1. DATA SUMMARY

1.1 Purpose of data collection What is the purpose of the data collection/generation and its relation to the objectives of your project? What is its relation to the objectives of euCanSHare? To whom might the data be useful ('data utility')?

1.1 Description of data What is the origin of the data? (origin (source project and funding) What is the size/ expected size of the data? (number of variables and subjects, volume in MB/GB/TB) What data types and formats of data have you generated/collected / will you generate/collect during the project? (Table 1)

1.1 Data security levels, confidentiality and potentially disclosive personal information For each data / dataset type or specific variable types (and metadata) please specify the corresponding levels of data security (access level) (Table 1/ (Table 2)

2. FAIR data management

2.1 Data collection/ generation Will data be deposited through euCanSHare Web Portal or through an external submission portal (Table 1)

If data is deposited through external submission portal (not through euCanSHare Web Portal): Specify the methodology of data deposition for user Define protocols for depositing new cohort raw data into the appropriate repositories and methods to provide rich metadata to foster a quality reuse of raw data for newly coming research projects.

2.2 Data storage Will data be deposited in euCanSHare centralized repositories or external repositories? (Table 1)

If data is deposited in external repository (not deposited in the euCanSHare centralized repository): Where will your data be stored? What are the provisions for long-term secure storage during and after the ending of the project?

2.3 Making data findable, including provisions for metadata Will you provide metadata for discoverability? Will data will be identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)? What naming conventions do you follow? (e.g for datasets names: Data_<WPno>_<serial number of dataset>_<dataset title>.) Will search keywords be provided that optimize possibilities for reuse? Do you provide clear version numbers? What metadata have been/will be created? (example: metadata standard Darwin Core). In case there is not a metadata standard for your variables, please outline what type of metadata will each of your data types contain.

2.4 Making data openly accessible How will the data be made accessible? (e.g. by deposition in the euCanSHare centralized repository /by links to an external repository)

· If data is deposited in the euCanSHare centralized repository: What data, metadata would be already visible from the platform (level 1) (for each> raw data, variable names, variable names and number of data available per variable, variable aggregated statistics?) (Table 2) Which data (data types, variables) will be accessible under an authenticated level (level 2)? (Table 1 and Table 2) Which data (data types, variables) should be accessible under a controlled-access level (level 3) (Table 1 and Table 2)

· If data is deposited in external repository (not deposited in in the euCanSHare centralized repository): What methods, technology/software tools are needed for access/ transfer of data? Will you include the relevant software (e.g. in open source code) and its documentation? How will the data and associated metadata be accessible? (levels) (Table 2) Have you explored appropriate arrangements with the identified repository? If there are restrictions on use, how will access be provided? Will you use EuCanSHare?s data access committee? If not, describe conditions for access to your data? Will you use EuCanSHare?s ADA-M as data access manager interface? If not, describe access (i.e. a machine-readable license)?

Will the metadata (and variable description) be made accessible by deposition in the euCanSHare centralized repository - centralized Mica-server)? · If metadata is deposited in external repository (not deposited in the euCanSHare centralized repository (EGA)- or centralized Mica-server): Define how links to metadata in external repositories will be provided.

2.5 Making data interoperable Are your data interoperable, that is allowing data exchange and reuse between researchers, institutions, organizations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)? What metadata you consider as ?required metadata? for each data type you handle (minimal required metadata - think of it as the metadata required for between cohort comparison and ultimately multi-cohort analysis) ? (Table 2) What data and metadata vocabularies, standards or methodologies do you follow to make your data interoperable?

Do you use standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability? In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

2.6 Increase data reuse (through clarifying licenses) Answer if data is deposited in external repository (not deposited in in the euCanSHare centralized repository) How will the data be licensed to permit the widest reuse possible? If it isn't currently, when will the data be made available for reuse? Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the reuse of some data is restricted, explain why. How long is it intended that the data remains re-usable? Are data quality assurance processes described?

3. ALLOCATION OF RESOURCES Answer if data is deposited in external repository (not deposited in in the euCanSHare centralized repository) What are the costs for making data FAIR in your project? How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions). Who will be responsible for data management in your project? Are the resources for long term-preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

4. DATA SECURITY Answer if data is deposited in external repository (not deposited in in the euCanSHare centralized repository) What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)? Is the data safely stored in certified repositories for long term preservation and curation?

5. ETHICAL ASPECTS Are there any ethical or legal issues that can have an impact on data sharing? Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?

6. OTHER ISSUES Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?