Case 4 - Professor Chartreuse

* Is the "science of science" considered part of social sciences? ( I need to know as to select the proper metadata standard for your research)
* "The science of science (SciSci) is based on a transdisciplinary approach that uses large data sets to study the mechanisms underlying the doing of science—from the choice of a research problem to career trajectories and progress within a field"
* What's the size of the data that you have so far? and how long do you think your research is going to last?
* "My data collection is about 20GB is size, and I plan on continuing until the day I am no longer here. "
* What type of metadata would you like to see happen with the files? Metadata is how you label your files. Example: title of the file, the date of the creation of the file. the subject of the file and so on.
* "I would like the metadata to include relevant information for me to be able to query it later on. So, definitely date, name, keywords, publications that used the data, the authors. Those a mandatory, if you think that you can add more, so that I can make more complex queries, let me know."
* What type of openness would you prefer to see for the data? For example, the database could be close to only your view or it can be open to everyone.
* "Since I collect my data from public sources (mainly Pubmed), I would like to keep their licensing and definitely make it available for non commercial use only."
* What data do you consider to be the most relevant in retaining?
* "Any type of data that will lead to a publication."
* Where is the data being stored?
* "Right now the data lives in Excel files in my hard drive."
* How do you organize your data in excel?
* "Simple excel files with a maximum of 3 worksheets."
* Is there a specific format structure in the way you store your data in your external hard drive, that you would like to keep (i.e Different Files structure)?
* "I store it by date. I have found that it makes it hard to go back and search for things. So I am not adamant in keeping it."
* Is there a specific budget range you are looking to maintain? Or spend no money at all?
  + “there is some money allocated for this. But if you can give me an option that would be free, I would be more than willing to take that.”
  + For how many years have you been collecting data? I am trying to gain a better understanding of how much data storage on average you require per year. Also, do you have data in your files that are from a previous area of study? ie. Is there data in your files that may not be necessary to include in a database centred around resources on the science of science?
  + I have been collecting data for four years. All the data is for this particular area of study
  + How far will the collaboration extend to, such as will you give colleagues access to editing your database if you were to adopt a database management system (i.e. SQL database)?
  + I would like for them to be able to view the data, but not be able to edit it. And they need to get permission to access it first.
  + Anyone of your intended audience (i.e graduate students, colleagues) can view the data at any time without permission, but need permission to access such as to use the data, like download it? Also, i know in the future there are plans for colleagues to add to the database, that means they will add to your own personal database? and they will have to be granted access to add to your database to do that? Correct me if I am wrong.
  + Yes, your assumptions are correct
  + I was hoping you could elaborate a little bit more on your issue with the keyword search. Why would you return the results by hand into an excel sheet?
  + I do it by hand because that is what works best for me and that's the way I know how to do it. It has nothing to do with PubMed or its infrastructure.