|  |  |
| --- | --- |
| **Roles**  | **Functions** |
| **Process owner** | Manager Lifelines Research Office  |
| **Author** | Data Manager Lifelines Research Office |
| **Review** | Manager Lifelines Research OfficeData Manager Lifelines Research OfficeStaff member Lifelines Research Office |
| **Authorisation** | Director Lifelines |
| **Document Control** | Staff Assistant Finance & Contracts Lifelines Research Office |

**Content**

[1. Aim 2](#_Toc17195049)

[2. Area of Application and Process relation 2](#_Toc17195050)

[3. Definitions and abbreviations 2](#_Toc17195051)

[4. Process flow 3](#_Toc17195052)

[5. Process description and actions 4](#_Toc17195053)

[5.1 Review UGLI proposal 4](#_Toc17195054)

[5.2 Rejected UGLI proposal 4](#_Toc17195055)

[5.3 Approved UGLI proposal: DMTA and code of conduct 5](#_Toc17195056)

[5.4 Grant Data Access 5](#_Toc17195057)

[Project folder access 5](#_Toc17195058)

[Release folder access 5](#_Toc17195059)

[5.5 Revoking access 6](#_Toc17195060)

[5.6 Support data access 6](#_Toc17195061)

[Questions on using the cluster 6](#_Toc17195062)

[Questions on phenotypic data 6](#_Toc17195063)

[Questions on genetic data 6](#_Toc17195064)

[Questions on cluster infrastructure 7](#_Toc17195065)

[Questions on access control 7](#_Toc17195066)

[Data availability from the Lifelines workspace to the cluster 7](#_Toc17195067)

[5.7 Backup/Restore Data 7](#_Toc17195068)

[Appendix A: Cluster folder structure 8](#_Toc17195069)

[Appendix B: RACI UGLI data access 10](#_Toc17195070)

1. Aim

This process describes the steps involved to give a researcher access to the UGLI-data as defined in the approved UGLI research proposal. UGLI consortium members receive temporary exclusive right to use the data for a period of 3 years. The 3-year period starts when UGLI-data has been made available to researchers. After 3 years, UGLI consortium members must submit applications for the use of the data generated within UGLI via the regular Lifelines application procedure. This SOP therefore applies to the 3-year period.

|  |  |
| --- | --- |
| **KPI** | **Standard** |
| Lead time between signed contracts and Data Access granted to the researcher | Genotype and phenotype data: Maximum 5 working days lead time between signed DMTA and code of conduct and data access granted to the researcher |

1. Area of Application and Process relation

The area of application of this process is the Lifelines Research Office, the UGLI-steering committee (UMCG SC) and the Genomics Coordination Centre (Team GCC).

1. Definitions and abbreviations

|  |  |
| --- | --- |
| UGLI-SC | UMCG part of the UGLI-steering committee. Contact person and head of this steering committee is Harold Snieder. Lifelines is the other 50% of the UGLI-steering committee. |
| HPC environment | A High Performance Computing Linux working environment, also known as ‘cluster’. UGLI-data and Lifelines phenotypic data from first and second assessment will be made available to researchers in the UMCG HPC environment |
| Backup data | UGLI data and phenotypic data will stay available on the UMCG High Performance Computing (HPC) environment and will not be archived. Files, scripts and syntaxes from project home directories on the HPC will stay available and will not be archived. An automated back-up is performed on all data on a daily basis with the purpose that (part of) the data can be restored if needed. |
| Data Access  | Process of giving a researcher access to UGLI-data. Access can be granted and revoked. |
| GCC | Genomics Coordination Centre (UMCG/RUG). |
| Genotype dataset | Dataset with genetic data of Lifelines participants derived from collected blood samples. |
| Phenotype dataset | Dataset with phenotypic data of Lifelines participants derived from interviews, laboratory assessments, measurements and questionnaires. |
| Link data | Matching datasets by using the same participant identifier. |
| UGLI Research Proposal | Proposal to request access to UGLI genotype dataset and Lifelines phenotype dataset from first and second assessment with a motivation and description of the requested data. The research proposal is written by the researcher and has to be approved by the Lifelines Research Office (see Process Management of Proposals) and by UMCG-SC. Further procedure of the application will only follow when both parties approve the proposal. |
| Restore data | Process of restoring (part of) backup data to the UMCG HPC environment. |

1. Process flow



1. Process description and actions
	1. Review UGLI proposal

**Actions:**

* When a UGLI proposal has been submitted, always send this to Lifelines data management: data@lifelines.nl.
* Check whether proposal has been submitted to UGLI-SC. If not, send the proposal to Harold Snieder for review by UGLI-SC. Replies by Harold Snieder should always be forwarded to data@lifelines.nl.
* Lifelines data manager/ Lifelines project manager or research office manager AND UGLI-SC review proposal.
* When UGLI proposal has been approved, continue with **5.3 Approved UGLI proposal: make DMTA and code of conduct**.
* When UGLI proposal is rejected, reply to researcher with review outcome, see **5.2 Rejected UGLI proposal**.

**Work documents:**

* Procedure IV/OV Aanvraag (iDocs 40005)
* SOP Procedure Review (iDocs xxxxx)
* Review Form Research Proposal (iDocs xxxxx)
	1. Rejected UGLI proposal

An UGLI proposal can be rejected due to a number of reasons. These reasons have to be explained to the researcher in the Review outcome email.

**Actions:**

* If proposal has been rejected because researchers are not linked to UMCG, use the following text in the review outcome email:

*A number of conditions to get access to the UGLI data have not been met with regard to your application. UGLI data will be made available on the HPC (Linux environment) of the UMCG and therefore the data can only be accessed by researchers that are affiliated with the University Medical Centre Groningen (UMCG), since accessing the HPC requires an UMCG account. The data will not be accessible through the Lifelines workspace. In the future the UGLI data will also be made available to non-UMCG researchers, though the timeline for this is unclear. Another requirement is that you have to be part of the UGLI consortium to gain access to the UGLI data. If you want more information about this consortium and the possibility to get access to the data once it is available, please contact the UGLI consortium contact person Harold Snieder (**h.snieder@umcg.nl**).*

* If proposal has been rejected by UGLI-SC, use the reasons given by UGLI-SC in the review outcome mail
	1. Approved UGLI proposal: DMTA and code of conduct

**Actions:**

* The Lifelines data manager makes an DMTA and code of conduct. These documents are send to the researchers.
* When applicant researcher requests access to a release folder of another project group, inform researchers responsible for requested release folder (see also **5.4 Grant data access**)

**Work documents:**

* DMTA (versie 7) (iDocs: 27722)
* Code of conduct (iDocs: xxxxx)

* 1. Grant Data Access

The researcher is granted with data access rights to the UGLI data on the UMCG HPC. For information about folder structure on HPC see Appendix A. For schematic overview of folder structure see figure 1 in Appendix A.

Project folder access

**Actions:**

* The Lifelines data managers contact the researcher to determine the project group/folder name.
* Lifelines data manager sends request to Cluster Administrators from GCC for creating a project folder for the researcher on the cluster, giving it the name that was discussed with the researcher. The Cluster Administrator will add the Lifelines data managers and the researcher to the group[[1]](#footnote-1).

Release folder access

**Actions:**

* When researcher has final results, the researcher will contact Lifelines data managers and will indicate which dataset is available for release.
* Lifelines data manager checks data/content and metadata of the dataset that is to be released
* Appoint responsible researcher (for questions from others)
* The Lifelines data manager contact the researcher to determine the release group/folder name.
* Lifelines data manager sends request to Cluster Administrators from GCC for creating a project folder for the researcher on the cluster, giving it the name that was discussed with the researcher. The Cluster Administrator will add the Lifelines data managers and the researcher to the group
* The researcher is responsible for transferring data from *project folder* to *release folder* and for creating / updating a README file for use of data
* When other researchers request access to the released datasets, contact GCC to add these researchers to the release folder
	1. Revoking access

As soon as a researcher is no longer part of a research, their access to project folders and release folders should be revoked. If the researcher is the contact person for a dataset that has been released, the corresponding information should be updated and a new researcher should be appointed.

**Actions:**

* The researcher has to indicate that he/she is no longer participating in research
* Lifelines data manager contact GCC to revoke access of the researcher to project and release folder(s)
* Researcher has to appoint a new contact person who will take responsibility for the release file (for instance, the main applicant)
* Lifelines data manager verifies new contact person for data in the corresponding *release folder*
* *The researcher* needs to update README file for use of data when a new contact person has been appointed, if applicable
	1. Support data access

Questions on using the cluster

Researchers that have questions on using the cluster can always send these to Lifelines data management. However, researchers are required to have basic knowledge and proficiency in using a cluster to perform analyses on phenotypic and genetic data. Lifelines and the Genomics Coordination Centre are not accountable and/or responsible for educating the researcher in using the cluster or working on a Linux environment. Researchers can consult the wiki to learn about basic cluster commands: <http://wiki.gcc.rug.nl/wiki/GccStart>

Questions on phenotypic data

Questions on phenotypic data will be answered by Lifelines data management. The phenotypic release folder(s) will contain a README file describing how to get in touch with Lifelines Data Management.

Questions on genetic data

Questions on genetic data will be answered by the respective researcher responsible for the genetic release folder. Each release folder will contain a README file describing how to get in touch with the responsible researcher.

Questions on cluster infrastructure

Questions on cluster infrastructure will be answered by Lifelines data management in consultation with the Genomics Coordination Centre. Lifelines data management will be able to support researchers within the umcg-lifelines folder. Cluster-wide operations or moving data between folders on the same, or higher level than umcg-lifelines, will be performed by the Genomics Coordation Centre.

Questions on access control

Questions on access control for project folders and release folders will be answered by Lifelines data management in consultation with the Genomics Coordination Centre.

Data availability from the Lifelines workspace to the cluster

Researchers that would like to have their data moved from the Lifelines workspace to the cluster are free to do so. Because researchers can not download data from the Lifelines workspace, Lifelines data management will have to provide assistance. For older projects the data files should be pseudonymized to Lifelines Omics identifiers, or a pairing file should be made available. Please note that the requested data should be placed in the *project folder* of the researcher, and not in a *release folder*, unless specifically indicated.

**Actions:**

* The researcher indicates which data needs to be transferred from the Lifelines workspace to the project folder on the cluster
* Pseudonymized to Lifelines Omics identifiers, or provide the data with a pairing file
* Lifelines data manager moves the requested data from the Lifelines workspace to the project folder of the researcher on the cluster (not the *release folder*!)
	1. Backup/Restore Data

All userdata stored on the UMCG HPC environment and backed up daily. Data can be restored from backups, but data will stay available on the HPC environment after project has finished.

**Actions:**

* No actions required

Appendix A: Cluster folder structure

The UGLI data is stored within a folder on the cluster which can be accessed by all researchers who gain access to this folder. The UGLI data can remain in this folder and does not have to be copied to the researchers own personal folder, since researchers will not have authentication to change any files in the UGLI main storage folder. They can access the data for their analyses.



Figure 1: Schematic overview of project and release groups on the cluster. The researcher gains access to a project group and a project folder. The researcher uses the project folder as his/her work directory. More than 1 researcher can get access to a project folder as can be seen by the number of “R” appointed to a project folder. The researcher can access the UGLI data indirectly to do their analyses. Once the research has finished (or part of), results, datasets, etc are added to a release folder. These release folders can be shared with other researchers who would like to use the datasets that are available in these release folders. A project folder can use more than one release folder.

The researcher will be added to a project group and a release group. Within these groups, the researcher will get access to a project folder and a release folder, respectively.

A project folder is used by researchers to work on data they’ve received access to. It’s a working folder in which scripts and results from analyses can be stored. The folder can also be used to store preliminary datasets which are being prepared before becoming a release. A project folder has at least one researcher that is responsible for its content. Other researchers can be added to a project folder upon approval of the main researcher and/or principle investigator.

A release folder can be the result of a project. Via releases other researcher will be able to access the resulting dataset from a project folder, both as raw data and as processed data (including quality control). A release folder always contains versions on which group rights will be attributed. The researchers responsible for providing the released dataset can use their project folder to prepare a new release version, if applicable. The raw and processed data is separated in different folders to avoid data duplication because of versioning. The UGLI folder, in which the UGLI data is stored, for instance, is also a release folder.

Appendix B: RACI UGLI data access

Table 1: Overview of all steps of UGLI data access, who is responsible (R), has to be informed (I)or consulted (C) and who has to execute the action (A).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Responsibility** | **Lifelines data managers** | **GCC DM** | **Researcher** | **Principle investigator** | **Lifelines reviewing committee** |
| *Step 1: Review proposal* |  |  |  |  |  |
| Submit Lifelines Omics Application Form |  |  | R | A | I |
| Review Lifelines Omics Application Form | R |  | C / I | C / I | A |
| *Step 2: DMTA and code of conduct* |  |  |  |  |  |
| Inform researchers from requested *release folder* | R |  | C / I | C / I | A |
| Collect DMTA | A |  | I | R |  |
| Collect Omics Code of Conduct | A |  | R | I |  |
| *Step 3.1: Project group and project folder* |  |  |  |  |  |
| Determine *project group/folder name* | A | I | R | I |  |
| Create *project group* | A | R |  |  |  |
| Add researcher to *project group* | A | R |  |  |  |
| Add Lifelines data managers to *project group* | A | R |  |  |  |
| Create *project folder* | A | R | I |  |  |
| *Step 3.2: Release group and release folder* |  |  |  |  |  |
| Indicate dataset available for release | I | I | R / A |  |  |
| Check data/content of dataset | R |  | A |  |  |
| Check metadata of dataset | R |  | A |  |  |
| Appoint responsible researcher (for questions from others) | I | I | R / A |  |  |
| Determine *release group/folder name* | A | I | R | I |  |
| Create *release group* | A | R |  |  |  |
| Add researcher to *release group*  | A | R |  |  |  |
| Add Lifelines data managers to *release group* | A | R |  |  |  |
| Create *release folder* | A | R | I |  |  |
| Transfer data from *project folder* to *release folder* | I / C | I / C | R / A | I |  |
| Create / update README file for use of data | I | I | R / A |  |  |
| **Responsibility** | **Lifelines data managers** | **GCC DM** | **Researcher** | **Principle investigator** | **Lifelines reviewing committee** |
| Add other researchers to release folder  | A | R | I / C | I |  |

Table 2: Overview of all steps of revoking UGLI data access, who is responsible (R), has to be informed (I)or consulted (C) and who has to execute the action (A).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Responsibility** | **Lifelines data managers** | **GCC DM** | **Researcher** | **Principle investigator** | **Lifelines reviewing committee** |
| *Revoke access* |  |  |  |  |  |
| Indicate that researcher is no longer participating in research | I |  | R | A |  |
| Revoke access from *release groups* | A | R | I | I |  |
| Revoke access from *project groups* | A | R | I | I |  |
| Verify contact person for data in the corresponding *release folder* | A |  | C |  |  |
| update README file for use of data, if applicable | I | I | R / A |  |  |

Table 3: RACI with questions related to data access asked by researchers and who is responsible (R), who is responsible (R), has to be informed (I)or consulted (C) and who has to execute the action (A).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Responsibility** | **Lifelines data managers** | **GCC DM** | **Researcher (responsible for release folder)** | **Researcher (user)** |
| *Support data access* |  |  |  |  |
| Questions on using the cluster | C | C |  | R / A |
| Questions on phenotypic data | R / A |  |  | I |
| Questions on genetic data | A |  | R | I |
| Questions on cluster infrastructure (within umcg-lifelines) | A / R | C |  | I |
| Questions on cluster infrastructure (cluster-wide) | A | R |  | I |
| Questions on access control (group management for folders) | A / R | C |  | I |
| Questions on access control (adding/revoking access) | A | R |  | I |

Table 4: RACI who is responsible (R), who is responsible (R), has to be informed (I)or consulted (C) and who has to execute the action (A) for tranferring data from the Lifelines workspace to the cluster.

|  |  |  |  |
| --- | --- | --- | --- |
| **Responsibility** | **Lifelines data managers** | **GCC DM** | **Researcher (user)** |
| *Data transfer LL workspace to cluster* |  |  |  |
| Indicating data transfer from Lifelines workspace to *project folder* on the cluster |  |  | R/A |
| Moving data from the Lifelines workspace to the cluster | R |  | A |

1. As long as Lifelines data managers do not have access to the IDVault [↑](#footnote-ref-1)