

# *3<sup>rd</sup> Call for Collaborative Data Science Projects*

2019

Submission Deadlines

Pre-Proposals: Jun 21, 2019

Full Proposals: Sep 30, 2019

Swiss Data Science Center  
EPFL, Station 14, 1015 Lausanne  
ETHZ, Universitätstrasse 25, 8006 Zürich  
Tel +41 21 693 43 88  
+41 44 632 80 74  
E-Mail [contact@datascience.ch](mailto:contact@datascience.ch)  
[www.datascience.ch](http://www.datascience.ch)

# Content

<u>1. OVERVIEW</u>	<u>4</u>
<u>2. RESEARCH THEMES</u>	<u>5</u>
2.1 MODULE A: DATA SCIENCE MEETS DOMAIN SCIENCE	5
2.2 MODULE B: DATA SCIENCE METHODS FOR THE REAL-WORLD	5
<u>3. SUBMISSION PROCEDURE AND PROJECT SELECTION</u>	<u>6</u>
3.1 GENERAL CONDITIONS	6
3.2 COVERED EXPENSES	6
3.2.1 COMPUTE AND STORAGE RESOURCES	6
3.2.2 SDSC TECHNICAL STAFF	7
3.2.3 STAFF FROM THE GRANTEES' INSTITUTION	7
3.3 PRE- AND FULL PROPOSALS: DEADLINES AND MATERIAL	7
3.3.1 SUBMISSION MATERIAL FOR PRE-PROPOSALS (MAX 5 PAGES)	8
3.3.2 SUBMISSION MATERIAL FOR FULL PROPOSALS (MAX 15 PAGES)	8
3.4 SELECTION CRITERIA	8
3.5 SCHEDULE	9
<u>4. ANNUAL SCIENTIFIC AND FINANCIAL REPORTING</u>	<u>9</u>
<u>5. CONTACT</u>	<u>9</u>

# 1. Overview

Data Science is one of the Strategic Focus Areas defined in the Strategic Planning 2017–2020 for the ETH Domain. The ETH Domain has launched the Initiative for Data Science in Switzerland to accelerate data science through education and research and the provision of infrastructure<sup>1</sup>. The Initiative created the Swiss Data Science Center (SDSC), whose mission is to accelerate the use of data science and machine learning techniques broadly within academic disciplines of the ETH Domain, the Swiss academic community at large and the industry. At the meeting of May 20/21, 2015, the ETH Board allocated 30 m CHF to the initiative over the next four years.

About two thirds of the budget are allocated to Collaborative Data Science Projects that are reviewed on a competitive basis. SDSC Collaborative Data Science Projects can either focus on the in-depth analysis of a particular interdisciplinary data science problem in a specific scientific domain, or on the development and the implementation of a technology or method that has the potential of broadly enabling data science research.

The motivations of this third call for SDSC Collaborative Data Science Projects are threefold:

1. Foster and accelerate the adoption of data science across the ETH Domain: The SDSC is expected to facilitate a strong synergy between data providers, data and computer scientists, and subject-matter experts, fostering scientific breakthroughs with significant societal impact. Projects showing an interdisciplinary character by linking research groups from traditionally separated disciplines are encouraged.
2. Explore the use of or expand on the SDSC hosted platform (Renku<sup>2</sup>) and machine learning tools for ETH Domain scientists: The SDSC is developing a platform for reproducible, reusable and collaborative data science. Projects must explicitly describe how they intend to leverage the SDSC platform, or expand its tools.
3. Promote Open (Data) Science: The SDSC platform also offers user-friendly tooling and services to help with the adoption of Open Science, fostering collaboration, research productivity and excellence. Projects are expected to follow guidelines and best practices to enable reusability and reproducibility of research.

An SDSC Collaborative Data Science Project is proposed and led by one main applicant from an ETH Domain institution. The Management Office of the SDSC will assist in the administration and coordination of the project and will ensure accountability of the project according to the directives defined.

The center will cover the operational expenses of the selected projects, including provisioning of compute and storage resources, and cost recovery of the center's R&D embedded staff, who can be temporarily located in any of the ETH Domain institutions, as well as cost recovery for R&D staff specialized in the research field of the proposal, who would be employed by the grantees' institutions for the specific project, if properly justified.

---

<sup>1</sup> "The ETH Domain launches the Initiative for Data Science in Switzerland", ETH Board Fact Sheet, May 2016 - [http://www.ethrat.ch/sites/default/files/FactSheet\\_Data\\_Science\\_E.pdf](http://www.ethrat.ch/sites/default/files/FactSheet_Data_Science_E.pdf)

<sup>2</sup> The Renku Platform - <https://datascience.ch/solutions/>

## 2. Research Themes

Candidate Data Science Projects are expected to focus on one of two themes: (i) the in-depth analysis of a particular interdisciplinary data science problem in a specific scientific domain, or (ii) the development and the implementation of a technology or method that has the potential of broadly enabling data science research.

The two themes are detailed in the sections below.

### 2.1 Module A: Data science meets domain science

We invite proposals for data-driven science projects requiring collaboration in any area of data science involving data hosting, curation and management for better data sharing, and statistics, machine learning and visualization for insights discovery.

This theme primarily targets research groups from domain sciences. Projects showing an interdisciplinary character by linking research groups from traditionally separated disciplines are strongly encouraged.

The scientific domains include but are not limited to -

- (i) Environmental sciences
- (ii) Personalized health

The funded projects will get access to and use the SDSC hosted software platform (Renku). In particular, members of the SDSC technical staff will facilitate uploading/streaming research data from the grantees' institution to the platform, to perform the data-driven research on the platform, and to finally publish the results in the platform in a long-term reproducible and reusable way.

SDSC scientific and technical staff is proficient in the design and development of machine learning algorithm for custom data sources.

### 2.2 Module B: Data science methods for the real-world

We invite proposals for research and development of system services and/or analytics methods that have the potential of broadly enabling data science research.

This theme primarily targets research groups working on data science methodology, ideally motivated by a real application. The funded projects will get access to the SDSC hosted software platform and the data stored therein. Members of the SDSC technical staff will facilitate data extraction and conversion for best integration with the grantees' institution work flows. They will further ensure integration of the newly developed tools into the SDSC platform with appropriate tracking of re-use and further development of the elements in the long term.

Research areas include but are not limited to:

- a. Interactive Data Science – Combining human expertise and machine intelligence
- b. Robust Data Science – Making machine learning methods robust against uncertainties
- c. Interpretable Data Science – Designing methods for interpretable machine learning
- d. Data Science in a Security- and Privacy-Conscious World
- e. Data (Science) Governance – Formally managing data (science) assets at scale

## 3. Submission Procedure and Project Selection

### 3.1 General Conditions

- Calls for proposal are organized every year with opening of the call in the summer and actual starting of the projects in the first half of the next year.
- Only researchers employed by an institution of the ETH Domain are eligible to apply. Institutions member of the domain include: Eawag, Empa, EPFL, ETHZ, PSI, and WSL.
- Project duration is limited to 24 months.
- The budget of a project is expected to be between 200'000 and 600'000 CHF.
- A two-stage submission procedure is used: Pre-proposals are submitted first, a selection of pre-proposals is then invited so submit a full proposal.
- Pre- and full proposals are to be submitted in English.
- Proposals are to be submitted by e-mail to the following address: [admin@datascience.ch](mailto:admin@datascience.ch).
- Proposals must clearly demonstrate how the SDSC platform and machine learning expertise are used (data hosting and sharing, open science: reusability/reproducibility of research) and/or extended.
- Proposals that are continuation of the awarded projects from the 2017 call can be submitted. They will be considered as new proposals and be subject to the same review criteria as new proposals.

### 3.2 Covered Expenses

If properly justified, the following operational expenses of the funded projects will be covered: (i) compute and storage resources, (ii) recovery of the cost of SDSC employees required for the execution of the project, and (iii) recovery of the cost of specialized staff from the grantees' institutions specialized in the research field of the specific project.

These expenses are detailed in the sub-sections below.

#### 3.2.1 Compute and Storage Resources

SDSC relies on existing infrastructures of the ETH Domain or national infrastructures (in particular by leveraging resources at the Swiss National Supercomputing Centre - CSCS), SWITCH, as well as those of cloud providers.

The following compute and storage resources can be requested:

- CPU Cores and RAM (through virtual machines or compute time on HPC clusters).
- Data storage.
- GPU boards.

The acquisition of computing and storage equipment is not covered, nor is the use of compute and storage resources not managed by the SDSC.

The SDSC platform ships with tools typically found in big data and machine learning systems, including tools like Hadoop, Spark or TensorFlow (to name a few). If properly justified, other tools can also be deployed within the software platform. A feasibility assessment will be made following the pre-proposal.

### 3.2.2 SDSC Technical Staff

The SDSC personnel specializes broadly in both large-scale processing of massive data sets, at rest and/or in motion, and in data science, in particular in the fields of data curation, data privacy, machine learning, artificial intelligence and statistic.

The proposals must clearly list the skills in computer and data science required for the execution of the project, and associated expected effort. These expectations will be further elaborated on and validated together with the SDSC team in preparation of the full proposal.

### 3.2.3 Staff from the Grantees' Institution

In many cases, specialized staff employed by the grantees' institutions are critical for the successful execution of the project. The cost of a postdoc, technician or PhD student with necessary skills outside the SDSC staff's expertise is eligible. If needed (no other source of funding) and properly justified, the cost of this staff can be partly recovered.

Proposals are expected to justify the need of a specialized member of the grantee's staff, and quantify the associated effort.

## 3.3 Pre- and Full Proposals: Deadlines and material

Project proposals are to be submitted in English as a PDF file.

Stage	Pre-proposal	Full Proposal
Material	<ul style="list-style-type: none"> <li>- Project description (see details below)</li> <li>- Short CV of all applicants (max 2 pages)</li> <li>- List of publications (10 max)</li> </ul>	<ul style="list-style-type: none"> <li>- Revised project description</li> <li>- Budget request with guidance from organizers</li> <li>- Short CV of all applicants (max 2 pages)</li> <li>- List of publications (10 most relevant)</li> <li>- Supplementary documents, if applicable; e.g., support letters, confirmation of co-operations</li> </ul>
Deadline	June 21, 2019	September 30, 2019

### 3.3.1 Submission material for pre-proposals (max 5 pages)

The pre-proposals should include the following sections using the template available on the call web page<sup>3</sup>:

- Summary (1 page): Concise statement of the goals, milestones and significance of the project.
- Background and significance to Data Science.
- Scientific goals and objectives, explicitly highlighting alignment to the call's objectives.
- Research plan:
  - General description of the scientific approach.
  - Description of the data sources and their use, if applicable.
  - Work packages, milestones and deliverables.
- Requested resources:
  - Compute and storage resources (minimum CPU cores & RAM, storage, GPU).
  - Software packages (Hadoop, Spark, TensorFlow, etc.).
  - Data science expertise (skills expected from SDSC technical staff).
  - Specialized staff (name(s), CV(s), task(s), expected effort and associated cost).
- Contributed resources.

### 3.3.2 Submission material for full proposals (max 15 pages)

All applications invited to submit a full proposal will receive guidance from the organizers of the program, in particular on validating the requested resources, selecting appropriate compute and storage infrastructure, assigning SDSC staff, and converting to actual budget request.

## 3.4 Selection criteria

Proposals must reflect the call's objectives as outlined under Research Themes.

A Scientific Review Committee, appointed by the Steering Committee of the SDSC and composed of members of the Steering Committee and external experts, possibly international, will be responsible for the evaluation and selection of the proposals.

Proposals will be evaluated based on the following criteria:

- Scientific merit;
- potential impact of results;
- potential contribution to Open Science;
- potential value to the SDSC platform;
- proposed collaboration with SDSC;
- novelty of approach;
- technical feasibility; and,
- balance of funding.

all relative to the expected operational complexity (e.g., infrastructure cost, data compliance).

---

<sup>3</sup> Project pre-proposal template available here:

[https://datascience.ch/wp-content/uploads/2019/05/SDSC-preproposal\\_template\\_2019.docx](https://datascience.ch/wp-content/uploads/2019/05/SDSC-preproposal_template_2019.docx)

### 3.5 Schedule

Call for pre-proposals:	May 15, 2019
Submission of pre-proposals	June 21, 2019
Invitation to submit full proposals	July 26, 2019
Submission of full proposals	September 30, 2019
Final decision on full proposals	December 20, 2019
Start of projects	March-June 2020

## 4. Annual Scientific and Financial Reporting

Every year, a scientific progress report and financial report are to be submitted to the Management Office. The reports will be consolidated and passed to the Steering Committee where they will be reviewed.

Scientific reports should include the progress made during the period, activities foreseen for the following period and any changes or deviations from the initial project plan.

Financial reports should detail all spending during the period.

## 5. Contact

For questions regarding the submission of pre-proposals and full proposals, please contact the Program Manager: [program.manager@datascience.ch](mailto:program.manager@datascience.ch)