**Streamlining Privilege Logging with Email Threading**

Alisa McLellan, Esq. and Brett Beeman

DTI

Recently, New York State’s Commercial Division adopted Rule 11-b[[1]](#footnote-1) that advocates for a categorical approach to privilege logging. Documents that fall within a privilege category may be logged together as opposed to the current method of individual log entries. This approach has been touted as a cost-saving improvement to creating privilege logs[[2]](#footnote-2), but does not go far enough in the use of available technology.

**Login Instead**

I recall fondly my early days working as a paralegal on “electronic” document review: when potentially privileged documents were blown back in banker’s boxes, and I manually logged each document into an Excel spreadsheet or Word document. The attorney whom I assisted would put a Post-It note on each document indicating the reason for privilege and would mark with highlighters what needed to be redacted. I would then apply those redactions on the electronic document. This paper-electronic hybrid model was time-intensive and did not fully utilize the benefits of technology.

Now, as a vendor project manager, I assist attorneys with setting up coding layouts in order to electronically create privilege logs. The reason for privilege is tagged in a coding field or typed into a notes field, and these fields along with necessary metadata are exported as an Excel spreadsheet. Duplicate documents can be grouped using MD5Hash for more efficient review and coding or redaction. Minimal manual clean-up may be necessary but can be mostly automated through Excel formulas. This method is far more technological than manual but still does not fully utilize available technology.

**Categorically Speaking**

Courts are beginning to recognize the time and cost associated with creating a privilege log. In both scenarios listed above, there is a review of each and every document that is being logged. Duplicates are identified in the latter scenario, and propagation of coding to duplicates will reduce the review population, but the reduction may not be significant. Also, the disposition of an item-level duplicate may change based on the treatment of its parent in the case of an email attachment, making propagation to duplicates a potentially risky shortcut.

Rule 11-b was created to reduce time and cost involved with privilege logging.[[3]](#footnote-3) While the proposed categorical approach might reduce the size of the eventual log, all documents still need to be reviewed in order to be categorized. Rule 11-b goes on to state that uninterrupted e-mail chains can be treated as a single document for purposes of listing such chains on a document-by-document privilege log. This is where things start to look interesting for document populations comprised of mostly email communications.

On the one hand, a categorical privilege log sounds like a major advancement. This log could be condensed and would be easier to use in regards to knowing why a group of documents were withheld and would show a type of relationship between documents that goes beyond duplicates or family. On the other hand, this type of log would not reduce the cost involved with reviewing documents.

**Last Cull**

An alternative to categorical logging is treatment of document by terminal (i.e., last-in-time) email(s) in a conversation, considered as a single entry rather than multiple individual entries. With the textual analytics tools on the market, grouping emails into conversations for review represents a well-adopted practice for ensuring first level review consistency because each successive email contains the content of the previous email and reduces the overall time-spend for reviewing a conversation. In the same way, the terminal email contains all of the email content necessary for logging.

The workflow for this process is most efficiently executed at the outset of the review: (1) identify the review population; (2) apply email threading; and (3) batch email conversations with attachments for review. Once the potentially privileged universe is identified, analysis of the population identifies a subset that represents the unique content to batch for privilege review and logging. Here we define unique content as an inclusive email, a unique attachment (based on MD5 or SHA-1 hash value) from each email conversation, and a unique loose file (based on its hash value). A more aggressive approach might involve negotiating with the receiving party to review and produce only the unique content; otherwise, this analysis can be limited to privilege review and logging only. All non-inclusive emails can be treated as duplicates with their basis for privilege claims represented as a part of the conversation’s log entry without redundant, individual logging.

Now, the argument against this alternative approach may be that the rule only states that ‘uninterrupted’ email chains can be treated as one document, and how often when dealing with privilege communication are the email chains ‘uninterrupted’? This, however, is just the beginning of the wildfire, a turning of the tides, and a recognition of a better way to use technology. Uninterrupted email chains may be the threshold today, and tomorrow all email chains that contain the same privilege communication may be proposed. A categorical, email threaded approach is the future – the cost savings in review alone would be worth it.

**Sample Results – Enron Data Set**

The ubiquitous Enron data set serves as an excellent example of how limiting review to inclusive email content only can drastically reduce the review population. From a population of 74,042 records, 45,730 non-inclusive emails can be removed from the review universe (see Appendix A). This represents a reduction of more than 60% of the content, not by search terms or predictive coding where the balance of precision and recall are competing goods. This reduction improves precision by 60% while maintaining a recall level of 100%. Certainly, gains realized from this process can vary greatly by collections, but any reduction is valuable, especially at scale for large data sets.

**Appendix A**



1. 22 NYCRR §270.70(g) (see http://www.nycourts.gov/RULES/comments/orders/AO-114-14.pdf). This categorical approach has also been promoted by the Seventh Circuit E-Discovery Pilot Program (http://www.discoverypilot.com/content/model-discovery-plan-and-privilege-order (Case Management Order)). [↑](#footnote-ref-1)
2. http://www.natlawreview.com/article/new-commercial-division-rule-seeks-to-streamline-privilege-log-requirements-new-york. [↑](#footnote-ref-2)
3. http://www.nycourts.gov/RULES/comments/PDF/PCPacketPrivilegeLogs.pdf. [↑](#footnote-ref-3)