Data Management Plan (DMP)

1 Data collection and documentation

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

Our project will generate individual-level data of retired top-level athletes in Switzerland containing the following information: Socio-demographic information (e.g., gender, age, occupation, educational level, and income), psychological profile (e.g., athletic identity, resilience, and self-esteem), and other variables (e.g., general health, reason of the career termination, sporting success).

Most of the data will be typical survey data (numerical data) and only very few parts will consist of text data (e.g., from open survey questions). Therefore, the raw data will be formatted as UTF-8 CSV-files. For further analysis those files will be imported into IBM SPSS, R, and Microsoft Excel (see point 1.2). All the data will be collected by the projects own online surveys. We anticipate that the data produced in all categories will not exceed 500 MB.

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

The project will collect data from around an estimate of 1000 top athletes in Switzerland (e.g., members of national teams) using online questionnaires. The participants will be fully informed about the study (what data will be collected, what will be done with the data, and who will have access to them; where and how long they will be stored; when and how they will be deleted).

Data quality of the survey will be ensured by standard measures (i.e., validated questionnaires), and analyses of item and scale reliability after the data collection. Additionally, we will pre-test the whole questionnaire with selected athletes in order to receive feedback for the questionnaire.

We will use LimeSurvey for conducting our online-surveys. The information provided by each participant will be downloaded from LimeSurvey in a CSV-file and will be saved on our own institutional servers (see point 3.1).

For all three data measurement points, we follow the same standardized protocol regarding naming and coding of variables, screening and cleaning the data. The data analysis will be made using different statistical programs, notably IBM SPSS, R, ROPstat (Vargha et al., 2015).

All files will be named according to the corresponding measurement times and will be stored in a folder structure applying the same logic. The statistical analyses will be stored in a separate folder.

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

The data will be stored as UTF-8 formatted CSV-files (compatible with the majority of statistical software packages) and accompanied by a text file with metadata and codebooks. The metadata will contain information on study design, sampling methodology, fieldwork and information on additional variables derived from the raw data. The codebooks will contain variable-level detail, labelling schemes as well

as the number of answers and the frequencies of answer options. Both the metadata file and the code-books will be stored as text files. They will provide all information necessary for an accurate and effective secondary analysis of the data.

2 Ethics, legal and security issues

2.1 How will ethical issues be addressed and handled?

National regulations as well as all ethical standards recommended by the Declaration of Helsinki will be adhered to in the present study. Prior to data collection, approval will be obtained from the Ethics Commission of the Faculty of Human Sciences at the University of Bern. In addition, in the beginning of the data collection the participants will provide their informed consent (see also 1.2). Amongst others, we guarantee that all personal data will be published anonymously and treated confidentially.

Most of the data generated by the project contain sensitive information about the athletes (i.e., personal data about income, health, self-esteem, etc.). Therefore, those data are handled with particular caution. Our study is facing great challenges concerning the protection of the personal data in general, and the athletes' anonymity and confidentiality in particular. By nature of the collected data, a complete anonymization is impossible due to information unique to an individual athlete (e.g., sport domain, ranking, gender, and age). This unique information allows for identification of individual athletes due to supplementary information available in the media and internet (e.g., who won the gold medal in a specific sports domain at the Olympic Games in 2016?). Moreover, due to the finite and publicly known population (i.e., top athletes who end their careers), it is very easy to establish a link between athletes and study participants. Nevertheless, the names of the athletes will be replaced by a code (see 2.2). In publications the data will only presented in aggregated form and all personal details will be removed or made anonymous, which prevents drawing any conclusions about specific individuals.

2.2 How will data access and security be managed?

All data will be stored in a folder on internal server of our institute. Because of the high sensitivity of the data and the severity of a breach in security, only the project working team will have access through a password. The involved individuals have to sign a contract about discretion. During the whole project, a limited number of just four people will have access to the data with a password. The list of the names and corresponding codes will be stored in a separate password-protected file.

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

The University of Bern is the owner of the data. The intellectual property belongs to the principal investigator of this research proposal. We will not use any third-party data nor a CC license, as the data will not be shared (see point 4.1).

3   Data storage and preservation

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

The project datasets will have a manageable seize of less than 100 MB per dataset. As the collected data capture highly sensitive information, there is a need for a special IT infrastructure for the duration of the project. The data will be stored on dedicated password protected servers of the involved university institute. These servers are secured in an identical manner as all other university servers against digital or physical threats (firewalls, automated backups every 24 hours).

3.2   What is your data preservation plan?

As already mentioned, the datasets of this project are especially sensitive with regard to their content. Therefore, no data will be shared with one of the established data repositories (see point 4.1). With regard to the long-term storage of the data after completion of the project, all the data will be saved as UTF-8 formatted CSV-files on the university's campus archive (long-term storage). Additional information like metadata and codebooks will be saved as text files (compatible with any text editor software). We will preserve the data for 10 years on the institutional server. The principal researcher will be responsible for the data until his retirement on 31 July 2024, thereafter and until the expiry of the 10-year period the Managing Director of the Institute will be responsible for the data.

4   Data sharing and reuse

4.1   How and where will the data be shared?

In view of the sensitivity of the collected data, the all-inclusive datasets will not be shared on a public data repository. However, some data will be made available upon request. Therefore, the metadata of the dataset together with a contact e-mail and the conditions for access will be shared in a repository (i.e., Boris repository of our university).

4.2   Are there any necessary limitations to protect sensitive data?

As already mentioned, the data will not be made publicly available (see point 2.1). The investigators affirm and uphold the principle of the participants' right to privacy and complies with applicable privacy laws. Especially, anonymity and confidentiality of the participants will be guaranteed when presenting the data at scientific meetings or publishing them in scientific journals.

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

Yes.

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

Yes.