INTELLIGENCE, COLLABORATION, AND ANALYTICS FOR DIGITAL INVESTIGATIONS

Evolving beyond traditional forensics practices to efficiently examine and understand the contents of multiple evidence sources and data type

Nuix has advocated for many years that investigators need to evolve beyond traditional forensic tools and workflows, so they can efficiently examine the contents of multiple evidence sources at once

EXECUTIVE SUMMARY

Investigators must deal with large and growing volumes of digital evidence across an increasing number and variety of sources.

Criminals and wrongdoers have grown skilled at using technology to conceal their activities. We would argue that some are more effective at covering their tracks than investigators are at applying technology to uncover them.

Nuix has advocated for many years that investigators need to evolve beyond traditional forensic tools and workflows, so theyd can efficiently examine the contents of multiple evidence sources at once. But just as the key facts may not be located within a single evidence source or connected to just one person, they may not even be in the same investigation, or the same agency, or the same country. As a result, efficient investigation must enable people to share intelligence, to collaborate across geographic and jurisdictional boundaries, and to find seemingly hidden connections across very large numbers of evidence sources.

Technology has stood in the way of these vital abilities. Digital forensic tools have burrowed further and further down the rabbit hole of deeply examining single evidence sources. They can tell you everything you need to know about the binary structure of data on a hard drive, but nothing about how the instant message history stored in that data connects with a mobile phone seized in another investigation on the other side of the country.

This paper will examine technology-enabled processes for making those connections. It will discuss:

- Automatically extracting intelligence items such as email addresses and credit card numbers, correlating them across all available evidence sources and sharing this information efficiently with other investigators
- Providing a way for multiple investigators, subject matter experts, and external agencies to review and collaborate on the evidence you have collected
- Applying data analytics to progress rapidly from a bewildering array of information to highly relevant details.
- In this way, investigators can apply technology where it is most suited, free themselves from tiresome menial work and make best use of their brainpower and intuition.

DIGITAL EVIDENCE DOESN'T SHARE ITS SECRETS EASILY

Investigators face many challenges when dealing with digital evidence. Modern telecommunications technologies make it easier for criminals to operate across jurisdictional and national borders, hide their activities, and evade detection and prosecution. To combat them, law enforcement agencies need an efficient legal and technological framework to exchange intelligence.

Law enforcement agencies must also deal with large and growing volumes of data in an expanding number of devices. In addition to computers, potential evidence sources may include cloud email, social media, digital cameras, media players, smart household appliances, GPS devices, portable storage devices, and smart wearable devices such as fitness bands, eyeglasses, and watches. The prevalence of smartphones and tablet devices on one hand, and low-cost anonymous burner mobile phones on the other, means a single suspect may be connected to 10 or more mobile devices. Large-scale investigations in areas such as counterterrorism and organized crime may involve data from multiple suspects, each with a dozen potential evidence sources. As we have discussed, the traditional linear methodology of forensically examining each data source individually can never hope to keep up. The combination of slow forensic tools and case backlogs means that by the time forensic technicians examine an evidence source, it may be months old. By this stage much of its intelligence value may be lost.

A LACK OF SHARED INTELLIGENCE

For many years, Nuix has advocated an investigative methodology that makes it possible to examine and cross-reference multiple evidence sources at the same time. However, it is not uncommon for crucial information to reside outside the evidence gathered for a specific investigation. It may be in a previous or concurrent investigation conducted by the same personnel or someone else. It may be from a different agency, office, location, or country. According to the UK National Policing Improvement Agency, "Intelligence management involves linking information from a wide range of sources in order to build up a composite picture. It aims to highlight links between people, objects, locations, and events that are essential in supporting [policing purposes]. Identifying links enables decisions to be made about priorities and resources needed to manage risk."[1]

Many law enforcement agencies are acutely aware of how important it is to share intelligence internally and externally. But these efforts often fail at a practical level. For example, more than a decade after the US Government's 9/11 Commission Report made a series of recommendations on sharing intelligence information between agencies,[2] there are still many technical and procedural barriers to effective counterterrorism intelligence sharing in the United States.[3] Using traditional methods, investigators struggle to compare information between individual evidence sources in a single investigation. Sharing intelligence across multiple investigations and between agencies is an even more manual and labour-intensive process.

TECHNOLOGY TOOLS AN IMPEDIMENT TO COLLABORATION

Another major impediment to efficient investigations is the difficulty investigators face making digital evidence available for review to internal or external personnel. Investigative technology vendors have tried to solve this problem by bolting legal review platforms onto forensic investigation tools, or adding forensic processing and analysis capabilities to an existing review platform.

INABILITY TO FIND HIDDEN CONNECTIONS

In investigative organizations, forensic technicians often work in isolation from case investigators and other stakeholders.

The forensic specialists must make critical decisions about which data sources to examine, how to process them, and what information to extract—often without knowing the broader details of the case. This approach can only be helpful if the case investigators and forensic technicians have a clear understanding of what to look for. The connections between people, objects, locations, and events can be critical in providing intent or collusion, but often they are not immediately obvious.

It would take superhuman skill to mentally correlate connections from a single suspect's hard drives, mobile devices, instant messages, cloud email, cloud storage, and social media interactions. Multiply this by the number of suspects in an investigation and technology becomes the only answer.

Under such circumstances, the ability to visually represent and analyze data can be a rapid shortcut to locating the key facts and connections of the case. But most investigative tools have limited abilities to visualize data, especially across multiple evidence sources.

Law enforcement agencies must also deal with large and growing volumes of data in an expanding number of devices

USING TECHNOLOGY TO AUGMENT HUMAN BRAINPOWER

A common theme behind the problems we have discussed is that they require considerable human effort to conduct tasks that computers are much better at.

This situation has evolved because the developers of investigative tools have focused on increasing the forensic depth with which they can analyze individual evidence sources. However, solving crimes very often requires finding the connections across multiple individuals, places, events, and evidence sources. Human intuition has its place in the process, no doubt. But much of the laborious work involves picking out and matching specific pieces of information from massive volumes of data. Computers, when applied judiciously, have a natural advantage in intelligence sharing, collaboration, and data visualization and correlation. Here's how you can apply technology in the right places to assist human investigators.

INTELLIGENCE

Using the traditional digital investigation model, investigators must manually compare intelligence items across each evidence source. Even something as simple as proving person A and person B both used the same stolen credit card number is a complex matter of identifying credit card numbers in their various evidence sources, printing out lists of those numbers, and then poring over those lists to find the matches. Even semi-automating this process, by exporting the data sets into a database and running comparisons, is time consuming and error prone.Nuix's advanced investigative tools automatically extract a range of standard intelligence items using a "named entities" model to identify particular patterns of letters and numbers such as:

Names; Email addresses; IP addresses; Company names; Credit card numbers; Bank account numbers; Social security or identity numbers; Amounts of money.

This list is not comprehensive. Investigators can easily define their own named entities, using regular expressions, to extract locally relevant or investigation-specific types of information such as passport, phone, vehicle identification, or contract numbers.

Having identified relevant intelligence items, investigators can instantly see which suspects within a current investigation have those items in common, across all the evidence sources in the entire case. Typically, they can also identify who shared what, with whom, and when, using techniques such as timelines and network diagrams. However, as we have discussed, the connections between suspects and intelligence items may not sit neatly within the current investigation. It may be in a past or concurrently running investigation, or in one another organization has conducted. Using Nuix's investigative technologies with the highly scalable distributed database Elasticsearch makes it possible to create a centralized library of intelligence. Once each investigation is complete, investigators or forensic analysts can package up the intelligence items they have extracted and add them to the library. Next time, when they identify a potentially relevant name or phone number, for example, they can use it to query the intelligence library and see if it has appeared in previous investigations. If they had previously translated a document or video and added the translation to the intelligence database, and the same document appears again, they can immediately access the translation. Extracting intelligence and compiling a database of known items is an extremely rapid and automated way to uncover hidden connections between multiple evidence sources, people, locations, and investigations.

Intelligence does not only come from seized devices or suspect interviews; it also stems from a variety of real-time and real-world sources including:

- Open source and cybersecurity intelligence feeds
- Communications data and phone records from mobile carriers
- Machine data generated by human activity, such as building access logs
- Real-time data from computers that indicates what is going on at the device level
- Flight manifests
- Shipping port logs.

Nuix software makes it possible to combine historical digital evidence with real-time data sources, and draw correlations between the two. For example, digital forensics may show that someone accessed a computer using Paul Slater's credentials at a particular time but building access logs and mobile phone records reveal that Paul was not in the building at the time.

Computers, when applied judiciously, have a natural advantage in intelligence sharing, collaboration, and data visualization and correlation

Using these techniques also makes it easier for agencies to share intelligence with each other. For example:

Agencies can share lists of names, email addresses, or other intelligence items. For example, police investigating a suspected terror cell in Birmingham can provide a list of suspect phone numbers, email addresses, online aliases, and bank account numbers to their colleagues in Manchester.

- The Manchester investigators can run this list through their existing case files to see if any connections emerge
- Investigators can build lists of relevant words and phrases within case material such as documents or illegal or inflammatory propaganda material, and share these word lists with other agencies. Using near-duplicate analysis, they can identify identical and similar items within different evidence collections held in completely separate investigations.

These techniques also alleviate the issue of the delays in extracting intelligence using traditional methods. As soon as investigators uncover relevant and timely intelligence, they can run this across other potentially relevant evidence sources and very rapidly find and delve into any matches.

COLLABORATION

In the past, I've discussed an operating model that enabled investigative teams to divide up digital evidence and spread the review workload between many people. At a basic level, this is a way to share work between multiple investigators to complete the task faster. They may choose to divide the evidence by date ranges, custodians, location, language, or content.

It can also be a way to distribute different types of evidence to the people most qualified to understand it and its context. 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 multijurisdictional investigations, investigative teams can produce evidence or intelligence packages for third parties to review, comment on, and return.

A growing number of Nuix customers are using Nuix Investigate® to extend this collaborative model to tens or hundreds of investigators across multiple locations. Its simple interface enables people with minimal training or technology expertise to search, review, tag, and analyze data from any web browser. Role-based security controls ensure you can make evidence easier to access while protecting confidential and sensitive information.

Larger law enforcement agencies, advisory firms, and enterprises are using this model to set up centralized evidence processing facilities that can provide access to the results to any desktop across the organization.

This model has considerable advantages for sharing intelligence. A centralized lab which stores relevant case data related to all current investigations makes it easy to cross-reference intelligence items. It is also easy for one location or agency to provide in-depth access to their case data for colleagues in other locations.

Another useful approach for setting up a lab is defining templates and workflows for processing data consistently and repeatably, which Nuix lets you do. Adopting a templated workflow can help forensic laboratories show compliance with the quality assurance and testing accreditation schemes under which many operate.

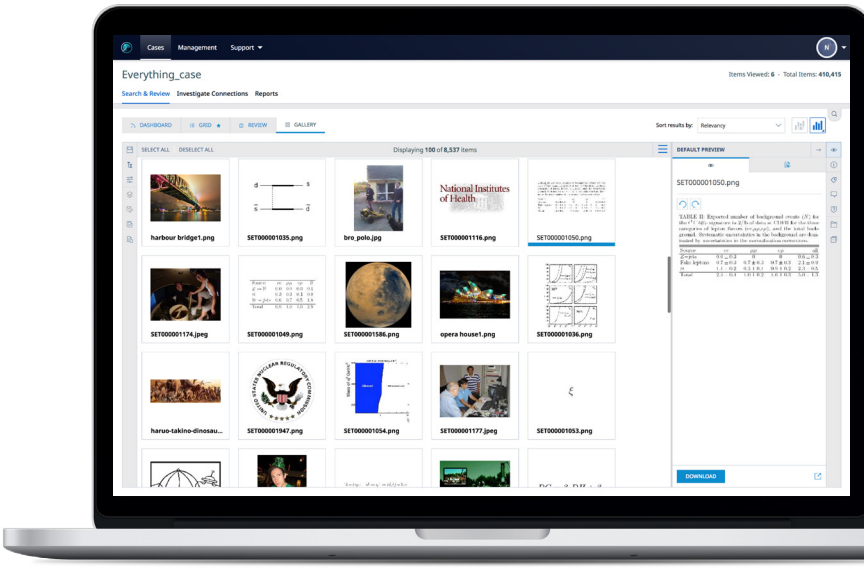

Figure 1: An image gallery in Nuix Investigate

APPLYING RELATIONSHIP-DRIVEN ANALYTICS & INTELLIGENCE

Nuix Investigate takes link analysis further by applying the well-known investigative framework of people, objects, locations, and events (POLE). Our software automatically correlates data to show connections between these four dimensions and displays them in a graph-like visualization—augmenting digital evidence with human context to simplify investigations and data analysis. This allows professionals of all levels of expertise to quickly see and drill down into the hidden relationships in their data.

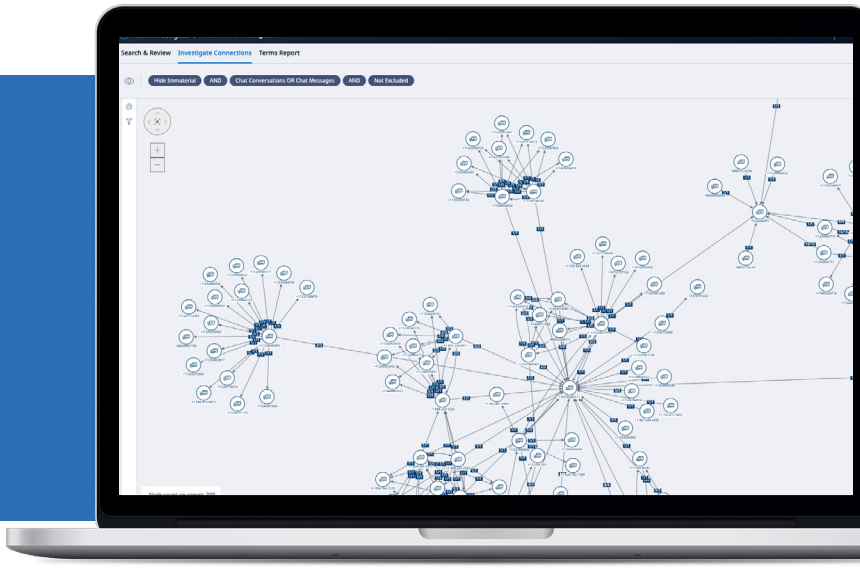

Figure 5: Examining the links between people, objects, locations, & events.

BUT WHAT ABOUT FORENSICS?

Investigators and forensic technicians reading this paper may be asking themselves, “But what about forensics? Will any of this stand up in court?”

The techniques in this paper offer a way of working that augments forensic analysis but does not eliminate the need for it, particularly in the areas of provenance and authenticity. However, as we have previously argued, the volume of evidence in most cases makes it too time-consuming to conduct deep forensic analysis on every data source. Instead, we offer a much faster and more efficient way of identifying the evidence sources that contain the data required to prove or disprove the case—and putting that information in front of the people who need to see it. The investigative team can then pass a small number of evidence sources back to digital forensics specialists so they can conduct in-depth analysis that will satisfy courts and authorities. This enables technical specialists to do the “clever stuff” they are best suited to rather than getting bogged down in repetitive grunt work such as running keyword searches across every exhibit—even those that contain no relevant data.

It is also worth noting that Nuix offers a full range of in-depth forensic capabilities, the equivalent of traditional specialist forensic software. Keeping the data within a Nuix case file removes the need to convert and move data between formats and tools during the investigation process. It is therefore much easier to maintain provenance and trace critical evidence identified during the investigation back to its original source.

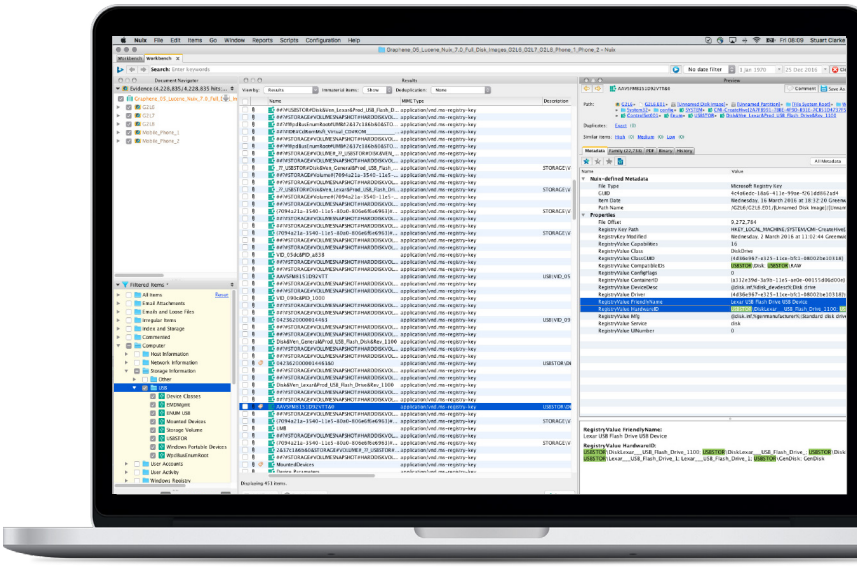

Figure 6: Examining USB device forensic artifacts.

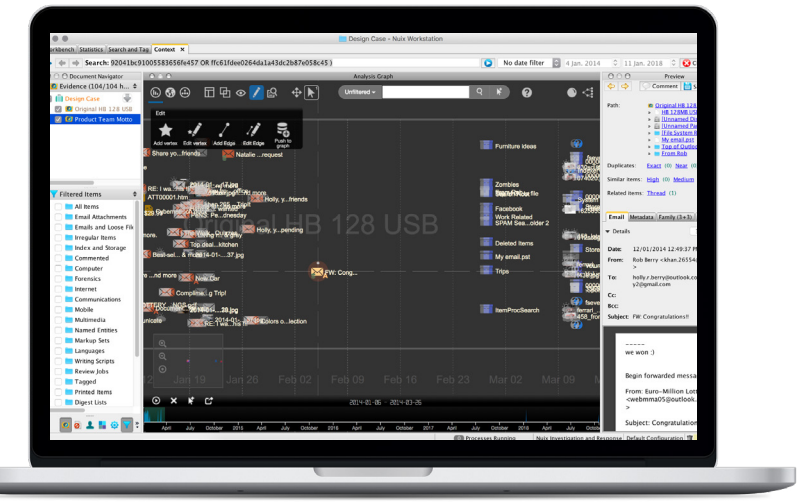

Figure 2: A timeline of events showing linked items.

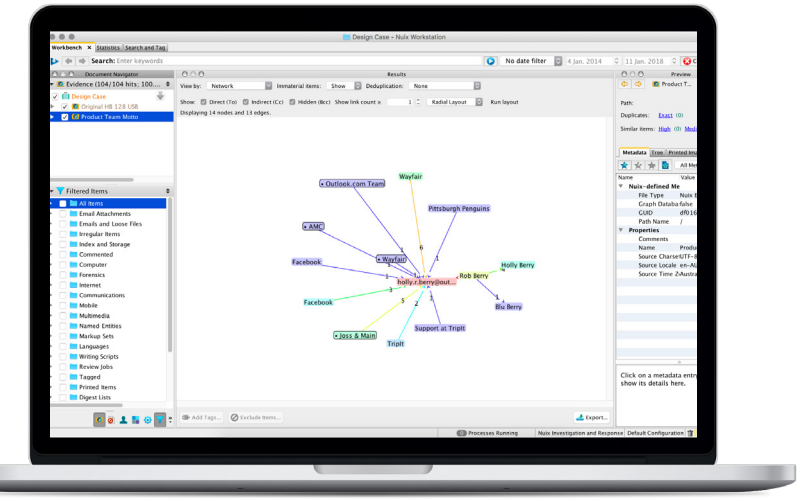

Figure 3: Communication analysis of mobile device, email, and Skype communications.

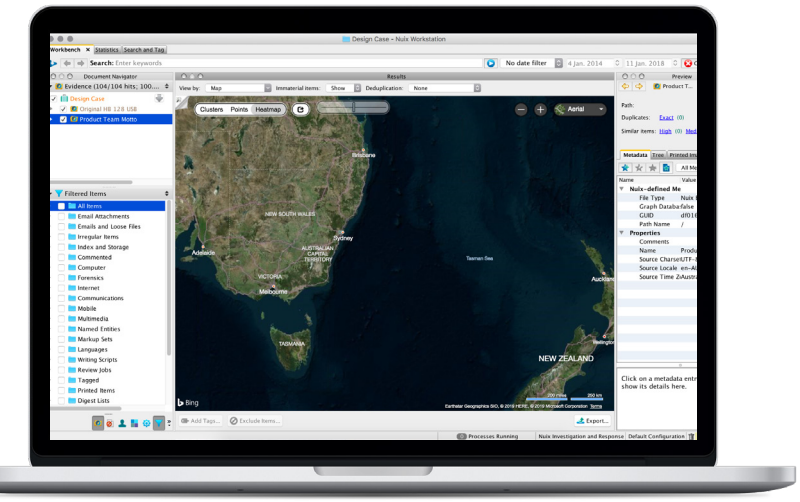

Figure 4: A heat map view showing the location and frequency of items with geographical data.

ANALYTICS

As digital evidence becomes larger and more complex, investigators' greatest struggle is not a lack of information, but having too much to make sense of. Visually representing large volumes of data can be a fast way to locate the key facts and connections within the case. It enables people, even with limited technical knowledge, to follow a hunch or idea down to very specific details in a matter of seconds. Common analytical techniques include:

- **Timeline.** Reviewing the content of emails, documents, phone calls, or other communications from multiple sources or custodians in the order they happened.
- **Link analysis.** Understanding the connections between people and intelligence items such as credit card numbers, IP addresses, organizations, and sums of money. As I've already discussed, finding links between seemingly unrelated items is made more powerful when applied across multiple investigations.
- **Map.** Plotting items with geographical information—including digital photos, mobile device logs, and IP addresses—to show their location and frequency.
- **Date trending.** Visualising the frequency of data over the entire case or any filtered subset, then drilling down to year, month, or day views.
- **Communication network.** Showing the interactions between persons of interest with an interactive network diagram that shows the number of connections for each link.

Combining analytical techniques can help investigators progress from a bewildering array of information to highly relevant details very quickly. For example, you could filter an entire evidence set to just email messages within a relevant date range that contains credit card numbers. If that still returns too many results, you could use other techniques such as suspect names or keyword searches to further filter the evidence. Now you can use a communication network diagram to see who is emailing credit card numbers to whom. A timeline view, traditionally used for email messages, is also useful for SMS messages, mobile device call logs, instant messages, Skype chats, and social media messages. In my experience, many people say things in instant messages that they would avoid in email.

Individual investigators can apply timelines, date trending, communication network, link analysis, and mapping analytics in Nuix. You can also use this application for textual analytics such as shingles and word lists. Investigators and other stakeholders can also apply a full range of analytics to digital evidence using Nuix Investigate.

ABOUT THE AUTHOR—PAUL SLATER

Paul Slater is a subject matter expert with over 20 years of experience in investigations, digital forensics, and eDiscovery. Paul has held senior roles within law enforcement, corporate, and Big Four advisory organizations. He was a member of the review board for the Association of Chief Police Officers Good Practice Guide for Digital Evidence.

Paul served for two years as interim head of the digital forensics unit in the primary UK agency for investigating and prosecuting serious and complex fraud. He designed workflows and implemented technologies that enabled the agency to process 20 times more electronic evidence each year. Paul now uses his expertise to help Nuix customers master their data through designing, building, and implementing digital forensic and eDiscovery solutions.

REFERENCES

1 National Policing Improvement Agency Guidance on the Management of Police Information, 2nd edition, 2010

2 National Commission on Terrorist Attacks Upon the United States, The 9/11 Commission Report, 26 July 2004

3 See for example Richard A. Best Jr., Intelligence Information: Need-to-Know vs. Need-to-Share, Congressional Research Service, June 2011

To find out more about Nuix digital investigations
visit nuix.com/solutions/government/forensic-investigation

Nuix (www.nuix.com, ASX:NXL) creates innovative software that empowers organizations to simply and quickly find the truth from any data in a digital world. We are a passionate and talented team, delighting our customers with software that transforms data into actionable intelligence and helps them overcome the challenges of litigation, investigation, governance, risk and compliance.

APAC
Australia: +61 2 8320 9444

EMEA
UK: +44 203 934 1600

NORTH AMERICA
USA: +1 877 470 6849