



From Open Access to Open Data collaborative work in the university libraries of Catalonia

Mireia Alcalá & Lluís Anglada (CSUC)

supported by the Research Support Working Group of CSUC

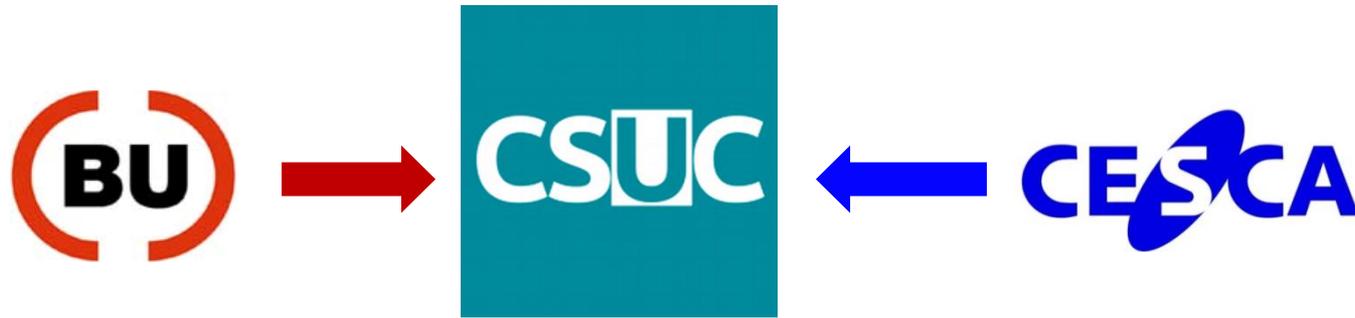
1. Who we are

2. Work done with research data

3. Work in progress with research data

4. Conclusions

1.1 New merged consortium in 2014



- for Catalan universities



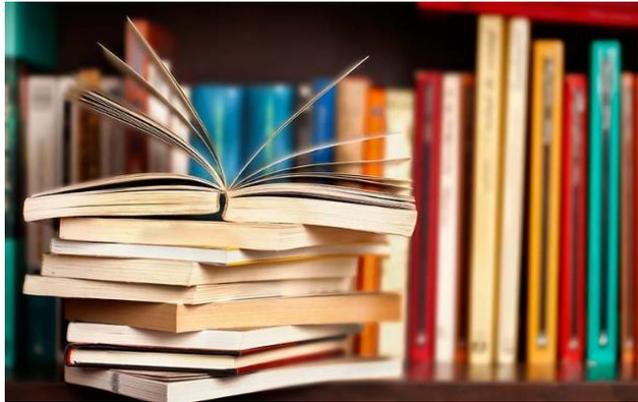
- the CBUC activities
 - Union catalogue, ILL, ILS...
- the CESCA activities
 - Network, supercomputing...
- CSUC new activities
 - join purchases (electricity, printing, cleaning, facilities, etc.)
 - common data center
 - **portal for the research output (PRC)**

1.2 Libraries evolution: from print to digital

from an **object** centred organization and services



to a **process** centred organization and services



1.3 Evolution of Catalan university libraries activities and organization supporting research

- Traditional activities supporting research:
 - Collection buildings
 - Acquisitions, Journals collections, Special collections...
 - The big change of the networked digital information
 - The content is on your PC, tablet, mobile...
 - New tools and rules are completely strange to the majority of researchers
 - Identifiers, OA, mandates, altmetrics...
 - Open Science as the new landscape
- Our evolution supporting research
 - Since 2005 (approx.), Catalan university libraries made changes in their services and organization to adapt to the new situation
 - subject librarians, reference management tools, open access...
 - In 2013, CBUC had also changed and created a new strategic line for supporting research
 - repositories, Mendeley, portal for Catalan research, OA...
 - but Internet is disruptive, and in 2017 (Feb) we decided to be more radical and create a new (small) area devoted to 'Open Science'
 - open metadata (portal for Catalan research), open access, **open data...**

1.4 The CSUC and research

OPEN SCIENCE AREA

Commission of vice rector for research

Research support working group

Commission of Research Portal



UNIVERSITAT DE BARCELONA

UAB

Universitat Autònoma de Barcelona



UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH



Universitat Pompeu Fabra Barcelona



Universitat de Lleida



UNIVERSITAT ROVIRA I VIRGILI



Universitat Oberta de Catalunya

UVIC

UNIVERSITAT DE VIC UNIVERSITAT CENTRAL DE CATALUNYA

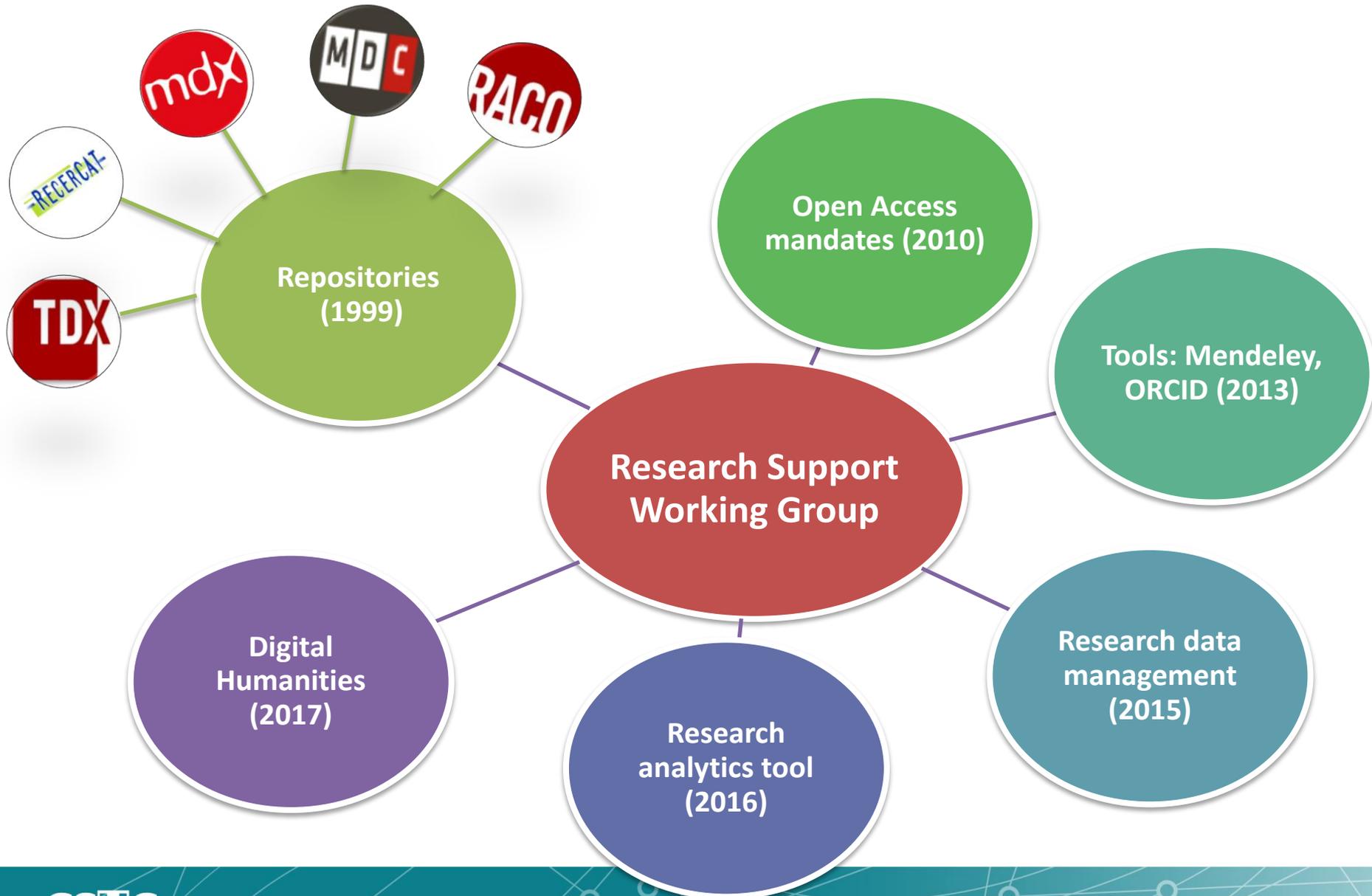


UNIVERSITAT RAMON LLULL



UNIVERSITAT JAUME I

1.4 Research Support Working Group

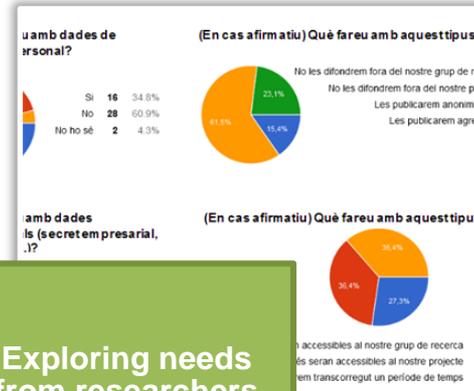


2.1 Work done with the research data

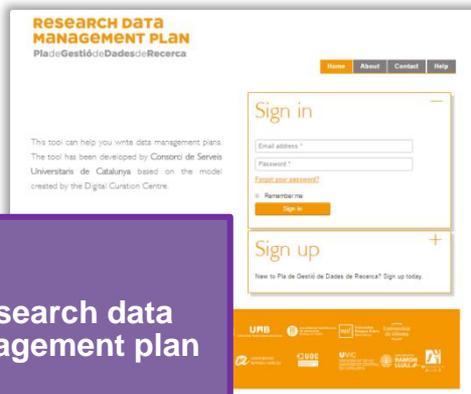
- In 2015 the working group decided to focus on RDM:



Training



Exploring needs from researchers



Research data management plan



Recommendations on data repositories



Framework agreement of data policy

2.2 Training and exploring researchers needs

- Training
 - At CSUC level, inviting experts from other countries
 - At university level, organizing staff training courses



- Survey
 - We sent a survey to the leaders of the H2020 projects in the Catalan universities (included or not in the Open Research Data Pilot)
 - Aim: to know their needs RDMResults: nothing clear (due to our inexperience in RDM but also - probably- because the researchers don't know exactly what they have to do)
 - Report is available at RECERCAT and the associated data at ZENODO (<http://hdl.handle.net/2072/268186> & <https://doi.org/10.5281/zenodo.183129>)

2.3 Data Management Plans

- Guide on what a DMP should include according to the guidelines of the H2020 framework programme, which includes the FAIR principles and is accompanied by selected examples from several real DMPs
- This guide is available in text format in English (hdl.handle.net/2072/266523) and Catalan (hdl.handle.net/2072/266502)
- Also, in the online tool “Research Data Management Plan” (www.dmp.csuc.cat), an adaptation of the DMPOnline tool of the Digital Curation Centre

The screenshot shows the homepage of the Research Data Management Plan (RDMP) tool. The header features the title "RESEARCH DATA MANAGEMENT PLAN" in orange and black, with the subtitle "Pla de Gestió de Dades de Recerca" below it. A navigation menu includes "Home", "About", "Contact", and "Help". The main content area has a "Sign in" section with a text box for "Email address *", a text box for "Password *", a "Forgot your password?" link, and a "Remember me" checkbox. Below this is a "Sign up" section with a plus sign icon and the text "New to Pla de Gestió de Dades de Recerca? Sign up today." The footer is orange and contains logos for "Powered by: DMP ONLINE" and "Coordination: CSUC", along with logos for various Catalan universities: UB (Universitat de Barcelona), URV (Universitat de Regensburg), UR (Universitat Politècnica de Catalunya), UPF (Universitat Pompeu Fabra), UdG (Universitat de Girona), UJI (Universitat Jaume I), UAB (Universitat Autònoma de Barcelona), UDL (Universitat de Lleida), URV (Universitat Rovira i Virgili), UOC (Universitat Oberta de Catalunya), UVIC (Universitat de Vic), and UR (Universitat Ramon Llull).

2.3 DMP – Collective template

RESEARCH DATA MANAGEMENT PLAN

PladeGestió de Recerca i Innovació Tecnològica

Signed in as Mireia Alcalá

Contact Help

Consult your university recommendations!

My plan (Home)

Plan details Initial

1. Data summary (1 question)

Provide a summary of the

B I

Save

Not answered yet

2. FAIR data (4 questions)

3. Allocation of resources (1 question, 0 answered)

4. Data security (1 question, 0 answered)

If the data are generated within the project, state the source of the data.

If the data are collected, state the source from which they were extracted.

Real example

1. *(data generated within the project)* Dataset produced by simulation tools and/or by real life trials will be used as a means to quantify the performance advantages that the project architecture offers compared with current practices.
2. *(data collected)* The data sources and solutions come from European research projects funded by the EU Commission and so have been collected from official portals, primarily those from the EC and from the project websites themselves.

2.4 DMP – Individual specifications

2. FAIR data (4 questions, 0 answered)

2.1 Making data findable, including provisions for metadata

B *I* [List Icons] [Link Icon] [Grid Icon]

Save

Not answered yet

2.2 Making data openly accessible

B *I* [List Icons] [Link Icon]

Save

Not answered yet

Guidance **Share note**

- Outline the discoverability of data (metadata provision) +
- Outline naming conventions used +
- Outline the approach towards search keywords +
- (any) +
- available +
- Specify what methods or software tools are needed to access the data +
- Specify where the data and associated metadata, documentation and code are deposited +
- Specify how access will be provided in case there are any restrictions +
- UPC specifications** +

UPC specifications

If you require more information to decide where to deposit and publish your datasets, please visit the [Publica!](#) webpage

2.5 DMP – Individual specifications

Ciutadella Campus Specifications (UPF)

Ciutadella Campus performs backups based on a type-class basis: user's data and virtual machines

- User's data (projects, HPC results, etc) are backed up:
 - Three times a day, locally (09:00, 15:00, 21:00), and retained for one day. Granularity: 3
 - Daily, weekly and monthly, once a day (23:00) and retained:
 - Daily backups, a whole week. Granularity: 7
 - Weekly backups, a whole month. Granularity: 4
 - Monthly, a year. Granularity: 12
- Virtual machines are backed up once a week, and copies are classified and retained on a time-basis:
 - 1 copy for weekly backup
 - 1 copy for monthly backup
 - 1 copy for a year's backup

4. Data security (1 question, 0 answered)

Address data recovery as well as secure storage

B *I* ☰ ☰ 🔗 📄 ▾

Save

Not answered yet

well as secure storage and
(general information) +

ications (UPF) +

ications (UPF) +

s (UPF) +

2.6 Recommendations to support researchers in selecting a repository

- The document provides sources for consulting:
 - disciplinary repositories (directories, publishers' recommendations, etc.)
 - multidisciplinary repositories (a comparative table showing the type of data allowed, the file size, the associated licences, the cost of depositing, etc.).



8. Cost	Gratuit. Però hi ha la possibilitat de més espai (EACS)	Gratuit	Gratuit fins a una certa mida: -5 GB pels arxius individuals -20 GB d'espai privat -Espai públic il·limitat	110 € per un conjunt de dades de fins a 20GB. Cost addicional per sobre de 20 GB	Gratuit	Gratuit
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 (R/D3a, SPSS, STATA, CSV, xls, FTIS)	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 (Publica)	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. Licències amb les que es volen difondre les dades	Permet varietat de licències encara que recomanen licències obertes. La descripció de les dades està sota CC0	Permet varietat de licències per a dades i per a software	CC0 per a dades (metadats & licències) CC-BY, per a no dades (audiovisuals, papers, etc.) MITL per a codi i software Altres licències: GNU, Apache 2.0, BSD 3 clause	CC0 (EACS)	Per defecte CC0 però es pot canviar	Permet varietat de licències: 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, quedaràn 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

<http://hdl.handle.net/2072/266502>

2.7 Research data policy

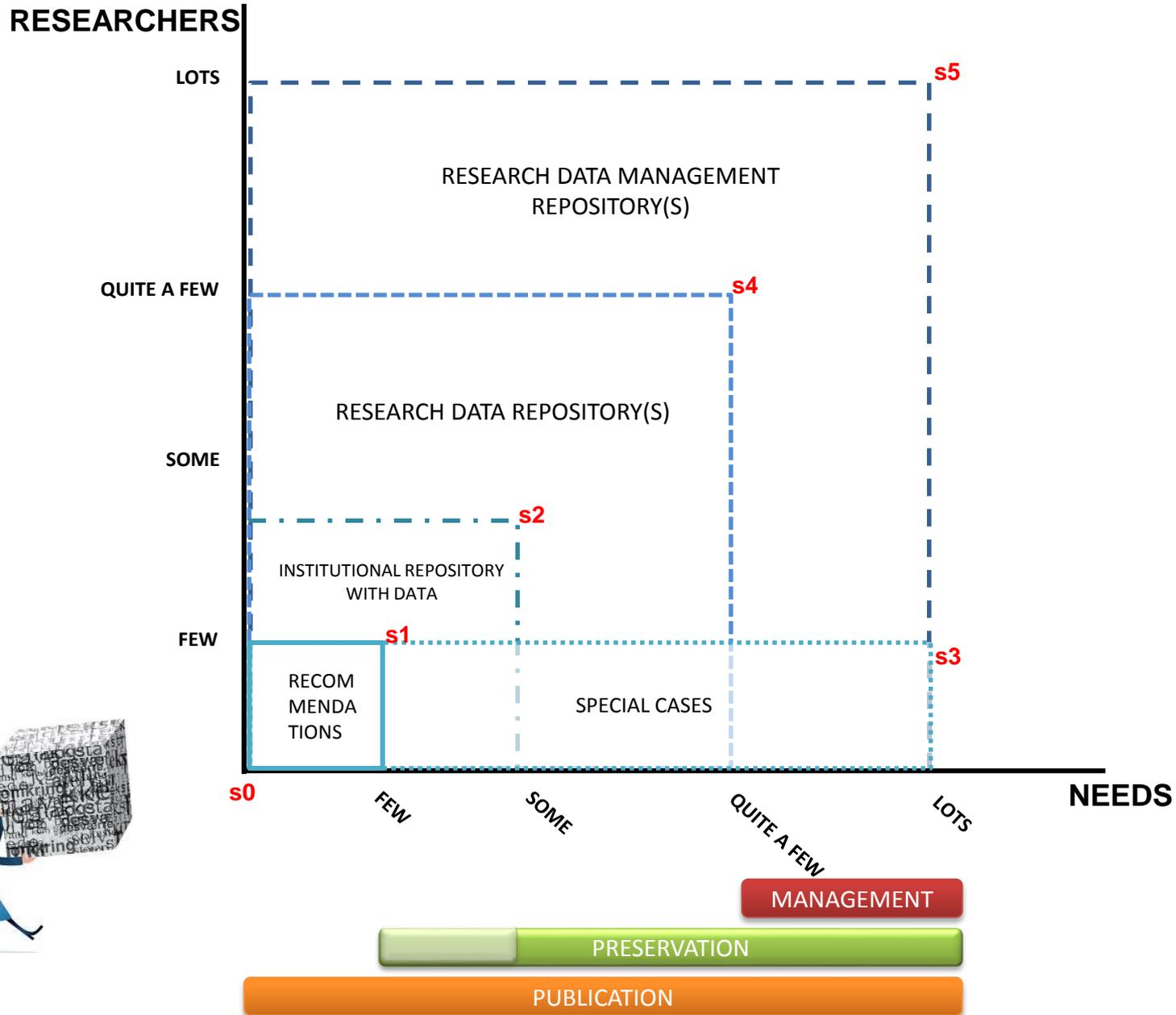
- Framework agreement for open access to research data supported by the vice rectors for Research of the Universities of Catalonia (<https://goo.gl/w6LcJk>)
- Models: “Policy RECommendations for Open Access to Research Data in Europe ([RECODE](#))” and “Model Policy for Research Data Management (RDM) at Research Institutions/Institutes” ([LEARN](#))



2.8 Monitoring

- Since the service officially started (October 2016) we have been collecting the actions associated to the RDM in universities
- These actions were measured with provisional indicators after the first 3 months of the service. They are currently collected periodically every 6 months
- The indicators used are:
 - H2020 projects
 - Staff involved in RDM
 - Training and RDM activities
 - RDM queries received
 - Visits received on the RDM website
 - Users registered in the DMP tool
 - Real plans created in the DMP tool
- **The big gap: a lot of actions, very few plans created!**

3.1 Our vision: different needs



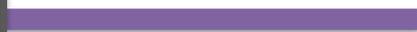
3.2 Our vision: different steps

RDMP

F1 **Recommendations**



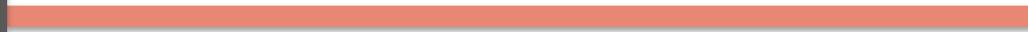
F2 **IR with data**



F3 **Special cases**



F4 **Research data repository(s)**



F5 **Research data management repository(s)**



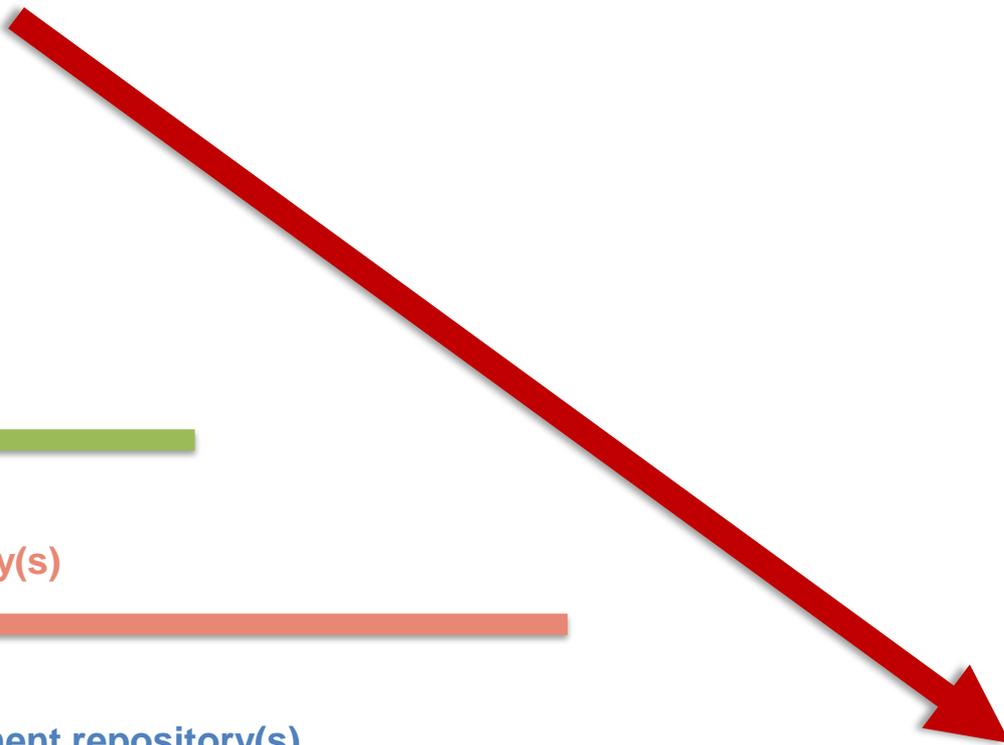
Short

Short-Medium

Medium

Medium-Long

Long



Step 1 - Recommendations

- As the needs differ a lot between disciplines, and, for the moment, we are not able to offer a university or Catalan repository...
- we will continue improving and getting up to date the recommendations
- This, added to RDM Plans, is a small but real service

8. Cost	Gratuit. Però hi ha la possibilitat de més espai (EACS)	Gratuit	Gratuit fins a una certa mida 5 GB pels arxius individuals 20 GB d'espai privat Espai públic il·limitat	110 € per un conjunt de dades de fins a 20GB. Cost adicional per sobre de 20 GB	Gratuit	Gratuit
	 zenodo GERN	 EUDAT	 figshare Digital Science (Macmillan)	 DR2AD NGSL	 Dataverse IQSS (Harvard University)	 MENDELEY DATA 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.)	<u>Tot tipus de formats</u> i també documents publicats	<u>Tot tipus de formats</u>	<u>Tot tipus de formats</u>	<u>Tot tipus de formats</u> (amb exclusió de software) i amb preferència per formats de codi obert	<u>Tot tipus de formats</u> amb funcionalitats addicionals per alguns (RData, SPSS, STATA, CSV, xls, FITS)	<u>Tot tipus de formats</u>
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 (València)	Obertes Restringides Embargades	Obertes Tancades (per defecte fins que no es publiquen) Restringides Embargades	Obertes <u>Embargades</u> (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 (metadades de dades) CC-BY per a no dades (audiovisuals, papers, etc.) MITL per a codi i software - Altres llicències: GNU, Apache 2.0, BSD 3 clause	CC0 (EACS)	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 quedaràn retinguts .	Per retirar contingut contactar amb EUDAT	Es pot sol·licitar la removal 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 .

Step 2 - Adapt the existing repositories to hold data

- Some Catalan universities have adapted their institutional repositories to permit research data deposit in order to respond:
 - requirements of the H2020 Open Research Data Pilot
 - editorial policies that request authors a permanent link to the data
- Modifications:
 - Use of HANDLE
 - The same tasks in terms of preservation and storage than the rest of the files
 - All universities have mapped OAI outputs to be interoperable with OpenAIRE
 - Differences to inform the typology of data in “dc.type” metadata
 - Data are grouped by type in collections
 - Volumetric issues were not considered
- This is not **the** solution, but
 - It’s easy and cheap to do it
 - Allow us to archive some type of data (in very restricted conditions)

Step 3 – Pilot project

- Some researchers (not a lot) have immediate and sophisticated needs in respect research data
 - some ones have disciplinary solutions (genetics, astrophysics...), but others, not
 - We don't have an institutional or Catalan repository for them (neither we will have it in the next two years)
 - These researchers are very motivated and they know data better than us
- **Characteristics of the 3 years pilot project (2018-20)**
 - **Central payment** (and only for 3 years) for a commercial solution + to hire a **data curator**
 - **A group of 10-15 research groups highly motivated** with one member of the group as **RDM leader**
 - RDM leaders + data curator = **learning group**
 - **Functions of Learning Group:**
 - share experiences and knowledge
 - annual accountability report
 - **help in determining functional requirements for the research data repository (step 4)**

Step 4 - Research data repository

- A commercial solution (step 3) seems not to be a feasible solution for all (step 5)
 - for cost,
 - for strategical reasons,
 - for interconnecting the RDR with other elements of the system
- 2017-18 objective: to determine functional requirements of the research data repository (RDR has to be a FAIR repository)
 - Permanent identifiers
 - Storage of high dimensions
 - Storage of different formats
 - Preservation of high performance
 - Interoperability
 - Management of special features
- Challenges of the research data repository
 - Involve all the stakeholders in the university
 - Vice rectors and researchers / Library, Research Office, ITC, etc.
 - To be one element of the system (jointly with IR, CRIS, Portal...)

Step 5 - Research data management repository

- As we have a lot of work to do before we arrive here, and
- a lot of unknown things can happen in the meantime
- we are not working on it (for the moment), except watching new developments



1. Who we are
2. Work done with research data
3. Work in progress with research data
- 4. Conclusions**

4. Conclusions, or, What we have learned

- There is a very big gap
- Cooperation is not altruism



When we talk about research data, there is a very big gap

- When we talk about research data...
 - a big gap between disciplines
 - a big gap between diagnosis and action
 - a big gap between what the bests are doing and what the rest can do
 - a big gap between senior and junior researchers
 - a big gap between Vice rectors and other ranking-worried-officers and researchers
 - a big gap between trendy and reality
 - a big gap between the view of IT people and librarians approach, and archivist vision and...
 - a big gap between..
- Two learnings
 - Start the way today but start it modestly
 - Don't stop but don't want to be the first,
- In the new paradigm (**Open Science**), probably, the big change is not on tools, but it's on processes
 - **and the key question is: where libraries have to be in the chain?**



Cooperation is not altruism

- Maurice B. Line (“Co-operation: the triumph of hope over experience?” 1997):
 - Cooperation (like second weddings) often represents “the triumph of hope over experience”, and
 - “cooperation should not be undertaken unless it is likely to produce better results than would be achieved by other means”



- Cooperation is costly
 - You need to align different needs and speeds between different institutions
 - You have to invest a lot in communication
- Two learnings
 - Sometimes cooperation is the unique opportunity to create a new service
 - Cooperation makes easy to get results, results are better and more accepted

Thanks!

Any question?



lluis.anglada@csuc.cat · @lluisanglada

<http://www.csuc.cat/en>