Case 3: Professor Pinkerton

Professor Pinkerton is an Excel wizard. She can make Excel do things no one else believes is possible. All of her research data ends up in Excel eventually. She has broad interests, but often ends up circling back to Management Information Systems-related work.

Professor Pinkerton collects data from a variety of sources using a variety of methods. Last year she conducted a systematic literature review and did a quantitative textual analysis; the data ended up in Excel. Two years ago she collected job descriptions for entry-level positions in her field; the complete text of each ended up in Excel. She has 12 years of student performance data from the courses she teaches, in Excel. She also loves data for the sake of data, and regularly saves interesting data sets for later use. Open data portals are her kryptonite; she can't stop downloading files that might one day be useful (she prefers Excel but accepts CSV). At last count she has a total of 17,384 spreadsheets saved to carefully organized folders on her laptop. The smallest has a single row of data; the largest has over 750,000 rows of data. The median size is about 1,000 rows of data (and 10 columns). The oldest spreadsheet is 9 years old and was created in an earlier version of Excel.

Professor Pinkerton is a well-known resource in her department; she is regularly asked about data files and provides copies when asked. Researchers in her area have started sending her their own spreadsheets in case she is interested, and her collection is rapidly growing. Professor Pinkerton has asked her dedicated postdoctoral fellow, Neil Gaiman, to build a folder structure somewhere in the cloud so she can use it to manage and share her spreadsheets.

Pinkerton hasn't used all of her spreadsheets, but she will one day; she is interested in answering research questions and writing papers based on the data. She also conducts annual data visualization seminars using one of her data files as a sample.

Of all of her spreadsheets, 95% are from external sources (other researchers, various governments, private corporations, etc.). However, Professor Pinkerton doesn't trust these sources to always provide this data into the future, so maintains local copies. People who work with Professor Pinkerton happily email spreadsheets back and forth. Pinkerton is willing in principle to share her collection of data, but is waiting for her postdoctoral fellow to finish with the folder structure in the cloud. In the meantime she is too busy to reply to all of the data requests she receives.