Event summary of Session 5

Overview of forensic technology

Presented by Baron Zhao and Michael Poon, both of EY

A. Overview: Key Areas of Technology in Forensic Projects

<table>
<tr><th>IT Forensics</th><th>Electronic Discovery/ Disclosure</th><th>Forensic Data Analytics</th><th>Cybercrime</th></tr>
<tr><td>Preservation of Electronically Stored Information</td><td>Investigation document review</td><td rowspan="3">Acquiring, merging, and transforming complex data from multiple sources to create critical business intelligence</td><td>Pre-incident preparation with cyber diagnostics</td></tr>
<tr><td rowspan="2">Digital investigations</td><td>Disclosure review and response</td><td rowspan="2">Investigative response to cybercrime incident</td></tr>
<tr><td>Discovery response planning and records management</td></tr>
</table>

B. IT Forensics

- It is the art and science of:

  - ✓ Identifying sources of electronic evidence (e.g. WhatsApp messages) potentially relevant to a matter under investigation
  - ✓ Preserving electronic evidence with an appropriate tool or methodology
  - ✓ Analysing electronic evidence to uncover information relevant to the facts and circumstances of the investigation
  - ✓ Reporting on the evidence in a manner admissible in a court of law or other legal proceeding

- Sources of IT Forensics Evidence

  - ✓ Incidental data: metadata, deleted/hidden files, system logs, mobile device logs, data in slack space
  - ✓ Interviews (if further investigations are required): Who did what? Who knew what? When?
  - ✓ Electronic documents: word documents, spreadsheets, presentations, email, intelligent, character recognition
  - ✓ Transactional data: database records, general ledger, payroll, AR, AP, inventory, production data

- IT Forensics process: scoping and identification. Work performed commonly includes:

  - ✓ Assessing the requirement (e.g. what is the underlying scope?)
  - ✓ Interviewing IT staff and key staff
  - ✓ Reviewing documentation policies
  - ✓ Requesting administration information from IT

- ✓ Interrogating IT systems
- ✓ Developing and testing extraction methodologies
- ✓ Mapping data sources, custodians and collection strategies
- ✓ Reporting on approach undertaken

- IT Forensics pointers

  - ✓ Unless explicitly instructed otherwise, assume any IT Forensics investigation will require evidence produced in legal proceedings
  - ✓ Forensic technology knowledge is required in collection of electronic evidence — both technical and procedural
  - ✓ No action should be taken that may change original data
  - ✓ An audit trail should be maintained of all activity
  - ✓ Ensure all laws and regulations comply with international laws (e.g. China State Secret Law), ensuring that you have proper authority to collect data (for example, collecting a mobile device from an employee could be illegal in certain countries)

C. Electronic Discovery/Disclosure

- The process of identifying, preserving, collecting and producing Electronically Stored Information (ESI) that may be used as evidence in legal proceedings

- Similar process also applies to electronic investigations, audits and fact finding exercises

- Key stage of Electronic Discovery/Disclosure is electronic review (eReview)

- Electronic Discovery/Disclosure: Where the services are needed -

  - ✓ Civil litigations
  - ✓ Regulator inquiries and investigations
  - ✓ Government investigations
  - ✓ Cyber incident investigations
  - ✓ Corporate internal investigations

- Electronic Discovery/Disclosure: The parties involved -

  - ✓ Corporate internal IT (Preservation and collection of the company/organization data from employees involved in the subject matter)
  - ✓ Corporate compliance team (High level management and project oversight for compliance and investigation matters, liaison with external legal counsel or regulators)
  - ✓ Corporate in-house legal counsel (Provide legal advice and directions on legal matters, liaison with external legal counsel)
  - ✓ External legal counsel (Under the guidance of internal compliance and legal teams, review data and disclose relevant information to involved parties, i.e., regulator, court or opposing party in legal matters.)
  - ✓ Corporate in-house or external counsel legal technology expert (Liaison with eDiscovery technology services provider to manage and organize the data review
  - ✓ eDiscovery technology services provider (Provide eDiscovery technology services including forensic data collection tools, processing and hosting platforms, searching, analytics and production support.)

- ✓ Managed document review services provider (Under the guidance of internal compliance, internal and external legal counsel, conduct document review to find relevant information)

- Electronic Discovery/ Disclosure process: Forensic Expert's role

  Forensic Professional's involvement in each step of the Electronic Discovery/Disclosure process — steps to take and questions to ask

  1. Strategy and identification (Identify scope of collection)
  2. Data extraction (Determine if any custom filters need to be used)
  3. Data processing (Determine rules for processing)
  4. eReview (Determine review strategy)

- eReview: teaming with forensic technology

  - ✓ Cooperative approach required
  - ✓ Selection of keywords
    - ➢ Iterative process to get final listing
    - ➢ Make use of operators to reduce volumes
    - ➢ Consider noise words
    - ➢ Allow for language variances
    - ➢ Allow for informal words, abbreviations and acronyms
  - ✓ Refine keyword listing using an initial sample
  - ✓ Final listing can be amended after reviews start

D. Forensic Data Analytics

- It can be classified into proactive and reactive reviews.

- Quickly and cost-effectively identifying the most relevant information

- Leveraging the knowledge inside large-scale structured and unstructured data sets

- Involves elements of business modelling skills, fraud investigation knowledge and advanced data analytics.

- Forensic Data Analytics techniques:

Rules-based analytics	Testing data against known behaviors/ patterns through e.g. transaction monitoring system	• Suited to well-defined processes • Ideal for anomaly testing • Used for identifying data not meeting expected behavior
Risk scoring	Using scoring models to identify and determine areas of higher risk	• Using rules-based tests with weighted scores • Combining weighted scores from different datasets to determine high risk areas

Text analytics	Analysing natural language text: • Keyword searching • Concept analysis • Sentiment analysis	• Search against keywords, e.g., fraud, bribery and corruption, blacklists • Analyse text to determine concepts discussed • Analyse emails for emotional tone insight
Visual analytics	Using interactive dashboards for reporting and trend analysis	• Visualizing facilitates understanding • Makes comparative analysis easier • Aids in spotting unusual trends
Predictive analytics	Using existing data to understand trends and predict future behavior	• Developing models based on existing data • Machine learning

- Forensic Data Analytics process

  1. Assess risk: What should be analyzed?
  2. Acquire data: Gather and load the right data
  3. Analyse: Execute the right tests
  4. Review: Assess and validate exceptions
  5. Follow up: Determine response and next steps

- Data analysis pointers

  ✓ Have a concrete understanding of the purpose for undertaking analysis before requesting data
  ✓ Understand what data is available and obtain enough of it to support your work
  ✓ Check that the data is complete
  ✓ Recheck the analysis procedures if the results seem odd — always perform basic common sense checks

- Teaming with forensic technology services

  ✓ Most data analysis is highly customizable
    ➢ Forensic experts or the client can request potential tests or visualizations that could help better answer questions or understand data
    ➢ Analyses can be developed and adapted to specific geographic locations, specific industries, specific companies, etc.
  ✓ Discuss testing plan and output expectations
    ➢ Collaboratively determine what tests to run, what data is to be collected and the scope of work
  ✓ How will the test results be delivered?
    ➢ Interactive dashboard
    ➢ Traditional spreadsheets or data extracts
  ✓ Review analytics results with the forensic technology team
      ➢ Determine if there is any follow-up work or additional analytics to be run

E. Cybercrime

- Cybercrime overview

  - ✓ IT security has been identifying, reducing and managing risks
  - ✓ The usual approach has been:
    - ➢ Close the risk up to when it is acceptable/feasible/cost-effective (e.g. 90%) through, for example, investing in advanced technology.
    - ➢ Transfer the residual risk (e.g. 10%), for example, having an insurance coverage in place
  - ✓ Cybercrime is related to:
    - ➢ the "unknown unknowns" which bother organizations during the normal course of their existence
    - ➢ negative incidents which happen because of the exploitation of the 10% non-covered risk

- **Concerns: cyber threat intelligence**

  - ✓ What is a cyber threat?
    - ➢ Any threat which manifests itself via or because an organization is connected to a public network (i.e., the internet)
  - ✓ What do we do about it?
    - ➢ Establish the real possibility that such threats will cause a negative impact
  - ✓ Identify the actor(s) that bring concerns to the organization

- **Events: cybercrime investigations**

  - ✓ What is a cybercrime?
    - ➢ Any crime which requires the active role of a computer or a network
  - ✓ What do we do about it?
    - ➢ We investigate using a variety of data types
    - ➢ We do this in order to find evidence of a compromise or a crime
    - ➢ We are usually requested to establish the timeline of events in order to identify the actor(s) of the negative events under investigation

- **Cybercrime: Forensic Expert's role**

  - ✓ Forensic Expert's involvement in each step of the Cybercrime process — steps to take and questions to ask

    - ➢ Cyber threat intelligence
      - ▪ Determine sources of potential threat
      - ▪ Provide specific knowledge to enhance threat assessments

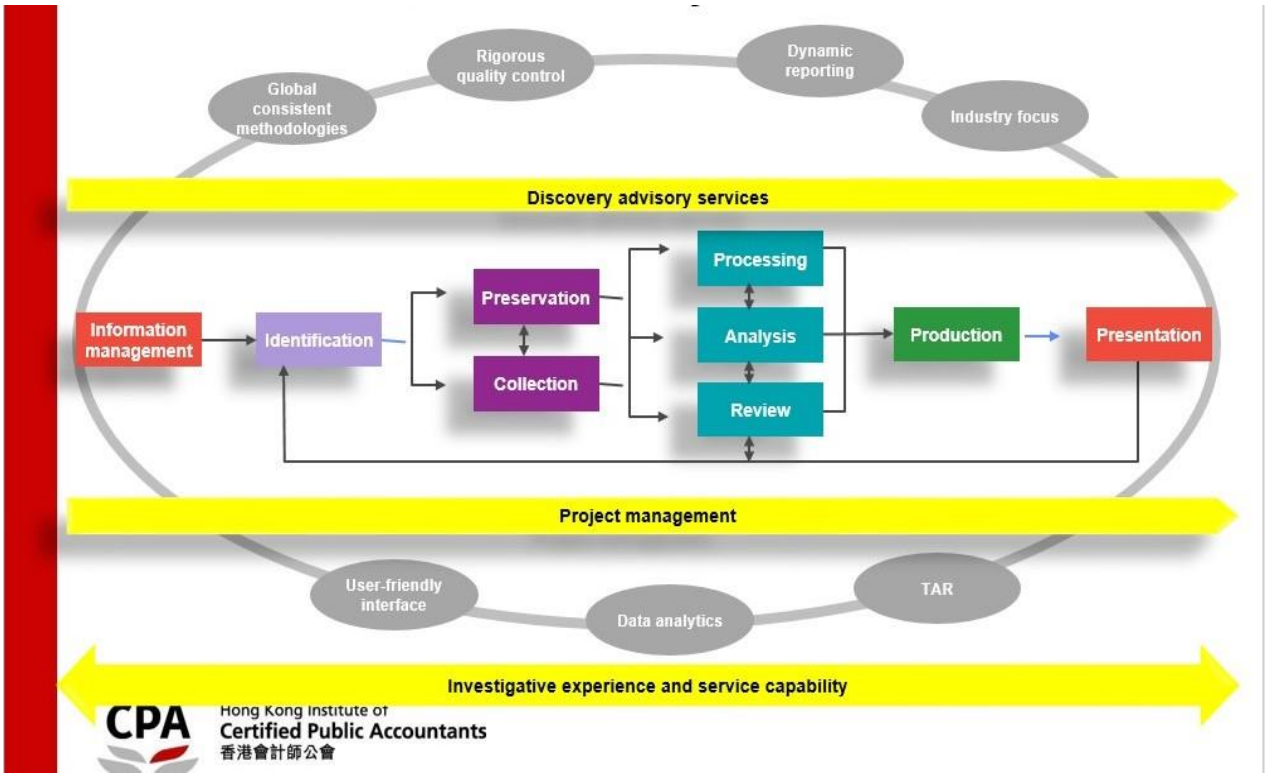
    - ➢ Cybercrime incident
      - ▪ Perform any follow-up investigations that may arise after response to the incident to determine timelines, negative events, etc.

F. Emerging Technological Tools

- **Evolving technologies play an increasingly key role**

	Robotic Process Automation (RPA)	Business Process Management (BPM)	Machine Learning / Artificial Intelligence	Advanced Analytics
Description	Software that mimics a human's interaction with systems	Software to model and manage end-to-end business process through orchestration and controls	Uses algorithms and large amounts of data to learn how to perform a task. Models designed to emulate human thought processes	Software used increasingly to visualize data trends and outliers and derive insights
Benefits	▸ Leverages existing system interfaces ▸ Enterprise grade software that is highly scalable with strong logging capabilities ▸ Supports tasks that involve multiple platforms	▸ Strong workflow and process monitoring ▸ Supports modelling complex processes with multi-step approvals ▸ Improves task coordination and audit trail	▸ Iteratively learns from data to find hidden insights without being programmed where to look ▸ Streamlines processes that are not deterministic ▸ Interprets contextual information and provides adaptive reasoning ▸ Applies human-like reasoning at scale	▸ Integrates data from structured or unstructured sources ▸ Improved decision making through predictive and prescriptive analytics ▸ Promotes knowledge sharing by serving data rapidly ▸ Near real-time interactive visualizations for quicker insights
Challenge	▸ A new skill that is emerging in the market ▸ Only suitable for highly deterministic and well-defined tasks	▸ Requires highly skilled resources ▸ Long implementation cycles ▸ Limited integration with business applications	▸ Limited by what it has been taught ▸ High complexity ▸ Selective use cases	▸ High complexity requiring very skilled resources ▸ Assumes ready infrastructure setup such as big data with access to large amounts of data ▸ Selective use cases
Suitability / Criteria	▸ Process repeats in a distinct manner ▸ Higher volume/frequency process ▸ Consistent outcomes of process	▸ Process requires sequencing and/or queuing of tasks ▸ Complex routing and distribution of process ▸ Higher number of areas/persons involved in process	▸ Process requires decisions to be made with varying outcomes ▸ The process requires validating the accuracy or outcomes of the automated task (apply human-like logic)	

- **Artificial Intelligence (AI) can assist across the Electronic Discovery Reference Model**



- **Electronic discovery tools that utilize AI, Machine Learning, Natural Language Processing, etc.**

  - Concept Clustering
    - ✓ Identifies concepts, topics and groups them together.

  - Predictive Coding
    - ✓ Also referred to as technology-assisted review (TAR) or computer-assisted review.
    - ✓ Used to find responsive electronically stored information (ESI) documents during a legal case's review phase.
    - ✓ Uses AI to develop software that continues to learn and make better decisions while significantly expediting the review process.
    - ✓ Starts by training software with a seed set of data. Reviewers then code each document as responsive or unresponsive, and input the information into the predictive coding software.
    - ✓ As the training continues, AI allows the software to learn and make better, faster decisions.

  - Continuous Active Learning (CAL)
    - ✓ CAL is a central part of a newer approach to predictive coding.
    - ✓ "Continuous", i.e., training doesn't end until the review finishes.
    - ✓ "Active", i.e., computer feeds documents to the review team with the goal of making the review as efficient as possible. As reviewers progress through documents, their judgments are fed back to the system to be used as seeds in the next ranking process.
    - ✓ When the reviewers request a new batch, the documents are presented based on the latest completed ranking.

- **Analytics**

  - Email threading
    - ✓ Reduces the complexity of reviewing emails by gathering and sorting the order of all forwards, replies and reply-all messages together allowing linear email thread review.
    - ✓ The technique can also be used to reduce the number of emails reviewed. Email threading identifies inclusive emails in the thread, meaning it identifies the email which contains unique content and should be reviewed.

  - Near de-duplication
    - ✓ Used to identify textually similar documents to assist and speed up the review process.
    - ✓ Grouping textually similar documents allows reviewers to see similar documents at the same time, this can have many use cases:
    - ✓ Bulk review of similar document versions – contracts, draft vs. final documents.
    - ✓ Used as a quality control mechanism - review teams can quickly see if review decisions and/or redactions are consistently applied across textually similar documents.

- **Electronic discovery tools that utilize AI, Machine Learning, Natural Language Processing, etc.**

  - Redaction tools
    - ✓ Simple redaction tools, which help replicate the original redactions on similar documents
    - ✓ More complex tools allow automation of redactions over files which are similarly formatted.
    - ✓ Tools to identify Personally identifiable information (PII) to help automate the process of redacting this. AI can be used to train the machine to pick up patterns e.g., phone number, driver's licence.

  - Audio Review tools
    - ✓ Conventional approach lexicon-based – limited applicability, with a high importance to develop and maintain the lexicon
    - ✓ More advanced cognitive tools able to identify sentiment, emotions and expressions, and provide contextual analysis for the conversation.

- **Advanced Analytics Services**

  - Named Entity Recognition (NER)
    - ✓ Capable of automatically identifying people, places, dates, times, credit card details, email addresses, and much more.

  - Topic Modelling
    - ✓ To identify the key underlying themes within a body of text for high-level classification – is this a conversation about manipulating the market? Transferring funds to an offshore account? A discussion that should be kept "under the radar"?

  - Sentiment Analysis
    - ✓ To identify the emotional tone within a body of text – is the author excited, angry, happy, or sad?