

SNSF Career Tracker Cohorts Data Management Plan

1 Data collection and documentation

1.1 What data will you collect, observe, generate or reuse?

This project produces questionnaire-based survey data and makes use of pre-existing data sources as described below. The volume of data for an individual survey wave including all files such as questionnaire data spreadsheets, administrative data spreadsheets, code of online questionnaires, or lists of standard classifications is around 200 MB. The main types of data used in this project are as follows:

- Administrative data provided by the SNSF. The administrative data is made available to the project team in Excel format. It contains personal identification data (names, gender, birth year, contact data) and administrative data related to the funding (instrument, discipline, grading of the application, funding decision etc.). The personal data will be used for the sole purpose of contacting the survey participants and for linking the survey data with those elements of the administrative data that are relevant for data collection and analyses.
- Survey data collected from the participating applicants to SNSF career funding schemes using an online survey software. The raw survey data produced by the survey software is in Excel and CSV format. The raw survey data will be processed by the project team to make them suitable for analysis and stored in several formats such as plain text format (CSV), Excel, and Stata format. This processing includes linking the survey data to administrative data that are necessary for data analysis.
- Data related to the implementation of the online questionnaire. This mainly includes Excel files containing standard classifications (e.g. classifications of occupations, lists of disciplines, or lists of research institutions).

1.2 How will the data be collected, observed or generated?

We collect data by means of online surveys with applicants to SNSF career funding schemes at the postdoc level. The online surveys are implemented using Limesurvey, an open source survey software, and are hosted locally on a server within the subnet of the Department of Social Sciences in the IT network of the University of Bern. The survey data is then partially linked with administrative data of the applicants (gender, birth year, discipline etc.), which is provided by the SNSF via safe transfer. In order to link the data without using and including personal identifiers like names in the dataset (for reasons of privacy and independence), we use a separate password-secured look-up table that matches the SNSF person number and the identification number used in the CTC project (CTC-ID). The CTC-ID is a random number consisting of eight digits and is different from the SNSF person number.

1.3 What documentation and metadata will you provide with the data?

For each survey, we compile a detailed documentation that contains information on the questionnaire development (e.g. pretesting), the final questionnaire (including screen shots), the survey participants and the field work (e.g. characteristics of the survey population, response rate, duration of the interviews), the data preparation and data quality (e.g. data cleaning, classifications, definition of variables, nonresponse analysis), and the final data (codebook). This documentation will allow external users to understand and work with the data (see Section 4). The documentation will be made publicly available at the project website. In addition, the documentation will be included in the data distribution archived at FORS.

2 Ethical, legal and security issues

2.1 How will ethical issues be addressed and handled?

The participants of the surveys are informed in advance by the SNSF about the aims of the Career Tracker Cohorts (CTC) study and the planned data collections. They are assured that their data will be treated with strict confidentiality and that their participation will be unrelated to the evaluation procedure of their application. Furthermore, they are notified about the transfer of their contact information and the required administrative data to the project team, and the use of this data by the project team throughout the project. With this information, they are given the possibility to opt out of the study. In this case, their data will not be transmitted to the project team and they will not be invited to participate in any of the surveys. Participants are also given the possibility to retract their consent at any later point in time, in which case they will no longer be contacted for any future surveys.

The SNSF provides the data of all participants who did not opt out to the project team via secure transfer. As mentioned above, for linking the survey data with the administrative data, we use a separately saved and password-protected look-up table that contains the necessary IDs (SNSF person number and CTC-ID) for the linking process. Separate IDs are used in order to guarantee privacy and independence from the evaluation and administrative processes at the SNSF.

The CTC project was approved on September 25, 2018, by the ethics commission of the Faculty of Business, Economics and Social Sciences of the University of Bern.

2.2 How will data access and security be managed?

Access to the project data stored on SWITCHdrive (see section 3.1) is limited to the members of the project team (edu-ID login). Files containing sensitive information (identifying information such as names and addresses) will be exchanged by secure-transfer between the SNSF and the project team and kept in an encrypted and password-protected format.

2.3 How will you handle copyright and Intellectual Property Rights issues?

The SNSF is the owner of the data and the deliverables. The property rights are set out in the mandate contract between the Swiss National Science Foundation (SNSF) and the University of Bern, acting through the Institute of Sociology and the Interdisciplinary Centre for Gender Studies ICFG.

3 Data storage and preservation

3.1 How will your data be stored and backed-up during the research?

The collected survey data, the administrative data from the SNSF and all other project relevant documents are stored in a project folder on the cloud-based SWITCHdrive. SWITCHdrive is a non-commercial cloud storage service for institutions of higher education in Switzerland. All data on SWITCHdrive is stored exclusively on servers that are part of the Swiss academic network. Access to this folder is limited to the members of the project team at the University of Bern. Moreover, all files that contain identification information (e.g. names, addresses) are secured by a password that is only known to the members of the project team.

Daily backups of the SWITCHdrive folder are stored on a server within the subnet of the Department of Social Sciences in the secured IT network of the University of Bern. Access to these backups is limited to the members of the project team at the University of Bern.

3.2 What is your data preservation plan?

We will preserve all data and documents related to the project on SWITCHdrive for as long as the project is conducted by us. Upon termination of the contract, we will return or destroy, as requested by the SNSF, all information that is not publicly accessible (confidential information), without keeping any copy thereof. The datasets and documentation published on FORSbase will be available according to FORS's regulations.

4 Data sharing and reuse

4.1 How and where will the data be shared?

The collected data will be made available to the scientific community in the form of scientific use files. The scientific use files, together with the documentation and plain text dictionaries of variables and labels, will be archived on FORSbase in plain text format (CSV), Excel, and Stata format.

The scientific use files will be anonymized in the sense that they do not contain any personal identifiers. The privacy of the respondents will be further protected by anonymization measures that concern the data directly as well as by procedural measures that concern access and use of the data.

As with most social-science survey data, identification of at least some of the individuals may be possible through elaborate procedures that exploit covariate patterns. Such attempts to identify individuals are prohibited by the data protection contract that researchers willing to use the data have to sign and are thus illegal. In general, the number of individuals that could potentially be identified by such procedures is small because surveys are typically conducted with samples from large populations. This is different for the CTC study where no sample is drawn and all applicants are invited to take part in the study. The number of individuals that could potentially be identified may thus be substantial. To safeguard against misuse of the data and ensure that the privacy of the respondents is adequately protected we thus apply de-facto anonymization and also take procedural measures controlling access to the data. De-facto anonymization is achieved by removing information from the data that could potentially be used to identify individuals (such as, e.g., the name of the host institution). Detailed information on the steps taken to anonymize the data can be found in the data documentation that is provided with for each scientific-use file (the documentation is made available on the project website at www.careertrackercohorts.ch and is included in the data distribution at FORSbase). The procedural measures controlling data access are as follows:

- 1) The scientific-use files will be archived on FORSbase. This ensures that only researchers and students affiliated with a university, a university of applied science, or a research institution for scientific research and/or education purposes can access the data.
- 2) Researchers requesting access to the scientific-use files have to submit a detailed description of the intended use of the data and the analyses and accept a user contract, before their request is approved. Use of the data will be limited to the purpose of the proposed project and the data will have to be destroyed after completing the project.
- 3) The SNSF will control access to the scientific-use files on a case-by-case basis. Data is only released after the SNSF gave its approval and the contract has been signed (see point 4). A steering committee consisting of the SNSF internal Data Governance Group will evaluate the submitted requests (see also annex 1 for more details). The committee will base its decision on the quality and authenticity of the research proposal, on whether the CTC data is suitable to answer the research questions raised in the proposal, but not on the research question per se.
- 4) The applicant has to sign a physical contract with the SNSF before receiving access the scientific-use files (see annex 2 for the example contract).

4.2 Are there any necessary limitations to protect sensitive data?

The entire data sharing process (see section 4.1) complies with the Federal Act on Data Protection (FADP).

4.3 I will choose digital repositories that are conform to the FAIR Data Principles

yes

4.4 I will choose digital repositories maintained by a non-profit organization

yes