Data Reduction

Narrowing a larger data set to a smaller data set for the purposes of review, based on a mix of objective and subjective criteria.

Data reduction methods include:

- Removal of immaterial items (e.g. empty files, folders, or other containers)
- Removal of known files (either using public lists such as NIST or project specific exclusion lists)
- Deduplication of exact or near duplicates
- Pre-filtering the data on key terms, custodians, file types, or date ranges

Data reduction requires carefully balancing the reductions of the quantity of documents to review, increasing the relevancy ratio of reviewed documents, and avoiding filtering out important documents.

If you are interested in this topic you can read our blog articles on Nuix Immaterial Items, the Grey Area or Deduplication Hidden.