



# THE INVESTIGATIVE LAB:

A MODEL FOR EFFICIENT COLLABORATIVE  
DIGITAL INVESTIGATIONS

## CONTENTS

Executive summary .....	2
The digital forensic investigation impasse .....	3
Lessons from eDiscovery.....	4
Case study: Serious Fraud Office.....	4
The investigative lab workflow model for digital investigations .....	5
Share the workload —a production-line methodology .....	5
Ensure the right data gets reviewed by the right people.....	6
Use advanced investigative techniques .....	6
Perform deep forensics only when necessary.....	6

## EXECUTIVE SUMMARY

In this paper, we will discuss a model for setting up an investigative lab that allows digital forensic specialists, non-technical investigators and subject matter experts to collaborate on digital evidence. The end result is a dramatic increase in the volume and quality of digital evidence an investigative team can analyze within a fixed time.

For many years, digital forensic investigators have used specialist forensic tools to “dig deep” into a handful of computers and other evidence sources. Typically, investigations rely on a single digital forensic specialist to examine these evidence sources one by one.

This approach relies on the digital forensic investigator, who is usually not familiar with all the details or context of the case, to extract the information he or she thinks is relevant from each device. As a result, non-technical investigators and subject matter experts must rely on an incomplete and subjective slice of the evidence.

By contrast, the model discussed in this paper enables investigative teams to divide up digital evidence and spread the review workload between many people. They can distribute different types of evidence to the people most qualified to understand it and its context.

This methodology also provides opportunities to apply advanced investigative techniques such as data visualization and near-duplicate analysis, helping investigators look at the evidence from different angles. While in-depth forensic analysis remains an essential tool, this workflow limits its use to specific circumstances when it can deliver the most value.

By adopting a collaborative investigation model like the one in this paper, the United Kingdom’s Serious Fraud Office (SFO) increased the volume of data it could process each year 20-fold. This methodology also enabled the SFO’s investigation teams to respond much faster to information requests from courts.

The end result is a dramatic increase in the volume and quality of digital evidence an investigation team can analyze within a fixed time

## THE DIGITAL FORENSIC INVESTIGATION IMPASSE

Digital investigators face a constant battle to find the truth in ever larger, more varied and increasingly complex stores of electronic evidence. At the same time they must balance business demands such as reduced budgets and resources, spiraling case backlogs and ever decreasing timescales.

These demands are made all the more challenging because many digital investigations rely on workflows, processes and tools that were designed before this mass explosion of data and devices—in fact, they often hark back to a pre-digital age.

As the growing volume of data has stretched traditional forensic tools to capacity, it becomes impossible to examine them all. Digital forensic investigators may take arbitrary decisions as to which evidence sources they analyze first—or at all.

Many investigative organizations do not follow standard processes for each stage of this investigative workflow. For example, digital forensic investigators may not handle all data sources consistently during processing and analysis. They might deliver the relevant information in different ways, such as printing it out, copying it to a CD or flash drive, or placing it on a computer where other stakeholders can search it.

A traditional investigation relies on two very different groups of people:

- **Case investigators**—such as police detectives and corporate fraud analysts—understand the wider context of the investigation and examine crimes or investigative matters from all angles. If an investigation has a digital component, such as a computer recovered at a crime scene, an investigator would call on the support of one or more digital forensic investigators who would examine the computer and they report their findings back to them
- **Digital forensic investigators** collect, image, process and examine the collected data. They typically identify and evidence digital material that they consider potentially relevant to the case.

This division of roles (see Figure 1) is a major source of inefficiency. More worryingly, it highlights a disconnection in the investigation process. Very often case investigators view digital forensic investigators as providing a supporting service, rather than working with them collaboratively. For example, it is common for a single digital forensic investigator to handle all devices involved in an investigation. This solo digital forensic investigator will typically examine each storage device in turn, extracting the information he or she thinks is relevant then preparing a report for the less technical investigative team.

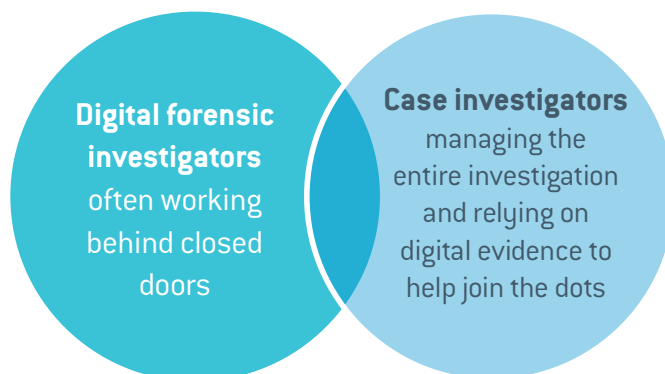


Figure 1: Typical division of roles in investigations with very little overlap between those performing technical and non-technical tasks.

This piecemeal approach is very inefficient, especially in larger cases. Digital forensic investigators must make critical decisions about the relevance of particular documents, email messages or images—or even entire evidence sources—often without knowing the broader details of the case.

This is something like the ancient parable about six blindfolded men trying to describe an elephant. One is touching the elephant's trunk and thinks the elephant is like a snake. Another is holding the elephant's leg and thinks it is like a tree, and so on. While they are all correct about the individual parts, none of them can see the bigger picture (see Figure 2).

The inability to understand the evidence in its broader context can have knock-on effects when digital forensic investigators must provide their findings to larger teams or other stakeholders such as prosecutors or human resources, legal or internal audit departments. These stakeholders must often rely on the pieces of the puzzle that the forensic investigator decided were most relevant.



Figure 2: Focusing on one part at a time makes it hard to see the elephant in the room.

Focusing on one part at a time makes it hard to see the elephant in the room

## LESSONS FROM eDISCOVERY

Some investigative organizations have streamlined their processes for handling digital evidence (see for example, the United Kingdom Serious Fraud Office case study). In this area, investigators have a great deal to learn from the way legal teams handle electronic discovery, which typically involves even larger volumes of digital evidence than investigations.

In many jurisdictions, courts require litigants to collaborate and consider eDiscovery at the earliest opportunity. Doing so can help keep the costs of the case proportionate to the amount at dispute. To achieve this, discovery practitioners must quickly identify potentially relevant material among massive volumes of digital evidence. They must then place this potentially relevant material into the hands of the experts such as senior lawyers, financial investigators or subject matter experts.

Legal teams often use a tiered review system, assigning junior staff to perform a “first cut” review of the material to eliminate the material that is clearly not relevant. Rather than allowing these reviewers to make arbitrary decisions, someone who has in-depth knowledge of the case would create a pre-defined set of guidelines for them to follow. This person may also review, validate or amend these decisions.

In this way, smaller and smaller volumes of more and more relevant material are passed up the chain. The highly knowledgeable—and usually highly paid—experts need only see the “hot” documents, safe in the knowledge that someone has reviewed and classified all other material.

This process is a very efficient way of classifying huge volumes of material into relevant or not relevant bundles. For it to work, legal teams must be able to:

- Divide up the available evidence into parcels for multiple people to review
- Ensure each reviewer understands the ground rules for deciding what is relevant
- Make the most relevant documents available for experts to analyze and examine.

This approach is not new, even in investigative circles. For example, in many complex criminal matters, rank-and-file detectives do the ground work, such as identifying witnesses and evidence, before passing on their findings to senior officers and subject matter experts for review.

However, it is rare for digital forensic investigators to follow this process when dealing with electronic evidence, often because traditional tools make it difficult to combine information from multiple sources and make it available to non-technical investigators or subject matter experts for review.

### CASE STUDY: SERIOUS FRAUD OFFICE

The collaborative investigation model and lab workflow discussed in this paper draws on the experience of Nuix’s Director of Forensic Solutions Paul Slater at the United Kingdom Serious Fraud Office (SFO). By adopting this model, the SFO increased the volume of data it could process each year 20-fold and made it possible to deliver timely responses to information requests from courts.

As interim head of the Digital Forensics Unit from 2009 to the end of 2010, Slater helped to standardize and streamline the SFO’s digital investigative processes. The SFO reduced its focus on in-depth forensics; created and automated investigative workflows; and developed a more collaborative approach to investigations. This change in approach helped to transform the SFO’s capabilities.

“While traditional computer forensics techniques dig deep into a handful of computers, [the SFO can now] quickly distil the huge volumes of data captured in our search operations and to focus on material likely to have greatest evidential yield,” wrote the SFO’s Chief Executive in its 2010-11 Annual Report and Accounts.<sup>1</sup> “We can now handle up to 100GB of new information each day—a 2,000% increase year on year.

“It is also allowing us to respond quickly to court requirements—in one case we were able to identify and produce over 47,000 emails overnight to satisfy a judge’s order. Such speed of response would have been impossible before.”

Investigators have a great deal to learn from the way legal teams handle electronic discovery, which typically involves even larger volumes of digital evidence than investigations

## THE INVESTIGATIVE LAB WORKFLOW MODEL FOR DIGITAL INVESTIGATIONS

The investigative lab model is a way for investigators to combine the efficiencies of the eDiscovery process with the forensic rigor and provenance of traditional digital investigation methodologies. It offers a more efficient way of utilizing available resources by making it possible to spread work between digital and non-digital investigators and subject matter experts, rather than being the sole responsibility of often a single person. It also ensures that digital forensic investigators handle each piece of evidence using an agreed set of repeatable processes.

Although this workflow model is technology agnostic, it requires a number of capabilities that are not available in traditional forensic tools, such as:

- Conducting a light metadata scan of data sources to rapidly assess their content and value to the investigation
- Combining multiple evidence sources into a single data set for analysis
- Processing multiple terabytes of evidence within a reasonable time
- Supporting a wide range of file formats including forensic image formats from multiple competing technologies
- Dividing a large data set into sub-cases based on criteria such as date, custodian, file format, location or language, then recombining the results into a single case.

### Share the workload—a production-line methodology

A key shortcoming of the traditional approach is that digital forensic investigators examine each evidence source individually. However, cases often hinge on a particular type of evidence—such as documents, emails or text messages—and the connections between them across multiple sources. In addition, certain evidence types must be reviewed by people who have particular expertise.

As a result, a necessary first step of this methodology is to assemble all available evidence—forensic images, email and mobile phone communications, loose files, documents and the rest—into a single location.

Conducting a light metadata scan on all these evidence sources can quickly help investigators choose the items they want to process in greater depth. This forms part of the content-based forensic triage process Nuix has discussed in a previous white paper, *Content-Based Forensic Triage: Managing Digital Investigations in the Age of Big Data*.<sup>ii</sup>

Once the investigative team has chosen the evidence sources most likely to be relevant, they can then process these devices following a set of previously agreed standards and settings. Investigative organizations can build a series of best practices or case-specific workflows which automate many of the time-consuming and error-prone tasks that are performed on each case. These include date range filtering, keyword searching and tagging, identification and optical character recognition (OCR) for non-text documents. The workflow can also include activities to filter out irrelevant information such as duplicate items or system files.

This approach significantly reduces operator-level decisions and inconsistencies around which files are processed and how, leading to a consistent and repeatable outcome. By using this methodology, investigative teams can rapidly trim down large evidence sets into small numbers of highly relevant items for expert review (see Figure 3).



Figure 3: A tiered approach to reviewing evidence.

Using the collaborative approach and tiered review methodology, investigators can quickly distil huge volumes of data into smaller, more manageable chunks

## Ensure the right data gets reviewed by the right people

The next phase of this investigative workflow involves dividing the processed evidence into review sets. At its most basic, it can be a way of sharing the work between multiple investigators to complete the task faster. They may choose to divide up the evidence by date ranges, custodians, location, language or content. However, there are many other options.

In a fraud case, for example, investigators could pass on financial records to forensic accountants and internet activity to technical specialists. In an inappropriate images investigation, detectives could package potentially relevant pictures and videos for specialist child protection teams, while leaving other file types for their digital forensic investigators. In multi-jurisdictional investigations, investigative teams can produce evidence or intelligence packages for third parties to review, comment on and return.

## Use advanced investigative techniques

As we have discussed, this workflow borrows a number of ideas from eDiscovery. Investigators can also use a number of technology-assisted analysis techniques from the legal world to look at the evidence from different angles.

For example, some investigative tools offer ways to visualize the data and metadata. Common visualizations include timelines of document or user histories; network maps of communications between people; and mapping locations of photos or phone calls. These allow investigators to quickly understand the who, what, where, when and why of the evidence.

Another emerging technique is the ability to identify similar or near-duplicate documents within a data set. By extracting and comparing lists of overlapping phrases called “shingles,” investigative tools can identify documents with similar content, and gauge how similar they are. This can help investigators identify who created, received or sent key emails, documents or attachments; analyzes how documents have changed over time; or find related documents that use similar language.

Near-duplicate technology can also speed up the process of identifying relevant evidence by connecting file fragments and recovered deleted data to similar content found within live files to build up the picture of events over time.

The list of shingles used to make near-duplicate comparisons has another powerful use: increasing the relevance of keyword searches. For example, fraud analysts search for words and phrases that indicate an employee may have motivation, opportunity and rationalization to commit fraud. However, these keyword searches may also bring up many results that are not relevant. By instead searching the list of shingles that contain words such as “cover up,” “write off,” “grey area,” “not ethical” or “off the books,” investigators can quickly locate longer phrases that will deliver highly relevant search results.

## Perform deep forensics only when necessary

Many digital forensic investigators have been reluctant to change their processes even as they struggle with masses of digital material. One reason is they believe the old-fashioned technique is the only way to achieve the forensic rigor and deep technical analysis that courts and other authorities require.

In our experience, nowadays the key evidence is more often found “hidden in plain sight”—typically in email, documents, spreadsheets or images. The major impediment for digital forensic investigators is that they can’t get across the volume of evidence to find the facts that will prove or disprove the case, especially if they conducting time-consuming deep forensic analysis on every data source. As a result, in-depth forensic analysis must become the exception rather than the rule.

Using the collaborative approach and tiered review methodology in the paper, investigators can quickly distil huge volumes of data into smaller, more manageable chunks. Once this technique identifies the smoking gun or the elephant in the room, investigators can pass this piece of data back to specialist digital forensic investigators for them to examine and provide provenance, validate authenticity and produce to the court or authorities. Alternatively, the review workflow may not locate the crucial evidence but may provide strong clues as to where digital forensic investigators should deep-dive to try and find it.

i <http://www.sfo.gov.uk/media/175084/resource-accounts-2010-11.pdf>

ii [http://nuix.com/PDFDownload.aspx?f=images/resources/White\\_Paper\\_Nuix\\_Content-Based\\_Triage\\_US\\_WEB.pdf](http://nuix.com/PDFDownload.aspx?f=images/resources/White_Paper_Nuix_Content-Based_Triage_US_WEB.pdf)

Key evidence is more often found “hidden in plain sight”.  
As a result, in-depth forensic analysis must become the exception rather than the rule.

TO FIND OUT MORE ABOUT NUIX'S INVESTIGATIVE LAB VISIT  
[nuix.com/investigator-lab](http://nuix.com/investigator-lab)

#### ABOUT NUIX

Nuix enables people to make fact-based decisions from unstructured data. The patented Nuix Engine makes small work of large and complex human-generated data sets. Organizations around the world turn to Nuix software when they need fast, accurate answers for digital investigation, cybersecurity, eDiscovery, information governance, email migration, privacy and more.

#### APAC

Australia: +61 2 9280 0699

» Email: [sales@nuix.com](mailto:sales@nuix.com)

#### North America

USA: +1 877 470 6849

» Web: [nuix.com](http://nuix.com)

#### EMEA

UK: +44 207 877 0300

» Twitter: [@nuix](https://twitter.com/nuix)

