1/9/23
Call: Knowledge Management
Meeting Summary

Structure: Matthew Dorocak moderated a full group discussion and led a breakout room along with Meetal Desai and Lee Green. There was report back after breakout room discussion and then a subsequent group discussion.

Subtopics covered in this meeting include:

- How to decide what types of data to keep in terms of content and date range.
- How R&D data factors into overall data mining, how these decisions reached, and what goes into this decision-making process.
- How do these decisions change based on business unit or department?

Breakout Room Summary-
Question 1: How do you decide what types of data to keep in terms of content and date range?
- Participants agreed that, for the most part, everything is on the table, although data that has not been validated or used is often set aside. An ongoing participant project cited that their data and reports are being reviewed from the 1950s onward and there are almost 8000 documents that have to be sorted through.
- Separation of derived knowledge vs. raw data is a key distinction to make, as well- derived knowledge is retained forever whereas raw data is typically unexamined and can be seen as expendable.
- HR and Legal have different rules for what to retain which may apply across the business but often a business unit may make the decision independently. The goal across most organizations is to retain as much important information as possible but developing a process poses a challenge.
- In terms of people with institutional knowledge that are retiring, a few participants mentioned that trying to retain about 80 percent of their knowledge is crucial and allowing 20 percent to be potentially lost is an achievable goal. A standing default is an organization trying to keep proprietary data for 5 years and anything legal in nature is kept permanently- in fact, database systems had reminders to clear through data and free up space to avoid having any liability. There is a distinction between personal documents and data that has been uploaded into IT maintained systems- these two are treated very differently.

Question 2: Who is involved in deciding what data to keep?
- Legal departments or librarians were mentioned by several participants but typically, this is a top/down approach whereby leadership directs business units, and those managers are supposed to implement a process.
- A point was made about what data is kept if business units are being spun off or if there is a new business acquisition and it was agreed that this is a very arduous process where the legal department must take an active role in setting policy here.
- It is very important for companies to make decisions and set a process- being able to create a firm action plan for staff to easily follow is critical.

- Participants have looked at whether this data had been used recently to decide whether to log it for future use. The example of lab notebooks was given, and a few companies have already digitized theirs and added year and scientist to the notebook for easier usability.
- "We are always looking for hidden gems [of knowledge], but what constitutes a hidden gem?" A participant estimated that they have about 5 years before no one will know where a certain document is. Which makes it crucial to capture at-risk knowledge soon, and it should be determined if scientists that are doing similar work have all knowledge captured. Remote working has made this a more pressing issue- an example was given that if a box comes in and you need someone to review this box, your staff might be working from another state full time at this point, which makes the default reaction to just "pitch it". There was a question of is it even functionally possible to do what we are trying to do- if companies don't address this problem now, the entire exercise of collecting old data may be a moot point. A participant mentioned having a scientist retire who indicated that people may not care about her data as no one had asked about it in many years and this person simply shredded it.
- The moderator mentioned that a particular company reported their staff cannot close out a project until all key takeaways are written down and edited by a reviewer in their final report.

Collection systems:
- Companies have been truly trying to digitize for 5-10 years and anything new that has been set up once all these systems have gone in have stayed there and there has been a higher build up of data in these systems- this does not appear to be cost prohibitive as of late. What are reputable people that digitize and what is the latest technology that will carry us through- we are indeed, sitting on this mound because many companies can't find a supplier that they can trust- it really depends on what the data is, and one system does not fit all.
- Many systems are built on other preexisting systems, and it seems like there is no one size that fits all solution on the market. Some systems may be great for contracts and not work as well for R&D. These data mining processes typically take place once a month for physical notebooks and data but this is different for polymers and lab data- it's possible to search how many polymers have been made via a database and can fill in gaps in information- a participant reports this being an ongoing process which does take place regularly.
- It was acknowledged that there are a lot of errors that can take place when staff try to digitize handwritten notes. Scanning software is not sensitive enough at this time- how can meta data be enhanced instead of reliance on really drilling down to individual handwritten notes.

Summary statement "A Day in the Library will Save you a Year in the Lab."