

Social Data Science Team (SoDa) Principles

Criteria and working principles

SoDa wants to be a **low-threshold** data science team: researchers can contact us and we will always try to meet with them to discuss their project. The projects come from ODISSEI grants, activities, and community events, as well as direct contact with researchers. As such, there will necessarily be a large variation in the type of research question, data, and solution required.

Currently, meeting with the SoDa team can result in **consultations** or larger **research projects**.

Research projects

If we notice in our meeting that the research project requires substantial involvement of the SoDa team we start defining a research project in collaboration with the researcher. The project should exhibit the following principles: social science, team science, open science, capacity building, and goal-orientation. SoDa projects should:

- bring **added value** to the field, either by answering interesting social-scientific research questions using (ODISSEI) data and/or computational methods that could not have been answered otherwise, or by enabling such answers through the development of research software and modelling – and ideally, both;

- are truly **collaborative**, in the sense that both the SoDa team and the substantive researchers are equal partners in decisions regarding the setup of the study, data analysis, software design, and peer-reviewed publication(s);
- adhere to the **highest feasible standards** of reproducibility, open science, and research integrity.
- benefit especially from other resources developed at ODISSEI, such as CBS microdata, the OSSC, or the CBS social network file.
- define a S. M. A. R. T. goal, resulting in **one or more of the following outputs**:
 - Reusable **software** packages (the team will measure usage & references). Reusable software become part of the infrastructure of available methods for computational social science. To this end the SoDa team will collaborate with NLeScience, for example to make research software FAIR through the portal research-software.nl. Following other parts of ODISSEI, code will be published under the Apache 2.0 license.
 - **Tutorials** to make software accessible (measure website visits). These tutorials are vital components of any infrastructure, as they document usage not only for existing users, but also for new users who will be able to assess whether this method is relevant to their situation. Any materials generated this way will be made available openly.
 - **Research proposal** contributions, i.e., being mentioned and/or co-applicants of research proposals. This in collaboration with both early career researchers and consortia. These contributions are an investment into the

longevity of the team beyond the last period of the current project.

- Research **papers** of one of the following types:
 - *Substantive* papers with SoDa involvement. These will typically be publications in the field of social science, where (some) computational aspects are partially done by one of the members of the SoDa team.
 - *Methodological* papers with substantive motivation. These will be byproducts of the computational and data-scientific insights of the SoDa team members when working on various projects. In other words, methodology developed at SoDa will always have a strong motivation of actual use in social-scientific applications.
- **Data products** with documentation (task 1.2). Some projects may result in datasets which can be used by other social scientists. These datasets will be documented and distributed in a FAIR (findable, accessible, interoperable, reusable) way in close collaboration with the data stewardship components of task 1.2 at ODISSEI.

Examples of such projects SoDa already is involved in are:

- Creating a statistical model to determine the success factors for housing search using large-scale internet data
- Creating an efficient parallel processing pipeline to speed up computation for <https://kansenkaart.nl/>

- Building a small prototype of a diagnostic questionnaire collection app with visualisations together with a psychologist

Consultations

If we notice that the researcher just needs to be pointed in the right direction, we will try to help them immediately. Sometimes this means scheduling a few more meetings to explain things or to be a brainstorming partner.

Examples of such consultations SoDa already does are:

- We are consulting the FIRMBACKBONE project on the implementation of their data infrastructure
- We did a one-on-one consultation about parallelization in R code for a sociologist
- We brainstormed with a social psychologist about determining perceptions about art from online text
- We discussed projects with several researchers before they submitted a proposal to the eScience grant