacenaje alta
3. Prestaciones medio-altas de preservación
- 4. Interoperabilidad con otros sistemas**
 - Repositorios (institucionales, temáticos...)
 - Almacenamiento en la nube (Unidisc, Dropbox...)
 - Discovery tools (EOSC, PRC, OpenAire...)
 - Protocolos comunicación estándar (OAI-PMH, API...)
 - Formatos estándar de datos (XML, JSON)
5. Gestión de características especiales

Requisitos funcionales mínimos y factibles

1. Identificadores persistentes
2. Capacidad de almacenaje alta
3. Prestaciones medio-altas de preservación
4. Interoperabilidad con otros sistemas
- 5. Gestión de características especiales**
 - Diferentes versiones de un dataset
 - Gestión de diferentes esquemas de metadatos
 - Diferentes tipos de accesos

Requisitos funcionales mínimos y factibles

- Aceptación de cualquier tipo de formato
- Diferentes tipos de *submit*
- Citación recomendada
- Difusión de datasets a través de redes sociales
- Gestión de distintos tipos de licencias
- Datos analíticos de uso de la plataforma
- Suministrar metadatos para su reutilización
- Fácilmente usable
- Cumplimiento de la legislación vigente

Buenas prácticas

1. Realizar una curación de los datos
2. Seleccionar los conjuntos de datos
3. Fomentar el uso de formatos abiertos
4. Usar estándares i protocolos i vocabularios controlados ampliamente aceptados

Resumen de otros países

País	Liderazgo	Infraestructura	Estadio
Bélgica flamenca	VLIR (órgano consultivo representante de 5 universidades)	<ul style="list-style-type: none"> Elaborada una propuesta 	En estudio
Finlandia	Finnish Ministry of Education and Culture y producción de l'CSC-IT	<ul style="list-style-type: none"> Centralizada Servei segons el cicle de vida de les dades: <ul style="list-style-type: none"> Descobrir (Etsin) Almacenar (IDA) Describir (Qvain) Preservar (Fairdata-PAS) 	Inicial
Países Bajos	DANS	<ul style="list-style-type: none"> Centralizada Servei segons el cicle de vida de les dades: <ul style="list-style-type: none"> Descubrir (NARCIS) Publicar i gestionar (DataverseNL) Preservar (EASY) 	Consolidado
Portugal	Fundação para a Ciência e a Tecnologia	<ul style="list-style-type: none"> Elaborada una propuesta pero paralizada por falta de financiación 	En estudio
Suecia	Swedish Research Council y consorcio de 7 universidades y centros de investigación	<ul style="list-style-type: none"> Híbrida (repositori central SND + repositoris institucionales) Servicio de preservación en desarrollo 	Consolidado

Índice

- El CSUC y la ciencia abierta
- El servicio de soporte a la gestión de datos de investigación (RDM)
- Requisitos para la creación de un repositorio consorciado de datos de investigación
 - Determinantes de contexto
 - Requisitos y buenas prácticas
- **Conclusión**

Conclusión

Actualmente todas las universidades catalanas disponen de un servicio de gestión de datos de investigación, con las siguientes debilidades:

- Existe **distancia entre la oferta i la demanda**, quizá por:
 - El mandato europeo aún no se ha hecho efectivo
 - Desconexión entre las OTT y el servicio de soporte a la investigación en las universidades
 - Novedad del tema
- **No tenemos (aún) la infraestructura** que la CE pide para publicar los datos en forma FAIR
 - Algunas universidades usan sus repositorios institucionales. Bueno mientras no exista una solución de más recorrido.
 - La creación de un repositorio de datos es más **importante ahora para generar experiencia y buenas practicas** que por su operatividad inmediata

Conclusión

- Se **creará de manera inmediata** en el marco del CSUC y con software libre existente un repositorio cooperativo de datos de investigación que gradualmente **adopte los requisitos presentados**.
- El diseño de esta infraestructura ha de considerarse no definitivo y **en evolución** ya que ha de tener características aún no definidas (como la compatibilidad con EOSC)

Muchas gracias por su atención



Ricard de la Vega

ricard.delavega@csuc.cat

 @rdelavega



Mireia Alcalá

mireia.alcala@csuc.cat

 @apmireia