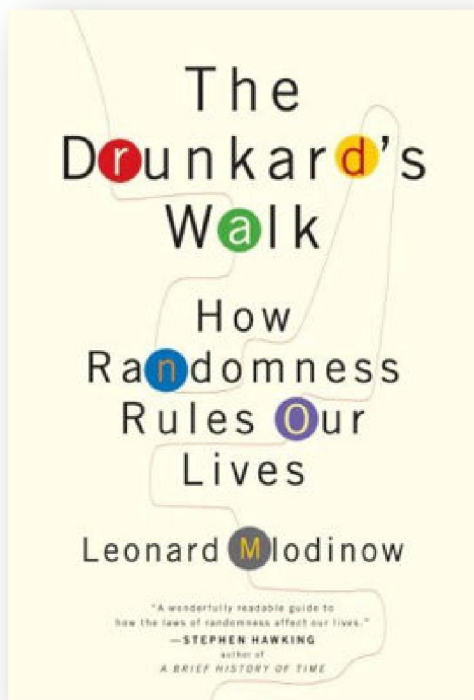


Drunks, DNA, and Data Transfer Risk in eDiscovery

"In law as in other realms, the understanding of randomness can reveal hidden layers of truth, but only to those who possess the tools to uncover them." Leonard Mlodinow¹



Over the summer I had a opportunity to begin reading a fascinating book, *The Drunkard's Walk – How Randomness Rules Our Lives*, by Leonard Mlodinow. Mr. Mlodinow received his doctorate in physics from the University of California, Berkeley, was an Alexander von Humboldt fellow at the Max Planck Institute, and now teaches about randomness to future scientists at Caltech.

In his book, Mr. Mlodinow shares an intriguing overview about the presentation of DNA evidence in criminal trials. In his discussion Mr. Mlodinow notes that DNA experts today regularly testify that the odds of a random person's DNA matching a crime sample are less than one in one million or one in one billion.² With those odds, Mr. Mlodinow notes, it is reasonable to think such a matching – if it occurs - may be beyond a reasonable doubt.

¹ Leonard Mlodinow, *The Drunkard's Walk – How Randomness Rules Our Lives* (Random House, Inc., 2008), p-40

² Ibid, p-36.

However, the interesting part of his overview on DNA centers around a discussion on *probability* and specifically on *what is not usually shared with the jury as part of such evidence*.

The *probability* being omitted as part of DNA evidence presentation is the *probability of human-based error* arising from labs making errors in both the collecting and handling of DNA samples and in the accidental mixing and swapping of samples. These human-based DNA sample transfer errors – which many experts put at 1% - significantly affect the probability of matching DNA - and thus can interject potential doubt in the minds of jurors where previously there may have been no reason for doubt based on the “half presentation” of the probability of a DNA match.³

After reading this overview, I began to wonder about the potential probability of human-based error in the transfer of data between the technologies and platforms available to conduct the complex tasks involved in electronic discovery.

While only a short overview of one risk factor and its probabilistic impact of interjecting error into the discovery process, I hope this overview provides legal professionals – both consumers and vendors alike – a new appreciation and consideration of the importance of integrated technologies and platforms in the conduct of electronic discovery.

Considering Data Transfer Risk

Electronic Discovery is a process that contains many complex tasks – tasks that in order to be accomplished accurately are reliant on the integrity of the data that they are acting on.

Just as there are many tasks in electronic discovery, many times in there are multiple technologies and platforms involved in the complete electronic discovery process. When there are multiple technologies and platforms involved, data must be transferred from disparate technologies and/or platforms to other disparate technologies and/or platforms. This data transfer can be considered a risk factor that affects the overall electronic discovery process.

Data transfer risk may be minimized by automation and standards or increased by the requirement of human intervention. As automation and standards are still slowly maturing in the realm of electronic discovery technology, it seems important that legal professionals understand and properly consider the impact on potential data transfer risk as they plan, source, and conduct their electronic discovery activities.

³ Extract from Google Books on DNA Overview In The Drunkard's Walk - <http://tinyurl.com/ma8wfm>

To Err is Human, but to Really Foul Things up Requires a Computer⁴

Since the error rate of data transfer between disparate electronic discovery platforms due to human-based error is difficult to measure, it appears realistic that courts would be extremely cautious in allowing human-based error arguments on this topic – unless such an error is one that is totally visible and documentable. However, many times human-based error may not be so readily visible or documentable. Because of this fact, seeking an understanding of probability of risk in this area seems a reasonable exercise any time disparate technologies and platforms are involved in the electronic discovery process.

Hypothesizing to Allow for Risk Comparison

Human-based error in transferring data between disparate electronic discovery platforms is difficult to estimate, but it seems reasonable to assert that it does occur. With that assertion in mind and to highlight the specific risk factor of *non-automated and integrated data transfer* (otherwise referred to as human-based data transfer), a reasonable hypothesis⁵ of this type of data transfer error might be to consider that at least one in every hundred – or 1% - of such data transfers interject human-based error into the electronic discovery process. The exact percentage could be adjusted for other reasonable estimates, but as one hopefully will see in this human-based data transfer risk assessment exercise, the human-based data transfer errors can be a major determinant and multiplier in relation to the overall risk of electronic discovery core tasks such as collection, analytics, processing, and review.

The Rule for Compounding Probabilities

*If two possible events, A and B, are independent, then the probability that both A and B will occur is equal to the product of their individual probabilities.*⁶

To evaluate the probability of human-based error in data transfer, we first need to determine where these data transfer points might occur in the electronic discovery process. In taking a high-level look at the electronic discovery process, and to simplify this risk assessment exercise, let us consider the following scenarios:

⁴ Farmers' Almanac, 1978. http://www.quotationspage.com/quotes/Farmers'_Almanac/

⁵ Hypothesis: an assumption or concession made for the sake of argument. hypothesis. (2009). In Merriam-Webster Online Dictionary. <http://www.merriam-webster.com/dictionary/hypothesis>

⁶ Leonard Mldinow, *The Drunkard's Walk – How Randomness Rules Our Lives* (Random House, Inc., 2008), p-33

Scenario 1: Traditional Electronic Discovery Approach (Marketing Level Integration)

While many times represented as “integrated” in marketing materials, the Traditional Electronic Discovery Approach consists of the use of different technologies and/or platforms for the electronic discovery tasks of collection, analytics, processing, and review. In the Traditional Electronic Discovery Approach data is collected and then must be transferred – with human intervention – at least three times prior to the preparation of data for production. (Figure 1)

Each of the transfer points in this approach may increase the potential for human-based error by a factor of 1%. When viewed in light of the Rule for Compounding Probabilities, the apparent human-based risk factor for this type of approach would be 3% or three in one hundred chances of human-based error.

$$\begin{aligned} &\text{Potential Human-Based Data Transfer Risk Error} \\ &1/100 + 1/100 + 1/100 = 3/100 = 3\% \end{aligned}$$

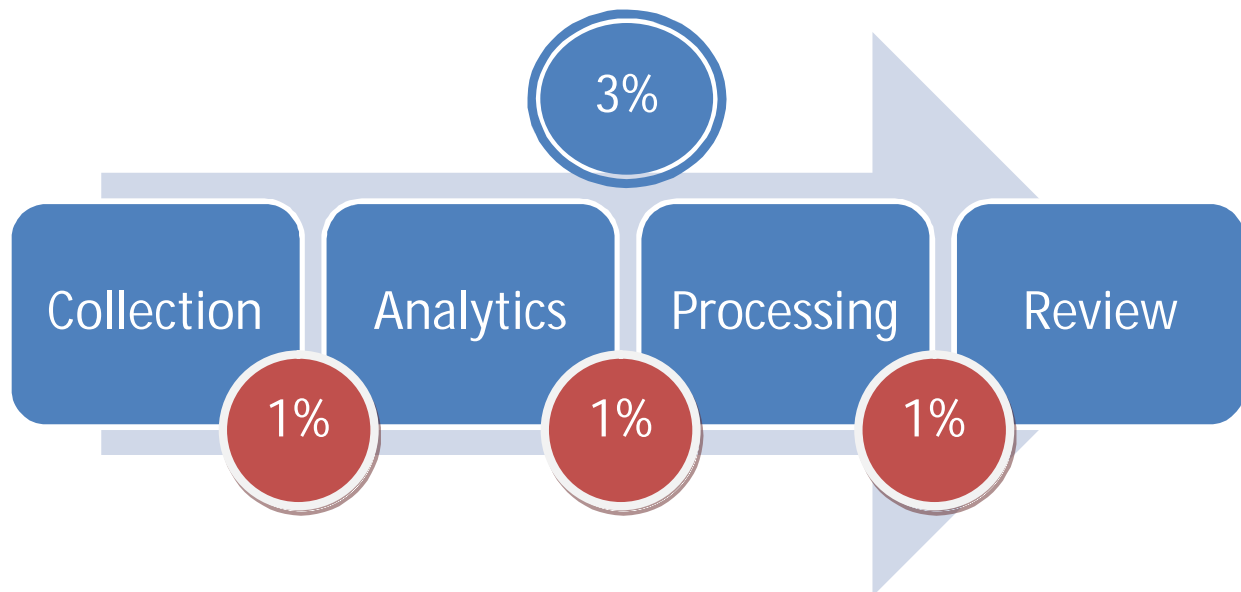


Figure 1 – Traditional Electronic Discovery Approach – Marketing Level Integration

Scenario 2: Quasi Advanced Electronic Discovery Approach (Platform Level Integration)

While many times represented as “integrated” in marketing materials also, the Quasi Advanced Electronic Discovery Approach consists of the use of different technologies and/or platforms for electronic discovery tasks, however this approach may combine two of the high level electronic discovery tasks into one technology or platform (example – the combining of analytics and processing in a single appliance). In the Quasi Advanced Electronic Discovery Approach data is collected and then must be transferred – with human intervention – at least two times prior to the preparation of data for production. (Figure 2) Within this approach there in fact may be some integration between two electronic discovery technologies and/or platforms, however there are still multiple points of human intervention required for data transfer.

As in scenario 1, each of these transfer points may increase the potential for human-based error by a factor of 1%. When viewed in light of the Rule for Compounding Probabilities, the apparent human-based risk factor for this type of approach would be 2% or two in one hundred chances of human-based error.

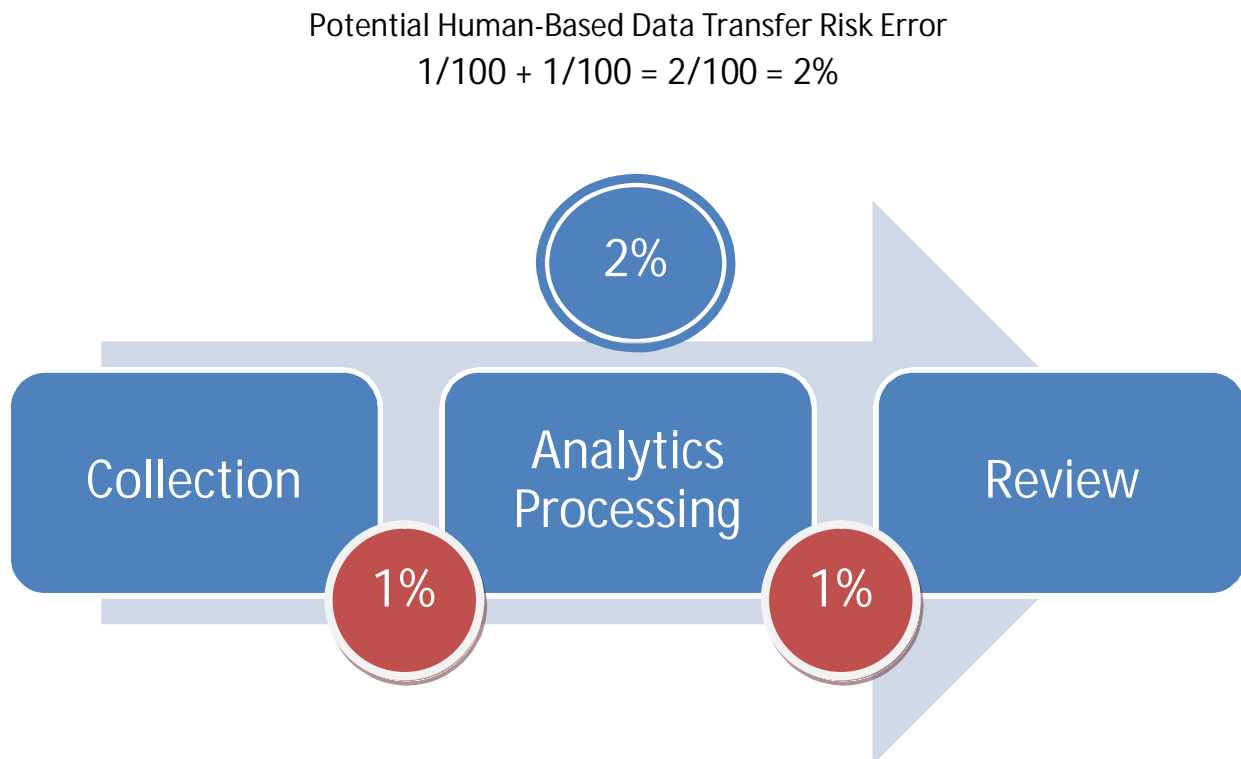


Figure 2 – Quasi Advanced Electronic Discovery Approach – Platform Level Integration

Scenario 3: Advanced Electronic Discovery Approach (Application Level Integration)

A truly integrated approach –meaning integrated at the application level – the Advanced Electronic Discovery Approach consists of the use of a single technology and/or platform to conduct the core electronic discovery tasks of analytics, processing, and review. The Advanced Electronic Discovery Approach requires no human intervention after the initial transfer of collected data.

As in previous scenarios, each transfer point may still increase the potential for human-based error by a factor of 1%. However, as there is only one apparent human intervention required prior to preparation of data for production, the human-based risk factor for this type of approach would be 1% or one in one hundred chances of human-based error. (Figure 3)

Potential Human-Based Data Transfer Risk Error
 $1/100 = 1\%$

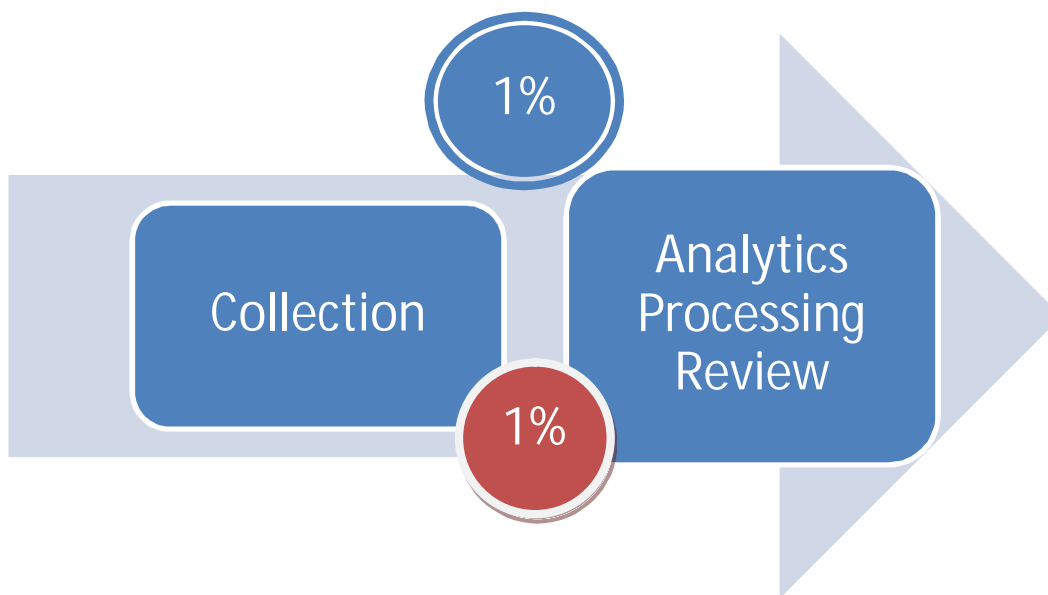


Figure 3 – Advanced Electronic Discovery Approach – Application Level Integration

The Risk Factor of Non-Integration

Comparing the potential for human-based data transfer risk - using the hypothesis that at least 1% of such data transfers interject human-based error into the electronic discovery process – it seems reasonable to assert that less human intervention during data transfer between technologies and platforms results in less potential overall risk for electronic discovery error.

Human-Based Data Transport Error Risk Factor Comparison

- *Level 3 Risk (3%) - Traditional Electronic Discovery – Marketing Level Integration*
- *Level 2 Risk (2%) - Quasi Advanced Electronic Discovery – Platform Level Integration*
- *Level 1 Risk (1%) - Advanced Electronic Discovery – Application Level Integration*

Considering Risk and Ethics

In all professional functions a lawyer should be competent, prompt and diligent.⁷

A lawyer shall provide competent representation to a client⁸.

Competent representation requires the legal knowledge, skill, thoroughness and preparation reasonably necessary.⁹

Understanding the potential risk factors associated with just this one aspect of electronic discovery, might it seem reasonable that one would seek to reduce human-based data transfer risk as much as possible – as it is a risk that can truly be reduced solely on the approach to electronic discovery one takes? Might there also in fact be potentially an ethical responsibility for electronic discovery professionals to reduce this type of known risk factor to the lowest level possible congruent with the resources available (time and money) for a specific audit, investigation, or litigation matter? These are questions that certainly warrant proper attention from legal professionals as they consider the best approach for their specific electronic discovery requirement.

Is there one single right approach? I might submit that the right approach is the one that properly balances resources available and risk and that also views any choice through the lens of what could be considered reasonable based on those resources and risks. While there may be no absolute *only acceptable choice*, there may be in fact an *absolute best choice* for approach given unconstrained time and resources.

⁷ ABA Model Rules of Professional Conduct, Preamble 1(4), (2004)

⁸ ABA Model Rules of Professional Conduct, Rule 1.1, (2004)

⁹ ABA Model Rules of Professional Conduct, Rule 1.1, (2004)

Beyond Data Transport Error

Yes, there are additional human-based risks in the electronic discovery process. And yes, each of these specific tasks (collection, analytics, processing, and review) may have multiple human-based risk factors – risk factors that can increase exponentially if data has to move back and forth between disparate technologies and platforms multiple times. But one irreducible fact appears to be that if one agrees that there is risk associated with human-based data transfer, and that the percentage of this risk is determinant by the number of times such human intervention in data transfer occurs, then it is imperative for legal professionals to understand the potential implications of such risk at the beginning of the electronic discovery process and reduce it to as low a level as possible. To do so is congruent with competent and reasonable preparation of data and this competence and reasonableness is critical to ensuring that electronically stored information (ESI) is of the highest quality.

*As a general rule, the most successful man in life is the man who has the best information.*¹⁰

Benjamin Disraeli

¹⁰ The Quotations Page - <http://www.quotationspage.com/quote/29221.html>