In Sharing We Trust: Taking Advantage of a Diverse Consortium to Build a Transparent Data Service in Catalonia

Clara Llebot
Oregon State University

Mireia Alcalá
Consorci de Serveis Universitaris
de Catalunya

Lluís M. Anglada i de Ferrer
Consorci de Serveis Universitaris
de Catalunya

Abstract

The Consorci de Serveis Universitaris de Catalunya (CSUC) is a consortium that serves 13 universities and 33 research centers in Catalonia and neighboring communities. In 2017 the Consortium created an Open Science department to collaborate with universities and research centers on facilitating the adoption of Open Science requirements. Even though CSUC also offers services to researchers directly (for example, its supercomputing resources), this report will focus on CSUC's work with its member institutions to create and offer data management services. We will explain how CSUC has led the creation of a robust shared governance system, and how it takes advantage of the diversity of its members to create useful, high quality, and transparent services for all researchers in the Catalan research system. Through sharing each other's experiences, values and priorities, the result is better than separate ad-hoc solutions. The process also creates a community of practitioners that develop expertise together with the help of professional development opportunities organized by CSUC, like recurrent self-learning labs focused on data curation tools, techniques and processes.

Introduction

The development of data services to encourage Open Science and to support the work of researchers with their research data is the responsibility of many. In the last few years, we can find many examples of such work. Funders and publishers have had a key role, creating requirements and guidance. It is well known that complying with publisher and funder requirements are some of the circumstances that motivate researchers the most to share their data (Digital Science et al., 2019). At the institutional level, the work has focused on providing support and resources to facilitate the task of managing and sharing data, primarily from university libraries, which often collaborate with teams from information technology and the research office (Ashiq et al., 2020; Tenopir et al., 2016). Regional and national networks, like consortiums, alliances or coalitions can also provide important services and effectively support other entities, such as research institutions with limited resources, expertise or infrastructure. Here we argue that these types of partnerships not only make it possible for some institutions to provide good quality services, but they also provide an added value. It takes work to find consensus among a large group of heterogeneous institutions, but the potential benefits are many: transparent and robust services, and importantly, a community of practitioners that support each other and develop expertise together.

This report describes some of the work of the Consorci de Serveis Universitaris de Catalunya (hereby referred to as CSUC or the Consortium), a consortium that serves more than 45 university and research centers in Catalonia. We will focus on two of the activities that the Consortium has engaged in that illustrate the creation of research data services as a collaborative process: the development of data management plan guidance, and the development of data repository policies. We describe the process, as well as highlight the challenges and advantages of this collaborative experience.

Context: the Consortium and Research Data Services

The Consoci de Serveis Universitaris de Catalunya is a public consortium that was created in 1991. Its core members are the Generalitat de Catalunya, plus the 8 public universities of Catalonia. Two other private universities participate in the consortium as affiliated members, and one university outside of Catalonia participates as an independent member (see Table 1).

The Consortium has the mandate to support research centers in Catalonia, even though these are not official members of the Consortium. These research centers are institutions that are generally research focused, discipline specific and smaller than universities, although the size of all research centers together is comparable to the size of a university. Each university or research center can choose to participate in each of the services organized by the Consortium independently from their participation in other services. At the time of writing this report more than 40 research centers are participating in the services organized and led by the Research, Learning and Open Science section of the Consortium. Throughout this article, when we talk about "participating institutions" we refer to the institutions, be they universities or research centers, that have chosen to take part in a specific service. Overall, the services of CSUC have the potential to reach about 25000 researchers.

Table 1. University members of CSUC, and their type of membership

University	Type of membership
Universitat de Barcelona (UB)	Core member
Universitat Autònoma de Barcelona (UAB)	Core member
Universitat Politècnica de Catalunya (UPC)	Core member
Universitat Pompeu Fabra (UPF)	Core member
Universitat de Girona (UdG)	Core member
Universitat Rovira i Virgili (URV)	Core member
Universitat de Lleida (UdL)	Core member
Universitat de Vic – Central de Catalunya (UVic-UCC)	Core member
Universitat Oberta de Catalunya (UOC)	Affiliated member
Universitat Ramon Llull (URL)	Affiliated member
Universitat de les Illes Balears (UIB)	Independent Member

The services offered by CSUC to its member institutions are varied.[1] They include joint procurement services, connectivity services for research, tools for e-administration, library services (like a shared catalogue and an inter library loan program), cloud and supercomputing services for research, and services related to Open Science. In particular, CSUC aims to provide research data services that will support researchers throughout the research lifecycle (Figure 1). The planning stages are supported by eiNa DMP, a tool for creating data management plans (DMPs) based on DMPRoadmap.[2] A collection of computational infrastructure and related services are provided to help researchers through their data collection and analysis tasks: storage, cloud services, and scientific computing. The Repository de Dades de Recerca (RDR), a shared institutional repository which is an instance of Dataverse,[3] allows researchers to publish, share and preserve data of any discipline. To support discovery and reuse CSUC coordinates the Research Portal of Catalonia (PRC),[4] a portal with information of current researchers and research groups in Catalonia, as well as their research outputs, including articles and datasets. All the services that support and promote Open Science are organized under the umbrella of the Catalan Open Research Area (CORA), which, among other things, allows each of the member universities to promote and display their resources and activities related to Open Science.[5] The Consortium provides instruction for researchers on how to use these services, and also for the administrators from each university or research center that liaise with them so that they can provide high quality support.

---

[1] See a description of the services in the CSUC website, https://www.csuc.cat/en/services

[2] https://github.com/DMPRoadmap

[3] https://dataverse.csuc.cat/

[4] https://portalrecerca.csuc.cat/

[5] https://cora.csuc.cat/en/open-science/open-science-at-universities-and-research-centers-in-catalonia/
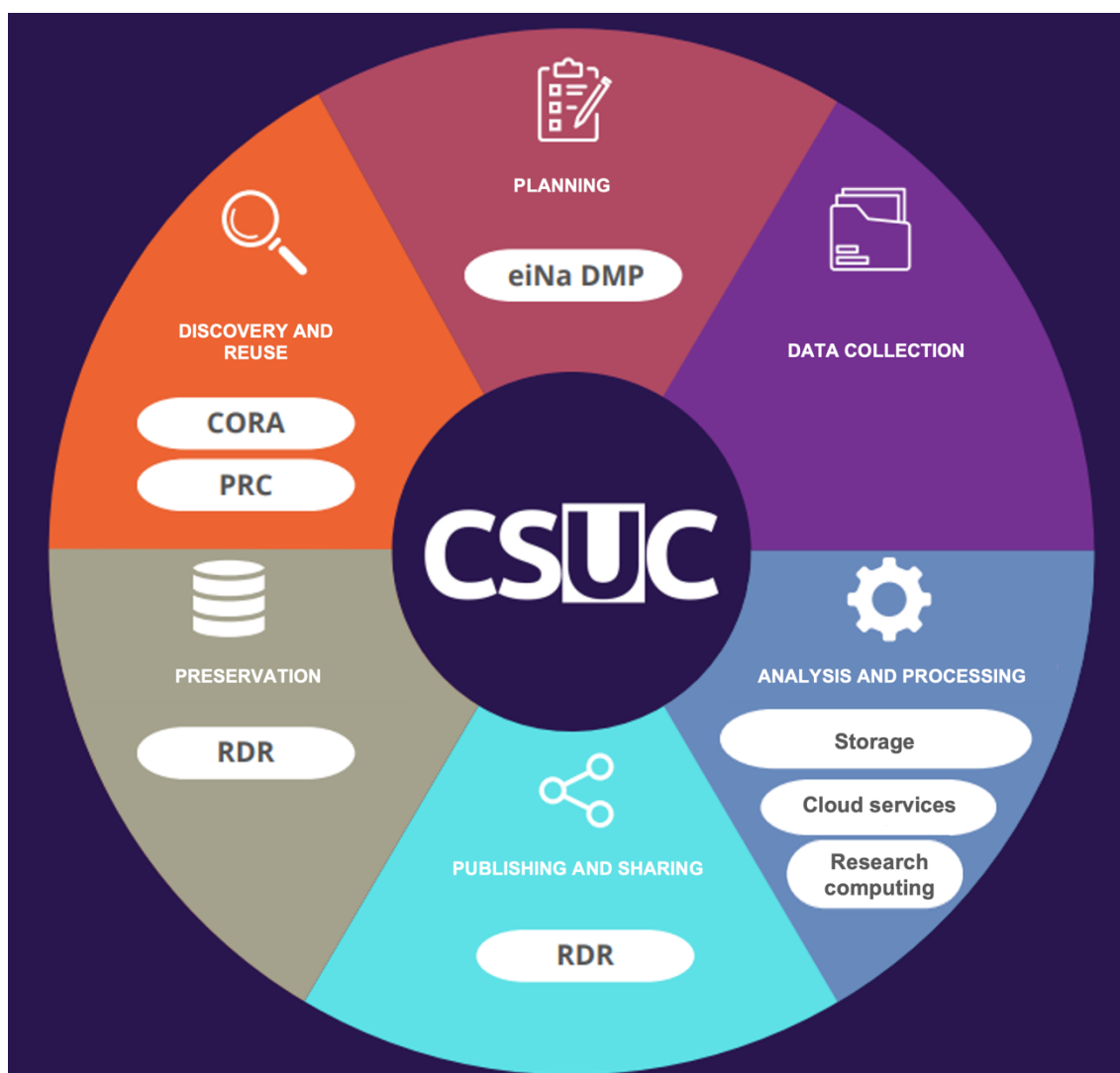
Figure 1. Diagram showing the services CSUC provides to support researchers to manage their data throughout the research lifecycle

Collaborative Development of Services

Data Management Plan Guidance

EiNa DMP is the tool that the Consortium offers to help researchers create, share, review and export data management plans. At the moment of this writing 12 universities and 26 research centers actively use this tool, and about 1300 data management plans have been created using it (this number excludes unfinished or test DMPs). The tool is built on DMP Roadmap,[6] the data management planning software provided by the Digital Curation Centre and the University of California Curation Center, making it very similar to other tools that also use the same platform, such as OnlineDMP or DMPTool. One of the main features of the tool is the collection of guides that a researcher has access to, that can be used by a researcher to learn how to write a strong DMP. These guides consist of general guides for each funder, accessible to all participating institutions, plus individualized guides that each university or research center can create about their own infrastructure, resources or policies. While the creation of individualized guides is left to the discretion of each institution, the general guides

---

[6] https://github.com/DMPRoadmap

had to be created for the tool to be launched. CSUC considered several options, such as translating and adapting guidance from DMPonline or other similar resources, or entrusting the development of the guides to an expert. Because of the nature of the Consortium, it was decided to opt for a collaborative model, in which representatives of each university would work together on creating the guides. These representatives were staff from the library or the research office of each university, who had data management support as one of their responsibilities. This process took place in 2015-2017, in the early stages of the work of the CSUC on Open Science, and research centers were not involved, only universities.

CSUC led the process, by organizing meetings and mediating discussions. A representative of the Consortium was present in all of the meetings that took place. The eleven universities were divided into three or four groups. Each group focused on a specific area of the DMP guidance for a particular funder, and wrote a draft document with explanations and examples. The document created with the contributions of all groups was then reviewed by the whole group to make sure that the content was accurate, and that the style was consistent. Finally, the Consortium was in charge of incorporating the content into eiNa DMP.

This system required high involvement from both the representatives of each university and from CSUC personnel, who convened and facilitated meetings. It was a slow process that took about six months from beginning to end. However, the overall outcome was positive. The resulting guides were of high quality and available to all researchers. Perhaps more importantly, by creating the guides, the participants had become experts in funder requirements and were able to provide strong support to researchers. They also had developed a sense of ownership of the tool, which they consistently promoted (and continue promoting) among their researchers. We believe this explains the high usage statistics that we see for eiNa DMP.

Creation of Repository Policies and Governance Documents

The Research Data Repository (Repositori de Dades de Recerca, RDR) is the discipline agnostic data repository that CSUC manages for all the participating universities and research centers. It is an instance of Dataverse that started operating in 2021. Staff members from each university or research center curate the datasets submitted into the repository by researchers of their institution.

In this case CSUC also opted for a model of shared governance that would ensure that all decisions were made collaboratively, and supported the participant institutions to make it happen. The most representative examples of this support are the organization of self-learning labs and the development of the repository policies.

The development of the policies that regulate the use of the repository is an iterative process in which new topics are added as problems and questions arise. For example, some of the questions that came up right away were deciding who can submit a dataset, when or why a dataset can be deleted, and the maximum size of a dataset. Every time that a new issue is considered the process is repeated and the policies are amended. The policies are, thus, a live document that is updated when necessary. To make decisions CSUC creates small groups. Each group addresses one of the topics in consideration, and writes a draft that is reviewed later by the whole group.

In this case the groups include staff from universities and from research centers. This fact is relevant because these professionals tend to be very different. Research center support staff often have a high level of disciplinary expertise and a strong background in research, but they often also wear many hats and data management is only a small part of their position description. Universities usually have staff with backgrounds in library sciences and information management that can devote a significant part of their work to Open Science and data management support. They typically have less disciplinary expertise, since universities also have a much broader research focus. Since the institutions are also heterogeneous it is useful to have representatives from each of them to ensure that the policies will reflect everybody's needs. For example, an issue that needed several iterations was 'who should be allowed to deposit a dataset in the repository'. Institutions that often work with interdisciplinary, international groups of researchers advocated for a policy that would allow a researcher not affiliated with one of the

participating institutions to deposit a dataset when they could demonstrate that they are collaborating with a participating institution.

Overall, this strategy is slower than it would have been to, for example, adapt and translate the policies of a similar repository, but it ensures that the policies reflect the reality of the repository and the researchers that it serves. Since the topics that are included in the policies are the result of issues that have been experienced, it is possible that there are gaps, but the system is flexible enough to address these gaps effectively when they occur. Because of the shared work, the curators and institution staff members share a sense of ownership of the repository, understand the reasons behind their limitations, and know how to take advantage of its potential.

All datasets that are deposited in RDR go through a curation process, and CSUC ensures that all curators are capable of guaranteeing a consistent level of quality. To that effect, the group developed a curation checklist, the REVISAT, inspired by the CURATED checklist from the Data Curation Network (2018). This enumerates the steps that all data curators commit to checking when reviewing a dataset for RDR.[7] Another strategy to ensure data curation quality is the organization of self-learning labs (Basalti et al., 2024), professional development opportunities that embody the collaborative spirit of the Consortium. These labs are organized and managed by CSUC staff, but they rely on the expertise of all the curators. Each session, which takes place every 2 or 3 months, revolves around a topic that CSUC personnel identify as interesting for the community. Often the lab starts with a short presentation that exposes basic concepts regarding the chosen topic. Then, several institutions are invited to explain the strategies that they are following, challenges that they encounter, and how they solve them, and at the end the whole group can discuss the issue and reach consensus when needed. For example, in fall 2023 the group met to talk about tools to transform Excel files to csv, and about controlled vocabularies. In this case two staff members of CSUC gave an introduction on these topics, and then three curators from research centers explained the controlled vocabularies that they use. During the session curators were encouraged to add the controlled vocabularies that they use into a shared document, organized by discipline, and there was discussion regarding the metadata "Keyword" and "Topic classification," which different institutions were using differently.

The large number of people curating datasets in RDR makes it hard to maintain consistency, and CSUC feels strongly that it is not their place to make important decisions for the group in a consortial repository. These self-learning labs are a powerful tool to talk about topics of interest, and they contribute to creating a community of practitioners that develop expertise together.

Conclusion

In this report we have described several examples of the work of a consortium to offer data services and support the personnel of research institutions to facilitate these service offerings. We have tried to capture the benefits and challenges involved. We have not covered technical or infrastructure work to support research data management, such as the establishment of a data repository, and we have not characterized the general advantages and disadvantages of library or higher education consortiums, which have been described by others (Burley et al., 2012; Galyani Moghaddam & Talawar, 2009).

The common denominator among these examples is that the initial collaborative work to set up a service is harder, and takes more time, than if it was done one-sidedly by consortium personnel, or for just one institution. And the work of a skilled coordinator and facilitator is essential. However, there are also many advantages. The resulting guidelines, policies, and regulations are generally strong and robust, because it was necessary to take into consideration the needs of a group that is heterogeneous. They are also transparent, because all the institutions that are affected by or that use these materials know how they have been developed, know why they state what they state, and had a role in shaping them. The creation of materials in a

---

[7] https://confluence.csuc.cat/x/BQkTCg

collaborative way also provides a sense of ownership of the results, and the institutions are more likely to promote them and to use them appropriately.

It is important to talk about the consequences of the work of a consortium for the people who participate in it. In the case of CSUC an important consequence of the collaborative process has been the learning that all participants have experienced, making this work a professional development experience. Another outcome has been the development of interpersonal trust among the participating members, and the creation of a strong community of practice. Participants help each other and share their struggles and their successes, similar to the experiences described by other communities of curators (Burns et al., 2024).

References

Ashiq, M., Usmani, M. H., & Naeem, M. (2020). A systematic literature review on research data management practices and services. *Global Knowledge, Memory and Communication*, *71*(8/9), 649–671. https://doi.org/10.1108/GKMC-07-2020-0103

Basalti, C., Fazekas-Paragh, J., Forni, M., van Gelder, C., Hasani-Mavriqi, I., Janik, J., Kalová, T., Kuchma, I., Lindroos, H., Lütcke, H., Pinnick, J., Raga, N., Thorpe, D., & Wildgaard, L. (2024). *Recommendations for Data Stewardship Skills, Training and Curricula with Implementation Examples from European Countries and Universities*. Zenodo. https://doi.org/10.5281/zenodo.10573892

Burley, D., Gnam, C., Newman, R., Straker, H., & Babies, T. (2012). Leveraging higher education consortia for institutional advancement. *International Journal of Educational Management*, *26*(3), 274–283. https://doi.org/10.1108/09513541211213345

Burns, H., Caldrone, S., & Narlock, M. (2024, February 20). *Curation is communal: Transparency, trust, and (in)visible labor (Slides)*. https://doi.org/10.5281/zenodo.10684435

Data Curation Network. (2018). *Checklist of CURATED Steps Performed by the Data Curation Network, v.1*. https://hdl.handle.net/11299/226891

Digital Science, Fane, B., Ayris, P., Hahnel, M., Hrynaszkiewicz, I., Baynes, G., & Farrell, E. (2019). *The State of Open Data Report 2019* [Report]. Digital Science. https://doi.org/10.6084/m9.figshare.9980783.v2

Galyani Moghaddam, G., & Talawar, V. G. (2009). Library consortia in developing countries: An overview. *Program*, *43*(1), 94–104. https://doi.org/10.1108/00330330910934138

Tenopir, C., Pollock, D., Allard, S., & Hughes, D. (2016). Research data services in european and north american libraries: Current offerings and plans for the future. *Proceedings of the Association for Information Science and Technology*, *53*(1), 1–6. https://doi.org/10.1002/pra2.2016.14505301129