Data Management Instruction GROMACS -1.1.1

<table>
<thead>
<tr>
<th>Name</th>
<th>GROMACS 2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acronym</td>
<td>GROMACS-1</td>
</tr>
<tr>
<td>Research group</td>
<td>KULTURA</td>
</tr>
<tr>
<td>Room</td>
<td>106</td>
</tr>
<tr>
<td>Work mode</td>
<td>1</td>
</tr>
<tr>
<td>Instruction owner</td>
<td>Lindsay Oakley</td>
</tr>
<tr>
<td>Last modification</td>
<td>2019-12-11</td>
</tr>
</tbody>
</table>

General description
GROMACS is a versatile package to perform molecular dynamics, i.e. simulate the Newtonian equations of motion for systems with hundreds to millions of particles. It is primarily designed for biochemical molecules like proteins, lipids and nucleic acids as well as polymers.

Software
Not applicable as GROMACS itself is a type of software.

Data
The data collected are atomic positions and trajectories. Internal algorithms also calculate system parameters such as volume, pressure, energy, etc.

Data is recorded in portable binary trajectory files (.trr, .tng) and human readable logfiles (.log). Depending on the number of atoms and the frequency of system snapshots the .trr file size can vary up to a few GB.

Post processing of the trajectory files can generate movies which also are usually on the order of 10-100MB depending on compression.

Metadata
Log files contain extensive header information about the GROMACS version, hardware, and initial input and simulation information as well as calculation performance and progress updates.

Data organization
Files are stored in the working directory for each project on the computing cluster. Each project has a dedicated folder. Internal structure of the folder is up to the project lead. Filenames include the type of model and relevant parameters for identification.

Data security
Access to the cluster is controlled by ACK Cyfronet AGH. Operators have their own accounts and can adjust read/write permissions for the files they create.

Produced data sets do not contain sensitive or personal data.

Data quality
Trajectory files require associated input and log files in order for simulations to be fully reproducible.
Software is updated regularly.

**Intellectual property**
Intellectual property management will be conducted according to the Institute IP Management Policy (“Regulamin zarządzania prawami autorskimi”).

**Data repository**
There is no specialized repository for GROMACS data.

**Backup policy**
Data generated on Cyfronet clusters is automatically backed up on the Storage Area Network. This SAN is characterized by high performance and availability, implemented by ensuring redundancy of devices and access paths as well as reliable power supply. The solution implemented in ACK Cyfronet AGH is resistant to failure of a single switch, RAID controller in the disk array, disk loop and power supply failure. Data can be transferred to an external hard drive or cloud via SFTP protocols and further archiving options on through Cyfronet are available on request.