INNOCENCE PROJECT

Executive Director
Christina Swarns, Esq.

Co-Founders & Special Counsel
Barry C. Scheck, Esq.
Peter J. Neufeld, Esq.

INNOCENCE PROJECT PUBLIC COMMENT ON NISTIR 8354-DRAFT

Digital Investigation Techniques: A NIST Scientific Foundation Review
July 11, 2022

The Innocence Project is pleased to respond to the National Institute of Standards and Technology (NIST) call for public comments regarding the NISTIR 8354-DRAFT report, Digital Investigation Techniques: A NIST Scientific Foundation Review (the "report"). For nearly thirty years, the Innocence Project has worked to exonerate the innocent and prevent wrongful convictions through systemic reform. Nearly fifty-two percent of the individuals exonerated by post-conviction DNA testing were convicted based, at least in part, on expert forensic evidence later shown to be erroneous. To improve the integrity of convictions and reduce the risk of an innocent person being found guilty, the Innocence Project urges robust gatekeeping and works to ensure that forensic evidence is admitted at trial only when it has strong scientific support, particularly from well-designed empirical studies.

With respect to digital forensics, this report embodies an opportunity to ensure that these tools are applied with transparency and proper safeguards. We commend the authors for actively disseminating information regarding their process at conferences across the country and now holding a public comment period to receive feedback.

However, we have some significant concerns regarding what we view as several oversights in the report that lead to overstated confidence in the results of digital forensic investigations and fail to account, in particular, for the role of human subjectivity and error or to recommend simple, noncontroversial strategies for the problems it identifies.

Moreover, this report does not seem to engage meaningfully with the reality that digital forensics is used within the context of the criminal legal system. It treats the application of digital forensics to life and liberty as incidental or separate from the technical issues of the digital investigative process. This lack of connection with the implications of the very serious problems of digital forensics (informal reviews, lack of validation, subjective analysis without cognitive bias guardrails, unnecessary intrusion into private information, and lack of transparency or documentation) is a major defect of this report.

We respectfully offer the following specific comments on the report.

INNOCENCE PROJECT

Comment No.	Page	Chapter	Text	Comment
1	1	Executive Summary	The overall finding of this report is that digital evidence examination rests on a firm foundation based in computer science. Several of the techniques had already been extensively studied and documented in the peer-reviewed literature. Others are documented more informally through community discussion forums	The mechanical process of manipulating computers is based on a firm computer science foundation, but the interpretation of the data may not be. How are attorneys supposed to assess the validity of techniques that are documented informally in community discussion forums? This report does not provide the public with a way to evaluate these informal techniques and seems to ask the reader to simply trust that they are sufficient.
2	1	Executive Summary	The application of these computer science techniques to digital investigations is sound, only limited by the difficulties of keeping up with the complexity and rapid pace of change in IT	Without a reference to the degree of uncertainty introduced by human examiners implementing the tools and analyzing the data, this statement is misleading. This sentence can be misapplied by parties seeking to deflect scrutiny from digital forensic examination. Please include the following statement to the last sentence to prevent misleading stakeholders regarding the validity and reliability of the entire digital forensic process: The application of these computer science techniques to digital investigations is sound, **but the reliability of the human interpretation of the gathered data is unknown. The digital forensic tool limitations include** ~~only limited by~~ the difficulties of keeping up with the complexity and rapid pace of change in IT. **The human interpretative component of digital forensic investigations includes incompleteness, inaccuracy, and misinterpretation.**

INNOCENCE PROJECT

Comment No.	Page	Chapter	Text	Comment
3	6	1	NIST also performed an interlaboratory study (Guttman et al. 2022) as part of its work on the scientific foundation of digital forensics. The study did not attract enough participants to draw meaningful conclusions but did demonstrate that digital forensic examiners could answer difficult questions related to the analysis of mobile phones and personal computers.	This passage refers to NISTIR 8412. It first states that the study did not attract enough participants to draw meaningful conclusions, but then draws the conclusion that "digital forensic examiners could answer difficult questions." While that statement is technically true, it is also misleading. Taken alone, the statement appears to reassure the reader of digital examiners' proficiency, however, Table 3 (p.8-10) of the black box study actually shows that examiners often get answers wrong. Twenty questions were asked in the study. In 9/20 (45%) of the questions, at least 27% of examiners gave wrong answers or skipped questions. In 3/20 (15%) of the questions, at least 49.4% of examiners gave wrong answers or skipped questions. Please revise this sentence to indicate that incorrect or skipped answers were frequent in this study.

INNOCENCE PROJECT

Comment No.	Page	Chapter	Text	Comment
4	6-7	1	After obtaining proper authorization and warrant for a search then a search can proceed. Digital evidence differs from physical evidence in the concept of search and seizure. For a physical search, the authorization covers searching the location and the seizure of objects of possible evidentiary value. In digital forensics an entire digital storage device, e.g., hard drive or flash drive is taken to then search it for evidence.	The analogy in pp. 6-7 to crime scene investigation is very helpful for understanding the specific digital forensics techniques covered in this report. However, this language should be revised to make clear that nothing inherent to the technology mandates that law enforcement look at every file on a digital device. Indeed, a frequent problem with warrants authorizing digital searches is that they are overbroad and insufficiently particular. While law enforcement may need to take possession of a physical device to perform a search, the warrant for a digital device—just like that for a physical search—must list the specific evidence that law enforcement has probable cause to believe is evidence of a crime, and there is no technological reason that law enforcement must look at individual files that go beyond those parameters. Indeed, the software addressed in this report allows for just such a targeted search. Moreover, it is essential that a defendant have a clear understanding of exactly how and by whom any search was undertaken, and this report should recommend further transparency with respect to disclosure and discovery.

INNOCENCE PROJECT

Comment No.	Page	Chapter	Text	Comment
5	7	1	In like manner, a digital investigation generates hypotheses, and the investigator searches for data artifacts, e.g., files, logged events with a time stamp, emails, etc., that can be used in evaluating observed evidence in light of alternative (opposing) hypotheses.	Using a hypothesis to drive a digital investigation presents a real risk that an investigator will seek out data to confirm that hypothesis—that is, a danger that cognitive bias will affect the investigation's outcome. Moreover, this approach exposes an individual's private information to unnecessary and unlawful exposure. This report does not sufficiently take these factors into account. Please include a recommendation that examiners be shielded from unnecessary biasing information and that the non-responsive personal information to which law enforcement has no legal right be protected from view. For example, the initial extraction of the entire contents of a digital device and culling down to items responsive to the terms of a search warrant should be undertaken by an examiner wholly unrelated to the investigation.
6	23	2	The capture-recapture method yielded a lower bound estimated population size of 11,000 with a 95% confidence interval of (9,900, 12,600). Due to the overlap between the lists and the fact that some of the total population has a zero probability of being selected in any list, the final value is interpreted as a lower bound estimate, rather than an absolute population size. This value of 11,000 US digital forensics organizations contrasts with the 409 publicly funded crime labs reported by the Bureau of Justice Statistics (Burch, Durose, and Walsh 2016). The decentralization of the digital forensics community in the United States is apparent in where digital forensics labs are found; they are not only in federal, state, and local crime labs, but also in prosecutor's offices, private consulting firms, and corporate cybersecurity operations.	This report does not communicate the urgency made apparent by the fact that there are so many digital forensics units in the U.S., but so little oversight in the form of accreditation, commissions, or laws to ensure the accurate and high-quality operation of these units. Moreover, the report fails to take into account the fact that it is private, for-profit companies that create the technology used in digital forensics and that these companies are incentivized to inflate the capabilities of their technology and disincentivized to be transparent.

INNOCENCE PROJECT

Comment No.	Page	Chapter	Text	Comment
7	20	2	KEY TAKEAWAY #2.4: The forensic examiner needs to be aware of key changes in computing technology relevant to the examination being performed. Frequent changes in digital technology introduces the possibility for incomplete analysis or for misunderstanding of the meaning of artifacts.	There should be a further recommendation concerning oversight and accreditation to address what this report identifies as necessary ongoing technical education. Moreover, it is important that any new understanding of a discredited approach or analysis be made transparent to the defense.
8	21	2	An examination of a mobile phone seized from a suspected drug dealer might begin by the examiner looking at contacts (possible customers and collaborators) and messages (setting up illegal transactions). To investigate a suspected espionage case the examiner might look for contraband (classified documents), removable device history (moving the contraband around), geolocator information (places the suspect has visited), contacts (identify collaborators), messages (extraction of planned actions) and deleted documents (hiding activity).	These searches would be appropriate only if specifically authorized by a valid search warrant. While it is true that all the searches described in this sentence might yield useful information, this sentence incorrectly implies that they are all automatic and legally proper in every investigation of a suspected drug dealer.

INNOCENCE PROJECT

Comment No.	Page	Chapter	Text	Comment
9	30	2	KEY TAKEAWAY #2.5: Not every digital forensic technique undergoes a peer review, formal testing, or error rate analysis. In general, the digital forensic community performs an informal review by providing feedback about the usefulness of techniques. This general acceptance process allows for techniques to be quickly evaluated and revised.	As we have learned through the widespread acceptance of now discredited forensic techniques such as bite mark analysis and hair microscopy, consensus and general acceptance does ensure valid techniques. This report does not communicate the urgent concern that much of digital forensic practice is informal and documented in fora that are outside the purview of the legal actors who must litigate and assess these techniques. This takeaway essentially asks criminal legal system stakeholders to simply put their trust in examiners without any way to verify their techniques. Please add language that communicates the problems that can occur in the criminal legal system if a discipline is built largely on informal reviews. Additionally, it is unclear why this Takeaway would not also include a recommendation to adhere to *IEEE 1012-2012: IEEE Standard for System, Software, and Hardware Verification and Validation* (IEEE 1012), a robust industry standard that already exists to set requirements for formal verification and validation of digital systems. The authors should definitively address in this Takeaway or in this section why they do or do not recommend the use of IEEE 1012.
10	33	4	4.1 Steps in a Digital Investigation	Please add a step to address the concerns regarding bias and unnecessary exposure of private information addressed in Comment 5, above. Moreover, the report should recommend that each of these steps be transparent to individuals whose information is being searched.

INNOCENCE PROJECT

Comment No.	Page	Chapter	Text	Comment
11	36	4	Cryptographic hashing is used to detect inadvertent or deliberate changes. Cryptographic hashing is a robust technique used in multiple high security applications. NIST publishes hashing standards as part of its cryptography program (NIST 2015a, 2015b).	This text comes from **4.3 Integrity Verification**. The integrity of the digital crime scene is essential to the validity of the examiner's analysis. If NIST publishes hashing standards that can help examiners or defense experts identify if inadvertent or deliberate changes were made to the extracted digital device data, why wasn't a recommendation made for digital examiners to use NIST's hashing standards? Please include a recommendation for digital forensic examiners to use NIST hashing standards as part of their digital forensic investigations.
12	37	4	KEY TAKEAWAY #4.1: When using techniques to recover deleted or hidden artifacts the examiner must determine the relevance of the recovered information as it may be incomplete or improperly merged with irrelevant information.	This takeaway makes clear how dependent the recovery of deleted data is on the judgment of the examiner. However, there is no discussion of how this process is documented, how examiners might make these judgments, and what protections can be made to insulate the examiner from cognitive biases. Since this process is so subjective, it will be essential for the defendant to have their own expert evaluate the forensic examiner's analysis. Please include language in this section regarding the need for documentation, transparency, cognitive bias protections, and defense expert access as mitigation for this very subjective process.
13	38	4	The main assembly of a narrative to describe the events of interest of an investigation or answering questions that arise during an investigation involves identifying, finding, and extracting relevant artifacts. A question of interest might prompt an examiner to select a specific artifact for examination. The examiner then tries to locate the selected artifact and then extract the artifact for examination.	This process has great potential to promote confirmation bias in the examiner, and this section does not describe any steps to mitigate confirmation bias. Please include language in this section regarding the need for documentation, transparency, cognitive bias protections, and defense expert access as mitigation for this very subjective process.

INNOCENCE PROJECT

Comment No.	Page	Chapter	Text	Comment
14	39	4	KEY TAKEAWAY #4.2: Searching tools have limitations based on the multiple ways that computers store information. Limitations include the type of files, types of encoding, and many other parameters. In general search tools are very effective at finding information, but there is a possibility that data will be missed because a tool does not have the capability to find it.	It is understandable that in conducting a digital search, not all data will be captured. However, in a criminal prosecution, missing data can have severe consequences. To mitigate any recovery problems, this section should recommend defense expert access to digital data. The Takeaway also references the limitations of the searching tools without offering parameters for assessing their conditions or impact. The limitations of any tool should be defined through validation processes. Please include language about the importance of defense expert access to digital devices given the fact that not all data may be captured in a search, as well as a requirement that technological limitations be documented through validation testing.
15	41	4	Artificial Intelligence (AI) tools use a technique called deep learning that can be used to uncover unseen relationships between case elements or search through data to recognize relevant items. Some AI applications have been controversial because of the introduction of unexpected, unintentional bias. Examples include facial recognition software exhibiting poor or misleading results for racial minority subjects (Grother, Ngan, and Hanaoka 2019). [...] AI tools are powerful, but not perfect and should be used with caution due to unexpected behaviors. What comes out depends on the data set used to train the AI and may not be relevant to the data at hand, and any results could be misleading and should be verified or confirmed. As with other techniques, examiner must use caution and check that AI based finding are used in the appropriate context.	The problems introduced by AI tools are serious, and the only recommendation that the authors offer is for the examiner to "use caution and check that AI based findings are used in the appropriate context." This recommendation is beyond insufficient and does not consider how cognitive biases may interfere in the "checking" process. Please include language in this section recommending the need for documentation, transparency, cognitive bias protections, and defense expert access as mitigation for these severe problems generated by AI tools.

Comment No.	Page	Chapter	Text	Comment
16	42	4	KEY TAKEAWAY #4.3: If someone has taken steps to change information in digital evidence to mislead an examiner, it may be difficult to detect the changes. Depending on the sophistication of the manipulation, identification of the changes relies on the skill of the examiner.	First, this Takeaway states that digital evidence can be manipulated, is difficult to detect, and that the criminal legal system is reliant on the skill of the examiner to identify this problem. However, this report does not include any language regarding how to proficiency test an examiner or determine their competence. NISTIR 8412 has demonstrated that examiners make mistakes and also indicated that 71-75% of participants passed a mobile or hard drive proficiency test in the past five years (see pp. 19 and 31). If we are to give examiners this much trust, there must be a way to demonstrate their capacity. Please include language in this section recommending the need to develop robust competency or proficiency testing programs and the frequency with which they should be given. Second, this Takeaway asserts that information can be changed to mislead a forensic examiner without any assessment of the technical capabilities necessary to undertake such measures or which kinds of files or devices would be most susceptible to such manipulation. The report should clarify what kind of technological capabilities would be necessary to, for example, disguise one type of file commonly found on a phone as another one. Such a scenario is often used by law enforcement to justify unnecessarily overbroad searches of digital devices.

INNOCENCE PROJECT

Comment No.	Page	Chapter	Text	Comment
17	41	4	For a forensic technique or method to be considered validated it should be shown to be fit for purpose otherwise defined as "the process of providing objective evidence that the method is good enough to do the job required by the end user". Validation can give a false indication of "fitness for purpose" that becomes apparent later.	This language flags an important problem, but does not provide more information about what happens if a technique or method is later determined to not be fit for purpose. What are the consequences of this? How might it impact an analysis? What happens to the people who have been adversely impacted by the flaws in the technique or method?
18	43	4	There have been several papers published on validation of digital forensics methods (Regulator 2020; Arshad, Jantan, and Abiodun 2018; Beckett and Slay 2007; Brunty 2011; Casey 2011a; Craiger et al. 2006; Guo, Slay, and Beckett 2009; Horsman 2018; Horsman 2019; Marshall and Paige 2018; Risinger 2018; SWGDE 2014; Wilsdon and Slay 2006). Some of these papers seem to confuse validation of a method and verification of a software tool and try to fold the two activities together instead of keeping them separate. The guidance from the UK Forensic Science Regulator (Regulator 2020) seems the most clear and includes consideration of risk assessment of the method, documentation of acceptance criteria and possible outcomes.	Given the central role of validation in the forensic science process, it would be important for the authors to assist the readers in better understanding which publications among those mentioned in this section properly or improperly describe validation and verification. The manner in which these reports are described here skirts an important issue, and the criminal legal system audience needs these experts to point out which publications we can rely upon rather than leaving it to a lay audience to debate. Additionally, if the UK FSR is the resource with the best guidance, then the authors should explicitly make that clear. It is also unclear why the authors do not include IEEE 1012 on this list. Please either recommend the UK FSR guidance if it is the best resource for validation and verification information for digital forensics methods or list the publications among those listed that provide accurate guidance on the topic. Additionally, the authors should definitively address why they do or do not recommend the use of IEEE 1012.
19	43	4	The general validation and verification for a given version of a tool can be done once. It does not need to be performed by every lab.	Is this true for every digital forensics tool? How do we know which digital tools require verification by specific labs and which do not?

INNOCENCE PROJECT

Comment No.	Page	Chapter	Text	Comment
20	45	4	Another problem is that the properties and characteristics of digital data changes with the software environment as the technology evolves over time and an error rate valid at one point in time might not apply at any other point in time. KEY TAKEAWAY #4.4: Digital processes tend to have systematic rather than random errors. Therefore, an error mitigation analysis provides more information and is the correct way to manage uncertainty. Asking for an error rate is only useful where there are random errors. KEY TAKEAWAY #4.5: When error rates are provided, it is important for the user to understand the context of the numbers. Errors in computer science techniques tend to be so small as to be negligible. For some forensic techniques, the error rates may vary significantly based on attributes of the technology and usage patterns.	The focus on error rates here is a bit of a deflection of the real problem. While the technical application of tools may not incur error, the interpretation of collected data is not well studied (see Black Box Study). It is this interpretive phase that introduces very serious biases and can mislead investigations. The statement in takeaway #4.5 that "Errors in computer science techniques tend to be so small as to be negligible" can be misunderstood and misused by examiners seeking to boost their credibility if it is not also paired with a statement that explicitly states that errors in the interpretive portion of the process are unknown. Lastly, this section of the report does not provide evidence for the statement "errors in computer science techniques tend to be so small as to be negligible." Even small errors have been demonstrated to have serious consequences, and courts have not been an effective arbiter of this risk.[1] Please provide evidentiary support in this section for the statement regarding negligible error rates. If no supporting evidence can be found, the statement must be removed from the Takeaway. Please also add language to Takeaway #4.5 to state the following: TAKEAWAY #4.5: When error rates are provided, it is important for the user to understand the context of the numbers. Errors in computer science techniques tend to be so small as to be negligible. For some forensic techniques, the error rates may vary significantly based on attributes of the technology and usage patterns. **Errors in the human component, the interpretation of the data gathered through the computer science techniques, is unknown. Therefore, a complete analysis of digital forensic techniques requires an evaluation of the application of the tools and the interpretation of that data. There is a dearth of research in the latter.**

INNOCENCE PROJECT

Comment No.	Page	Chapter	Text	Comment
21	46	4	4.10.2 Observed Errors	This section describes the problems with forensic tool implementations, namely incompleteness, inaccuracy, and misinterpretation. It does not offer a way to detect these errors, nor does it recommend that these errors be corrected when identified. This section is incomplete without addressing these two issues.
22	56	5	5 Conclusions	This section does not suggest cognitive bias protections, nor does it recommend a mechanism for identifying and remediating errors and notifying impacted parties when these errors occur. Please include language referencing these recommendations in this section.
23	56	5	The application of these computer science techniques to digital investigations is sound, only limited by the difficulties of keeping up with the complexity and rapid pace of change in IT.	Please see Comment #2, above.

[1] *See, e.g.*, https://www.pbwt.com/second-circuit-blog/second-circuit-oks-use-of-now-defunct-dna-testing-method