CERCAOPENDATA

DATA MANAGEMENT STRATEGY

Background

- Within the European Horizon 2020 Framework Programme (2014-2020), the European Commission initiated ORD Pilot, a project on research data, which in 2017 was extended to cover all projects. Under the Horizon Europe framework programme, it will be mandatory to publish open data and to produce a data management plan.

- On 27 March 2018, I-CERCA and the Barcelona Supercomputing Center (BSC-CNS) signed a framework cooperation agreement.

- On 4 May 2018, the Board of Trustees of I-CERCA agreed to instruct the director to implement an open data initiative for CERCA centres to facilitate and rationalise their data infrastructures, based on European FAIR criteria, and to produce a data management plan suitable for each institution and project.

- In November 2018 a Code of Conduct for CERCA centres was approved, point 2 of which referred to "Open Access to Research Data", stating that each centre must have an open science strategy and operate in line with the European FAIR principles, implement data management plans where appropriate, and curate and store research data for a recommended period of 10 years.

- In May 2019 a CERCA work group on data research management was set up. The work group was subsequently presented to the CERCA Conference and meets on a regular basis.

- On 6 June 2019, I-CERCA organised a workshop on open data and data management with the participation of speakers from The EUDAT Collaborative Data Infrastructure, The Dutch National Data Service Infrastructure and the Government of Flandes.

- In 2020 The National Pact for the Knowledge Society is approved. The agreement includes the implementation of a Catalan Open Science Strategy.

Action plan

The purpose of this plan is to ensure coordination with other agents in the system, such as universities, CSUC, BSC, etc., together with other cross-cutting initiatives that may take place in Catalonia, such as the National Pact for the Knowledge Society, once approved and in force.

1. Hold regular consultative meetings of the CERCA work group on research data management (every three months) to monitor progress on this strategy. It will discuss, propose solutions, and facilitate the implementation of the actions to be taken, and establish the parameters or indicators used to monitor said actions. A group of technical experts (GETEC) from the CERCA centres will meet in parallel to help ensure measures related to infrastructure and repositories are suitably refined.

    Schedule: Approximately every 3 months.

    Owner: I-CERCA

2. Develop a data management plan model or models, appropriately weighted for each scientific discipline, covering the data related to its scientific production. In the first phase, appropriate models and standard ontologies for metadata will be developed. These will be adapted for the CERCA centres in the second phase.

    Schedule: Urgent

    Owner: I-CERCA in coordination with CERCA centres, the CSUC and Catalan universities

3. Affiliate CERCA to the CSUC Open Data Repository, establishing what programming and formatting is needed to connect the computers on which the data is stored to the Repository. If this is not possible, a research data repository will be created for the CERCA centres.

    Schedule: 3 months

    Owner: CSUC - I-CERCA (GETEC)

4. Train researchers to develop research programmes that share data between R&D institutions. Researchers must be able to work with other researchers' data and make their own data openly available via a repository. These efforts will focus particularly on pre-doctoral and postdoctoral researchers (the next generation of researchers). Training must also be provided on the completion of data management plans.

    Schedule: 6 months

    Owner: I-CERCA - CERCA centres - CSUC

5. Train technical staff in data management so they can offer this service to CERCA centres (train the trainers), assisting in the creation of the post of Chief Data Officer, where required. This training will include the use of the DMPs proposed by I-CERCA (for training trainers in each CERCA centre).

Schedule: 6 months

Owner: I-CERCA

6. Prepare a CERCA data management guide to scientific research processes, propose file types, assign a DOI or equivalent persistent identifier to data, define general metadata (including ORCID researcher IDs) and forecast expected annual data volumes. Analyse how previously generated data is to be adapted and incorporated in a data repository. For those disciplines that already have a data repository, the existing dynamics must be respected. An assessment must be carried out into the possibility of I-CERCA becoming a member of the DataCite community in order to generate and assign DOIs, or whether it can work with another institution that can perform this system function.

Schedule: Approximately 6 months

Owner: I-CERCA with the support of CERCA centres

7. Implement a collaborative initiative to develop CERCA's data management system, network, physical storage and research data stewardship (infrastructures), including remote copies on tape. The storage capacity must be able to grow, anticipating likely scenarios for the generation of data at CERCA centres. Data must be kept for at least 10 years, as recommended by the CERCA Code of Conduct. GETEC, together with CSUC and BSC experts, if necessary, should assess whether it is desirable for this initiative to include the computational capacity of each CERCA centre, and data infrastructure nodes should be defined for the system (internal and external to CERCA) in Catalonia. Consider the following proposal for nodes:

a. Node I. Files up to approx. 10 GB (CSUC Nexus)

b. Node II. Files larger than approx. 10 GB (IFAE-PIC)

c. Node III (BSC).

d. Node IV. Geographically remote copies on tape (CSUC GEPA-Lleida or CERCA CTFC Solsona).

Schedule: 6 months

Owner: I-CERCA (GETEC) + Government of Catalonia + CERCA centres

8. Draw up a protocol of measures regarding data security and the security of the Anella Científica, with the support of CESICAT, the protection of sensitive data (personal details, data of commercial interest, etc.), and the temporary blocking of data where required.

Schedule: 1 year

Owner: CESICAT + I-CERCA (GETEC)

9. Determine the cost of common services linked to management, storage, access, and the recovery of research data from CERCA centres, according to the volume of data involved. Analyse financial viability compared with maintaining individual data infrastructures for each CERCA centre. Migrate progressively to a joint data storage system, limiting the purchase of large computers outside the nodes referred to in Action 7.

Schedule: 1 year

Owner: I-CERCA + CERCA centres

10. Define a programme of incentives for CERCA centres to migrate to a predominantly FAIR system in line with the CERCA Code of Conduct, facilitating management and, where appropriate, shared access to research data. Ask the Government of Catalonia to consider granting an additional budget allocation to facilitate the implementation of this pioneering strategy. Establish a protocol to determine what data should be open.

Schedule: 1 year

Owner: I-CERCA - Government of Catalonia

11. Reinforce access to Catalonia's Anella Científica to ensure there is sufficient bandwidth to meet the needs of CERCA centres. Plan future network demand scenarios.

Schedule: 2 years

Owner: I-CERCA - CSUC - Government of Catalonia

12. Systematically measure the volumes of data generated and made available via open access, queried, etc., to facilitate medium- and long-term data management planning.

Schedule: 2 years

Owner: I-CERCA

13. Curate the data (improve the quality and format). This will be done by each CERCA centre. A study will have to be carried out into the tools available, such as a data review robot, or an internal or common helpdesk for the CERCA system.

Schedule: Ongoing

Owner: CERCA centres

14. Prepare a proposal to enable research funding bodies and lenders (I-CERCA, AGAUR, AQU, ACCIO, etc.) to align with the main aims of this data management strategy (FAIR/ open data). I-CERCA will assess each CERCA centre's progress and participation in this strategy during the CERCA assessment period.

Schedule: Ongoing

Owner: I-CERCA

Action Plan

	1st year	2nd year	3rd year	From 3rd year
1. Kick-off meeting and creation of two working groups	■			
2. Development of a data management plan for each scientific discipline	■			
3. Construction of a repository of research data	■			
4. Training for research staff	■	■	■	
5. Training of technical staff	■	■	■	
6. CERCA data management guide	■	■		
7. Nodes and infrastructure. Management, networking, storage and preservation	■	■	■	
8. Anella Científica and Data Security Protocol		■		
9. Cost Management Model		■		
10. Incentive program		■	■	■
11. Strengthening access to the Anella Científica and planning future scenarios			■	■
12. Systematic measurement of data, traffic, Open data, FAIR data and planning			■	■
13. Data curation			■	■
14. Alignment of funding and accrediting bodies.			■	■