| | |
|---|---|
| **Aim of the Meeting:** | Gathering data overview info for euCanSHare's Data Management Plan v1 |
| **Meeting Called By:** | EGA-CRG (WP3) |
| **Date, Time, and Place:** | April 24 2019 10:00AM GMT, TC. |
| **Type of Meeting:** | Organizational |
| **Attendees:** | Tarja Palosaari (TLH), Teemu Niiranen (TLH), Kari Kuulasmaa (TLH), Ari Haukijärvi (TLH), Claudia Vasallo (EGA-CRG), Audald Lloret-Villas (EGA-CRG) |

**Agenda:**

1. Data summary
    1.1. Data purpose, origin, nature, collection procedures
    1.2. Data types, data formats
2. Data discoverability
    2.1. Naming conventions, metadata standards
    2.2. Approach towards keywords, raw data overview (aggregate statistics)
3. Data accessibility
    3.1. Levels of access
    3.2. Access procedures and conditions
4. Data interoperability
    4.1. Data and metadata vocabularies, standards or methodologies
    4.2. Required metadata
5. Data reusability
    5.1. Data utility, intended degree of reuse
    5.2. Safe long-term storage

**Outline:**

Following Horizon 2020's guidelines on Data Management Plan (DMP) we aim to a collect a description of data handled within euCanSHare, along with a comprehensive analysis of the nature of data on a dataset by dataset basis, including the needs for long-term storage and security issues, required metadata and other requirements for interoperability, as well as compliance with FAIR and EOSC data principles. The DMP can be considered as a reference for the resource and effort allocations related to data management. However, the DMP is intended to be a living document and can be modified and updated throughout the project.

*Data description.* DMP should include info on data origin, such as data obtention (collection, submission process) as well as the nature of data, in terms of data types classifications, size (and volume in MB/GB/TB) and formats of datasets, all of which will have implications in terms of storage, technical requirements and access.

*Data discoverability*. DMP should include the data owner's approach to unique identification for datasets with globally unique and eternally persistent identifier, versioning numbers, naming conventions, search keywords and metadata standards as well as metadata accessibility and data access levels, including data overview or aggregate statistics, all of which will facilitate discoverability of cohort data, comparison and selection of data of potential interest for users.

*Data accessibility.* DMP should address how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of accessibility conditions.

Access policies regarding each cohort, dataset or specific data types or variables should be gather in advance to implement the corresponding provisions for data accessibility and discoverability of data. These access policies (in close collaboration with WP1) will be defined and stored in a centralized platform to provide a simple framework to apply for data access, facilitate the procedure of granting and managing granted credentials by the committees (DACs tools). Also, if a different access methodology (e.g local, machinery licenses, novel blockchain technologies) will be implemented by any data owner, a description of this would be required for the DMP.

*Data interoperability.* DMP should include a description of data and metadata vocabularies, standards or methodologies followed by each data contributor to facilitate interoperability. To this end, an initial collection of this info from each data contributor will be valuable. Ideally, the collection about required metadata specific for each data type from data contributors/ experts will lead to a consensus on minimal required metadata collection for each data type, crucial for harmonization purposes, comparison and ultimately multi-cohort analysis.

*Data reusability.* DMP should address reusability issues such as data utility, the intended/allowed degree of reuse (marked by the establishment of licenses), data quality assurance processes and data safely storage in certified repositories for long-term preservation and curation.

*Data security.* DMP should describe the provisions for data security, including storage backup, long term preservation. For those cohort data stored locally in external repositories, storage capacity and provisions for long-term secure storage of data should be expressed by data custodians.

To go through these issues we would like to follow more or less a questionnaire based on H2020 guidelines https://github.com/clauw87/EGA/blob/master/questionnaire_data_providers_lite.docx in order to fill a table gathering most of these info https://github.com/clauw87/EGA/blob/master/DMP_Table_1.xlsx .