



Why Understanding the Data is Key to Cost-Effective Processing

In today's economy, everyone has become highly cost-conscious. At the same time, the volumes of data being collected for eDiscovery grow ever greater. How can these two opposing trends be reconciled?

Data culling prior to processing has become a key weapon in the struggle to contain cost. However, intelligent culling depends critically on understanding the data. Those working on an eDiscovery project need to know as soon as possible:

- (1) The volume and type of any non-processable or system files.
- (2) The volume of data that falls within agreed parameters such as a date range or set of priority custodians or domains.

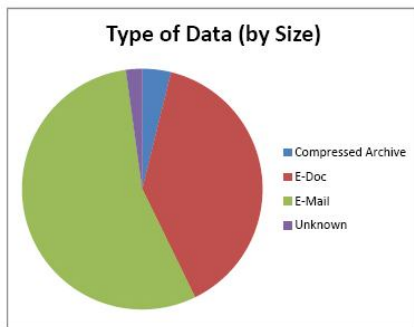
- (3) The amount of duplication within and between custodians and between email attachments and loose eDocs.
- (4) The probable percentage of privileged documents.
- (5) The probable percentage of responsive documents.

The first two items can be readily discovered by taking and analyzing an inventory of the files and emails prior to processing. A file level inventory will identify non-processable files e.g. music, video. An analysis of emails and eDocs by date will identify the volume falling within agreed date ranges.

Deduplication, while often one of the most effective ways to reduce data, can be a lengthy (and hence

costly) process. For large data sets, sampling a representative custodian set and analyzing the level of duplication within and between custodians and between emails and eDocs, can provide valuable input into developing an effective deduplication strategy.

A similar approach can be taken to assessing the proportion of privileged and responsive documents. Rather than indexing the entire data set (a lengthy process), representative subsets of the data can be sampled. Another approach is to run multiple searches across the entire data set, analyzing the results for single terms and complex compound queries to determine the most effective search strategy.



New for 2009

2009 is around the corner and LEX is preparing to open the New Year with extensions and additions to our service offerings.

In January we will be launching LEX iIdentify—a new Early Case Assessment and Intelligent Culling option designed to enable rapid, defensible early stage data reduction.

We're also upgrading our LEX Vault services to iConect 740 with Unicode support and, in Spring, integration with Microsoft SQL with additional custom reporting capabilities.

