Call for Collaborative Research Data Science projects

This document focuses on *Collaborative Research Data Science projects*, and describes the modalities of collaboration with the SDSC in these projects as well as the modalities of the corresponding call.

Please note that information and conditions applicable to all National Calls of the SDSC such as ***Project Duration, Eligibility criteria, Resources granted in Collaborative Projects, and Domains and tracks*** *are described in the* ***General documentation of the Second National Calls****.*

All second National Calls can be accessed here.

Collaborative Research Data Science projects

Collaborative Research Data Science projects are projects of up to two years, undertaken in collaboration with beneficiaries affiliated with Swiss institutions of higher education and research and/or with federal or cantonal administrations. In these projects, the SDSC commits to contribute competences in statistics, machine learning, and AI, including, but not limited to, computer vision, natural language

processing, statistical signal processing, and optimization. The SDSC is involved via the in-kind contribution of personnel (e.g., data scientist) working on the project.

The SDSC data scientists involved in our collaborative research projects typically have a PhD in one of the data science technical domains listed above. Many of them also have expertise in specific applied domains and their corresponding data science approaches. The contribution of the SDSC goes in most cases beyond the identification of relevant classical models and techniques required to solve a given problem. Instead, it typically consists in designing and implementing analyses and algorithms tailored to the problems and research questions considered, so as to best incorporate the prior domain expertise into state-of-the-art data science techniques. The work of the SDSC can also complement or support directly the contributions of other teams with methodological expertise in data science.

We are looking forward to supporting projects with strong scientific and societal impact, in which the contribution of the SDSC is important for the success of the project. The impact should be understood not only in terms of the immediate scientific results but also in terms of algorithms or technologies that can be operationalized after the project, or of open source algorithms and open research data produced as part of the project outcomes. A large fraction of our past and current research projects can be found on our projects page.

Given the mandate of the SDSC as a National Research Infrastructure, we particularly welcome projects benefiting a broad community of stakeholders aiming at one, or several, of the following: Developing Data Science/ML/AI tools and/ or generating new datasets/data sources that can contribute to the development of Data Science and/or Research Data Infrastructure platforms.

Expected mode of collaboration

Successful projects in data science often crucially rely on strong collaborations between domain experts and data scientists. The SDSC is fully invested in the projects in which it collaborates and typically leads or co-leads several work packages. We expect the domain expert collaborators to also be actively involved in the project, to help guide the development of the methods, and to participate directly in data exploration and analyses.

In most of our projects, the involved SDSC data scientists meet on a weekly or bi-weekly basis with collaborators from the partnering teams. These meetings can be at the SDSC, in the partnering lab, or via video calls. Given the importance of communication and informal exchanges of ideas in interdisciplinary research, for some projects, our data scientists can regularly spend time in one of the partnering labs. This possibility is examined on a case-by-case basis. We also organize bi-monthly progress meetings with all collaborators, co-PIs, and SDSC project coordinators.

Expected project outcomes

The typical outcomes of *Collaborative Data Science Projects* include:

- Scientific code, implementing methods or analyses, which will be made *publicly available* whenever possible in the context of the project;
- Curated and *openly accessible datasets* (to authorized users/organizations);
- Papers published in the journals of the community of the domain scientists;
- Communications in conferences in the domain sciences;
- Papers in data science journals or conferences, if the project requires specific advances in data science methods.

Application procedure

Overview

The application procedure for *Collaborative Data Science Projects* with the SDSC is divided into two phases: The submission of a pre-proposal and, upon acceptance of this pre-proposal, the submission of a full proposal.

The **pre-proposal** provides a general description of the planned domain science project and of the scientific questions that should be addressed using data science techniques.

After the invitation to submit full proposals (see the Timeline section), the SDSC will reach out to the applicants whose pre-proposals have been accepted, to offer to have **project discussions with the SDSC staff**. In particular, the objective will be to help define data science questions, modeling approaches, and potentially-relevant data science techniques. This interaction is by no means binding and is at the discretion of the applicants.

The **full proposal** should provide an in-depth description of the objectives and data science problem. At this stage, in addition to the data science questions, the applicants are encouraged to outline the data science approaches that will be considered. In both documents, the central role of data science in supporting domain science questions should appear clearly, as well as the collaboration scheme with the SDSC.

Timeline

Call for collaborative projects	Monday September 9th 2025
Pre-proposal **submission deadline**	Monday October 27th 2025
Invitation to submit full proposals	Friday December 19th 2025

Project discussions with SDSC staff	January 2026
Full proposal **submission deadline**	Friday February 27th 2026
Final decisions on full proposals	Friday June 19th 2026
Start of collaborative projects	September-December 2026

Evaluation criteria and procedure

The proposals will be evaluated based on the following criteria:

- ***Impact***: The project has clear scientific and societal impact. Contributions towards sustained and sustainable efforts in *open science* and *open research data which benefit a community of stakeholders are particularly welcome.*
- ***Data****: Most of the data* is available and accessible at the start of the project and is of appropriate quality and quantity for solving the research or technical questions. For any additional data that needs to be acquired, a conservative data acquisition plan should be proposed guaranteeing that data will be delivered in time. In cases where data transfer user agreements (DTUA) or ethical committee approval is required, accessibility of the data for a project in April 2026 is acceptable, provided that supporting documents demonstrating the feasibility of this timeline are submitted.
- ***Feasibility and risk assessment***: The objectives are attainable within the time of the project duration. The proposal includes a comprehensive identification of potential risks and outlines appropriate mitigation strategies to address them.
- ***Quality of collaboration***: The role of the SDSC is clearly described and should include scientific and technical contributions crucial for the success of the project. The collaborators should show that they will dedicate sufficient time to provide input and domain expertise for the development of the data science approach, data analyses, and to participate directly in the interdisciplinary work.
- ***Sustainability****:* If applicable, describe the long-term maintenance strategy of the data resources and data science tools produced as outcomes of the project.
- ***Clarity****:* The main challenges and objectives in the domain are described with sufficient detail and in terms accessible to non-domain experts.

The evaluation will be carried out by internal and external reviewers. The scientific merit and impacts of the project will be evaluated by external reviewers identified in the international research community of the domain, while technical feasibility on the data science side will be evaluated both by internal and external reviewers.

Pre-proposal and full proposal preparation guidelines

Pre-proposals and full-proposal should be written in English and include:

- A description of the scientific problem and its impact;
- A description of the data that will be used –characteristics, size, availability, etc.–, to assess the project feasibility from a data science perspective. A *data form* is supplied to this end;
- A specification of the research or technical questions to address using data science techniques, described as much as possible in data science terms. It should detail the following aspects:
    - The specific objective(s), such as the specification of what an algorithm should produce from what input, what is the underlying structure that the applicants wish to uncover in the data, etc.
    - Characteristics of the problems,
    - Key elements of domain knowledge to take into account when devising the methods.

Pre-proposal

Formatting instructions

Pre-proposals must not exceed 5 pages (A4 size, single spaced) excluding references, and with a minimum font size of 10pt, single line separation, and margins of at least 2.5cm. We recommend using the template provided.

Preparation guidelines

The pre-proposal should provide a high-level and accessible overview of the scientific project, from both the domain and the data science perspectives. It should clearly state and describe the precise data science questions that should be the focus of the collaboration with the SDSC. Furthermore, it should provide sufficient details to assess feasibility and relevance to the chosen track of the call for collaborative projects. Therefore, it should highlight what domain science problems will be addressed by data science techniques, without the need to specify precise models and data analyses. It should provide a description of the contributions of the PI and co-PI teams to the project. Furthermore, the main challenges and objectives should be outlined, specifying the roles of the SDSC and of all partnering teams on each of them, and key tasks assigned to the SDSC Data Scientist. At this stage, however, no detailed work package description or detailed resource allocation is required. Detailed templates for the pre-proposal document with additional guidelines and the data-related questionnaire are available on the webpage of the call in Word and Latex. Along with the pre-proposal, a *data form* should be filled to provide a clear description of the available datasets, and their intended usage throughout the project. If relevant, the applicants have the possibility to submit data samples and/or a metadata file to illustrate some of the characteristics or the structure of the data.

Pre-proposal submission

The application documents must be submitted before 23:59 CEST on the the day of the pre-proposal submission deadline on the CMT portal (Microsoft's *Conference Management Toolkit*), accessible from the webpage of the call and must include:

- A ***pre-proposal*** (max 5 pages);
- Short CV(s) of all applicants (max 2 pages)
- A ***data form*** (max 5 pages)

- A ***list of partners*** entered in the form on CMT
- Optionally, a subset of the data and/or metadata information that illustrates some characteristics or structure of the data.

Only pre-proposals receiving a positive evaluation will be invited to submit a full proposal during the second phase.

Upon acceptance of the pre-proposal, the applicants will be invited as well to suggest the names of three independent reviewers for the evaluation of their full proposal who have both domain science expertise, and, to the extent possible, data science expertise as it applies to their domain.

Full proposal

Formatting instructions

Full proposals must not exceed 15 pages (A4 size, single spaced) excluding references, and with a minimum font size of 10pt, single line separation, and margins of at least 2.5cm. We recommend using the template provided.

Preparation guidelines

Full proposals should contain a detailed description of the proposed research project. It should describe precisely the data science questions, the specific objectives and milestones, possible data science approaches as well as the expected outcomes. The full proposal should elaborate on the organization, specific roles, and interactions between the team of the different applicants (PIs and co-PIs) and with the SDSC. It should also describe the different work packages and deliverables, along with potential feasibility risks and mitigation strategies. The full proposal should include a detailed resource allocation proposal for the project.

A template for the full proposal document with additional guidelines and a ***resource form*** will be available on the webpage of the call. An updated version of the ***data form*** should be submitted along with the full proposal.

Full proposal submission

If the applicant is invited to submit a full proposal, it must be submitted before 23:59 CEST on the day of the full proposal submission deadline on the CMT portal, on the CMT portal linked from the webpage of the call.

The structure of the full application is similar to that of the pre-proposal and includes:

- An updated and detailed ***full proposal*** (max 15 pages, excluding references);
- A detailed ***resource form***;
- An updated ***data form*** (max 5 pages);
- Short CV(s) of all applicants (max 2 pages);
- Supplementary documents, if applicable; e.g., support letters, commitment letters, data usage agreements, etc.
- Optionally, a subset of the data and/or metadata information that illustrates some characteristics or structure of the data.

Contact

For questions on this call please write to research-call@datascience.ch .