Digital forensics research is an activity within the incident response process and is geared towards the identification, investigation, containment and remediation of anomalies in IT-infrastructure, as well as developing indictments against cyberattacks for litigation and further investigations. The scope of DFIR (digital forensics and incident response) is to answer questions like; ‘what happened?’, ‘what damage has occurred?’ and ‘what are the next steps?’ both in relation to immediate incident response, but also to hold perpetrators accountable and make our systems more resilient against future breach attempts.

This preparatory note presents a framework to shape our discussions on 27 October, and come towards an agenda that shapes the direction of cyberforensic science in the Netherlands for the coming decade. We can then accelerate the adoption of emerging technologies from the Dutch fundamental research base by its end users, such as the NFI, CERTS and NCSC, strengthening our joint resilience and contributing to solving crime and supporting larger criminal investigations.

The first part of this note will detail the elements of a DFIR process, aimed towards identifying the capabilities needed to deliver a comprehensive solution for the end user. Potential gaps along this process chain are then opportunities for research areas that emerging technologies or novel socio-legal approaches could address, cascading into an outline for a research agenda. Stylized, the DFIR process can be roughly synthesized as follows:

1) **Information Governance**
Malicious actors have been known to target organizations specifically, but also sometimes simply spot an opportunity to exploit a vulnerability. A good example of that is ransomware. It is thus important for every organization to be prepared for a cyber-attack by adhering to the base measures of the NCSC and DTC, but also by implementing an information governance infrastructure that enables digital forensics research when a breach does occur. Without this precondition being in place, it is unlikely for a DFIR-process to be successful.

2) **Identification and Scenario’s**
When an anomaly has been detected by your SOC or SIEM, the first step is to identify potential scenario’s that explain the anomaly and verify if indeed a breach has taken place, as well as what information is relevant for subsequent
investigation. There is a strong ‘human factor’ in this step, as algorithms tend to be very good in noticing anomalies, but less so in scenario planning and reasoning (yet). This is therefore likely an iterative process, where more practice leads to quicker pattern recognition, further sharpened by the insights gained in subsequent DFIR steps.

3) Data Collection
The identification of relevant data under the previous step should deliver a target list of data to collect, this step pertains to actually retrieving that data. Modern IT-infrastructures are however increasingly complex, which can make collection difficult. It is also not necessarily the system itself that has been compromised, meaning sometimes we have to look at data in transit and in use as well. Possible sources thus include file server storage, firewall logs, random access memory storage, network packets, etc.

4) Data Extraction & Filtering
The required data then needs to be extracted from the forensic artefacts (at rest), raw network data (in transit) or executed prompts (in use) and processed. As is the case with the selection of data sources, filtering must often be applied to reduce the noise and make the dataset easier to process, particularly if this was not sufficiently possible under the previous step due to complex and intertwined infrastructures.

5) Data Analysis
The (still relatively large) dataset needs to be analyzed, both operationalizing and refining the scenario as outlined in the identification phase. To aid this process, tools used often include sand box simulation, visualization, big data analysis and machine learning. While all these techniques increase efficiency, human specialists are always needed to interpret the output and guide the algorithm. In more complex cases, where the intentions of the isolated code cannot be identified, reverse engineering might be necessary as well.

6) Evaluation & Evidence
When the data has been analyzed, the findings need to be evaluated and processed into admissible evidence, meaning that it logically supports the thesis of the scenario outlined in the beginning (or potentially a refined version). These conclusions then further guide the incident response and recovery process and the start of a potential criminal investigation, but can also be used as points of reference for automated security operations and CERTs to detect and deter subsequent similar attacks.

For our workshop, we would like to see where – from your personal and organizational perspective – the main barriers lie, and how to overcome them, towards an effective execution of these six phases. In the afternoon we are then looking to cross reference these barriers with some signals of change we identified, to also be augmented with your input. Our signals of change are;
- open source technology;
- strategic autonomy (ensuring that we do not have to fully rely on foreign suppliers);
- standardization (e.g., of logging or forensic readiness);
- artificial intelligence (e.g., big data analysis and machine learning);
- the availability of human capital to work in cyberforensics and
- public private partnerships to accelerate knowledge and expertise exchanges

To give an example, we could identify a lack of capability and knowledge with non-critical infrastructure companies to increase forensic readiness and cross reference that with open source technology towards a research project for developing a technology that can be easily deployed. The JCSU for instance has published a Windows Event Logging & Collection Guidance ([https://github.com/JSCU-NL/logging-essentials](https://github.com/JSCU-NL/logging-essentials)).

The immediate deliverable of the workshop will thus be a filled out version of the matrix in the appendix, that – together with our report – will serve as the basis for the cybercrime chapter of the NFI’s National Forensic Research Agenda 2032.

<table>
<thead>
<tr>
<th></th>
<th>Open Source</th>
<th>Strategic Autonomy</th>
<th>Standards</th>
<th>AI</th>
<th>Human Capital</th>
<th>PPS</th>
<th>...</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Governance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identification &amp; Scenario’s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Collection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Extraction &amp; Filtering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation &amp; Evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>