MoSD meeting 1 - Notes

- Many different data types (working groups on algae, bacteria, etc.), metabolism data from the metabolomics working group
- data from GC-MS and LC-MS is basically similiar, even though some people might contradit that statement
- NMR data is diverse
- Sequencing data can come from two methods:
  - Genomics (DNA sequencing)
  - Transcriptomics (more complex)
- data structure: mostly strings consisting of (A,T,C,G)
- some image data (images of petri dished, plants, etc.), but these are the minority
- Brief history BioChemSys:
  - second funding period (4 years, 2018–2022)
  - in this time, the data management system shall be rebuild
  - infrastructure: one main server, two backup servers in different power networks
  - some labs use online repositories for their data mangement
  - not every working group participates in ChemBioSys
- up until now no existing data management plan (`did i get that right?`)
- aim of the platform: show the advantages of a data management system, upscaling is planned
- ISA:
  1. Investigate
  2. Study
  3. Assay

  - no data format restriction by ChemBioSys
  - metadata is not standardized (yet)
  - ISAcreator
  - up until now metadata is mostly text based (.csv, .txt, .docx)

  - the usage of ISAcreator is not part of the data management, but the working groups are responsible to implement it in their workflow (`did i get that right?`)
  - eventually, ISAcreator-files will be converted to JSON

- Denodo, a platform to publish data
- new trend: journals for data publishing, for example Journal of Physical and Chemical Reference Data
- we have two options regarding data access:
  1. master-account at ChemBioSys
  2. create account at fair-dom (read only), the data will be the same (`did i get that right?`)
- seek is mostly based on ISA
- maybe migration to [`i didn't get the name of the platform`]