5th meeting minutes 24/10/23 15:00 – 15:30

* What has been completed this week.
  + Accessed and explored dataset
  + Created subset of the first 1000000 entries
  + Started researching how to create knowledge graph
  + Subset only holds key info: pmid, published date, mesh tag names
  + Started trying to create smaller subset of knowledge graph
* Questions
  + Best way to create knowledge graph?
    - Itertools.combinations() will be useful for finding pairs
  + Can use cutoff of 2007 instead of 2010 if necessary, no problem
  + How to reduce dataset in a meaningful way?
    - Subset of most common terms
    - Create a scalable subset
    - Filter mesh terms by major/qualifier
* Next steps.
  + Build knowledge graph
  + Will build smaller one to visualize too
  + Look into filtering data
  + Implement scalable functions