Case 4: Professor Chartreuse

Professor Chartreuse is expanding his personal research to include more collaborative work with colleagues both here at JCU and at other institutions. His work focuses on the science of science, which is a very meta area of research that tries to better understand how researchers and scientists work together to generate new knowledge.

The data that Professor Chartreuse collects is mostly available from public domain resources. His most important collection comes from records published in a database called PubMed that indexes a wide variety of medical research. These records are available via a web API and are served in JSON. Thus far, Professor Chartreuse has used keyword searching to find the records of interest to him, and has transformed those JSON records into a series of large and complex Excel files. This has worked fine for Professor Chartreuse up to this point, but has already caused some headaches as he has tried to share his data with potential collaborators. The large volume of excel files has led to confusion as to which contains which type of data, and which is the most up to date. One of Professor Chartreuse's colleagues hopes to expand their dataset to include both records from other databases (such as Wikidata) as well as survey and interview data with scientists.

The data collected is used to write papers and create exercises for graduate classes. Professor Chartreuse also regularly visualizes portions of his data, both for substantive research reasons, and also in visualization demonstrations for his students.

Professor Chartreuse has a few specific issues that he mentioned would be helpful to include in a data management plan:

- He said he's interested in maybe consolidating his data, and has heard that something called an ER diagram or a schema might be helpful.
- He is concerned about how to make sure his data is consistent and protected. He once lost an excel file and all of the data associated with it, forcing him to go back and re-do a month of data collection.

- Professor Chartreuse has been gathering the public data that he uses by hand. He searches by keyword, and then types details from the returned results into his excel spreadsheets. He wonders if there's a better way for him to collect this public data.
- He is worried about how best to both share his data with collaborators, and also with other science of science researchers who might benefit from his curated datasets.