Cutting-Edge Evidence: Strengths and Weaknesses of New Digital Investigation Methods in Litigation

ALEXA KOENIG AND LINDSAY FREEMAN†

The increased use of digital technologies in daily life has led to a steep rise in the introduction of highly technical evidence and expert witness testimony in criminal and civil litigation. The growing use of novel, quickly-developing investigation methods for digital evidence presents several challenges related to the difficulty lay persons have in judging complex forensic methodologies. The lack of judicial and legal training in the underlying methods and their potential vulnerabilities can result in fact-finders who over-rely on experts' conclusions without properly interrogating the evidence themselves.

While many of the scientific and analytical methods employed by digital investigators can be promising additions to investigative toolkits, enthusiasm for these techniques should be tempered with healthy skepticism—and knowledge of the most helpful questions to ask about new investigative processes. In this Article, we identify the very real vulnerabilities in digital open source investigations and encourage careful analysis of each component in order to mitigate the risks. We recommend that investigators preserve digital material according to established forensic standards and carefully record the steps of their online investigation and analysis. Expert witnesses should be strictly prohibited from giving opinions on matters that stretch beyond the scope of their education, training, and well-established expertise. Lawyers and judges must be prepared to ascertain the reliability and validity of digital open source investigations and their findings through thorough interrogation of the underlying data. As a best practice, digital evidence should be triangulated with physical, testimonial, or other documentary evidence whenever possible. If conducted carefully and professionally, digital open source investigations can offer tremendous value for both civil and criminal proceedings.

† Lecturer and Executive Director, Human Rights Center, UC Berkeley School of Law and Technology Law and Policy Program Director, Human Rights Center, UC Berkeley School of Law, respectively. The authors thank Anthony Ghaly for his research and drafting support; the editors at *Hastings Law Journal* for their careful work; as well as Kelly Matheson, Yvonne McDermott, Annie O'Reilly, Eric Stover and the participants of the Pound Civil Justice Institute's 2021 Symposium on the Internet and the Law at UC Hastings, College of the Law for their feedback on earlier drafts. Any errors are, of course, the authors' own.

TABLE OF CONTENTS

INTRODUCTION

On the morning of January 6, 2021, the day Congress was set to affirm Joe Biden's victory in the United States presidential election, a large crowd of Trump supporters gathered on the Capitol lawn. At noon, then-President Donald Trump spoke to his supporters. Claiming—despite all evidence to the contrary—that he and not Biden had won the election, he told his followers to "show strength"[1] and to "walk [with him] down to the Capitol."[2] Many turned to march. By 1:00 p.m., the crowd had overwhelmed the building's security and breached its barricades. Angry rioters armed with camera phones and, in some cases, weapons, swarmed the building while hundreds more clashed with officers outside, forcing legislators and staff into hiding. It was several hours before the sergeant-at-arms was able to declare the building secure.[3] By then, the damage was done—the building trashed, five people dead or dying, and democracy degraded.

The events that day led to numerous civil and criminal cases. The United States Department of Justice responded by launching what has been described as the largest investigation in U.S. history, both in terms of the number of defendants and the volume of digital evidence.[4] Their work was supplemented by an army of online citizen investigators who provided useful tips and analysis to the Federal Bureau of Investigation and other law enforcement to assist in the identification of the insurrectionists. By October 2021, close to 700 people had been charged for their alleged role in the riots.[5]

Civil litigation has been similarly extensive. Cases have ranged from a lawsuit brought by Democratic Representative Eric Swalwell of California against former President Trump and others who spoke at the rally,[6] to a lawsuit brought by the former President to block the House January 6 Select Committee

1. Charlie Savage, *Incitement to Riot? What Trump Told Supporters Before Mob Stormed Capitol*, N.Y. TIMES (Jan. 10, 2021), https://www.nytimes.com/2021/01/10/us/trump-speech-riot.html.

2. *Id.*

3. *See* Lauren Leatherby, Arielle Ray, Anjali Singhvi, Christiaan Triebert, Derek Watkins & Haley Willis, *How a Presidential Rally Turned into a Capitol Rampage*, N.Y. TIMES (Jan. 12, 2021), https://www.nytimes.com/interactive/2021/01/12/us/capitol-mob-timeline.html (showing a timeline of the events of January 6, 2021); George Petras, Janet Loehrke, Ramon Padilla, Javier Zarracina & Jennifer Borresen, *Timeline: How the Storming of the U.S. Capitol Unfolded on Jan. 6*, USA TODAY (Jan. 6, 2021), https://www.usatoday.com/in-depth/news/2021/01/06/dc-protests-capitol-riot-trump-supporters-electoral-college-stolen-election/6568305002/.

4. Willy Lory, *January 6 Capitol Riot Anniversary: Biggest Criminal Investigation in US History*, THE NAT'L NEWS (Jan. 4, 2022), https://www.thenationalnews.com/world/us-news/2022/01/04/capitol-riot-anniversary-biggest-criminal-investigation-in-us-history.

5. *See, e.g.*, Madison Hall, Skye Gould, Rebecca Harrington, Jacob Shamsian, Azmi Haroun, Taylor Ardrey, & Erin Snodgrass, *At Least 800 People Have Been Charged in the Capitol Insurrection So Far. This Searchable Table Shows Them All*, INSIDER, https://www.insider.com/all-the-us-capitol-pro-trump-riot-arrests-charges-names-2021-1 (Mar. 17, 2022, 5:13 PM).

6. Carrie Johnson, *A Lawsuit Against Jan. 6 Rally Speakers Forces DOJ to Consider Who's Legally Immune*, NPR: POLITICS (July 26, 2021, 4:03 PM), https://www.npr.org/2021/07/26/1020786560/a-lawsuit-against-jan-6-rally-speakers-forces-doj-to-consider-whos-legally-immun.

from releasing information about his alleged involvement.[7] This was joined by a civil rights lawsuit filed by seven Capitol police against Trump, the Trump campaign, the Stop the Steal limited liability corporation, and members of the Proud Boys, the Oath Keepers, and others who were allegedly involved in the violence.[8]

Given insurrectionists' widespread use of smartphones and social media to share photos, videos, and other information online as events unfolded, such content has proven especially critical to finding and identifying suspects.[9] Digital detectives have gathered potentially useful information that is aiding both criminal and civil litigation. However, in addition to the clear opportunities, there are drawbacks to using this information as evidence in the courtroom. Having written extensively about the opportunities presented by digital open source information elsewhere,[10] in this Article, we focus on those limitations.

Digital open source information pulled from social media and other online spaces is playing an increasingly important role in establishing the who, what, where, why, when, and how of world events that result in civil and criminal cases, both domestic and international.[11] According to researchers, between 2010 and 2017, the Ninth Circuit alone saw a 350% increase in the use of social media evidence.[12] Even more dramatically, between 2007 and 2017, California's state courts experienced a 3,933% increase.[13] Nationally, 97% of Americans now own a cellphone, with 85% possessing a smartphone capable of recording videos and posting them to the Internet.[14] Globally, by 2018 humans were

7. Caroline Linton, *Trump Sues House January 6 Committee in Attempt to Block Release of Documents*, CBS NEWS (Oct. 19, 2021, 8:10 AM), https://www.cbsnews.com/news/trump-lawsuit-january-6-capitol-riot-committee-court-documents-executive-privilege.

8. Complaint at 9–14, Smith v. Trump, No. 1:21-cv-02265 (D.D.C. Aug. 26, 2021), https://www.lawyerscommittee.org/wp-content/uploads/2021/08/001-Complaint.pdf.

9. *Id.* at 15–53.

10. *See, e.g.*, Lindsay Freeman, *Prosecuting Atrocity Cases with Open Source Evidence: Lessons from the International Criminal Court*, *in* DIGITAL WITNESS: USING OPEN SOURCE INFORMATION FOR HUMAN RIGHTS INVESTIGATION, DOCUMENTATION AND ACCOUNTABILITY 48 (Sam Dubberley, Alexa Koenig & Daragh Murray, eds., 2020); Alexa Koenig, *Open Source Evidence and Human Rights Cases: A Modern Social History*, *in* DIGITAL WITNESS: USING OPEN SOURCE INFORMATION FOR HUMAN RIGHTS INVESTIGATION, DOCUMENTATION AND ACCOUNTABILITY 32 (Sam Dubberley, Alexa Koenig & Daragh Murray eds., 2020); Alexa Koenig & Lindsay Freeman, *Strengthening Atrocity Cases with Digital Open Source Investigations*, LIEBER INST. W. POINT (Apr. 1, 2021), https://lieber.westpoint.edu/strengthening-atrocity-cases-digital-open-source-investigations; Lindsay Freeman & Alexa Koenig, *Links in the Chain: How the Berkeley Protocol is Strengthening Digital Investigations and International Collaboration*, *in* VERIFICATION IN THE AGE OF GOOGLE (forthcoming 2022).

11. Open source information is defined as information that is publicly accessible on the internet and that any person can collect through observation, request or purchase. HUM. RTS. CTR., UC BERKELEY SCH.OF L. & U.N. OFF. OF THE HIGH COMM'R FOR HUM. RTS., BERKELEY PROTOCOL ON DIGITAL OPEN SOURCE INVESTIGATIONS 3 (2020), https://www.ohchr.org/Documents/Publications/OHCHR_BerkeleyProtocol.pdf [hereinafter BERKELEY PROTOCOL].

12. Lynne Graves, William Bradley Glisson & Kim-Kwang Raymond Choo, *LinkedLegal: Investigating Social Media as Evidence in Courtrooms*, 38 COMP. L. & SEC. REV. 1, 12 (2020).

13. *Id.*

14. *Mobile Fact Sheet*, PEW RSCH. CTR. (Apr. 7, 2021), https://www.pewresearch.org/internet/fact-sheet/mobile.

creating "more than 2.5 quintillion (eighteen zeros) bytes of data every day,"[15] a number that is expected to grow to 463 exabytes by 2025.[16] Litigants have increasingly relied on this digital content to strengthen fact-finding in a diverse array of cases, from personal injury cases where the scope of plaintiffs' harms are in dispute,[17] to trademark cases where social media may help establish consumers' brand confusion,[18] to defamation cases where the accuracy of negative claims are at issue,[19] to civil litigation brought under the Alien Tort Claims Act for alleged human rights violations[20]—as well as national and international criminal prosecutions.[21]

Given the growing use of digital information and communication technologies, the field of practice known as "digital open source investigations" is rapidly expanding.[22] Such investigations—which rely heavily on user-generated content such as videos and photos posted to social media, as well as on commercial satellite imagery[23]—differ both qualitatively and quantitatively from more traditional, analog forms of open source information, such as non-governmental organization reports, newspapers and radio broadcasts.

While the use of digital open source investigation techniques and the engagement of lay investigators can offer tremendous value to both private litigants and law enforcement, the introduction of any new scientific or technical investigative methods and any resulting evidence into legal proceedings also comes with significant risks.[24] In seeking accountability for the events of January 6, lawyers and investigators have faced a "glut of social media evidence," a phenomenon that has both helped and hindered their ability to parse fiction from fact.[25] The events of that day have also underscored the difficulties

15. *Forensic Examination of Digital Devices in Civil Litigation: The Legal, Ethical and Technical Traps*, AM. BAR ASS'N (Mar. 1, 2016), https://www.americanbar.org/groups/professional_responsibility/publications/professional_lawyer/2016/volume-24-number-1/forensic_examination_digital_devices_civil_litigation_legal_ethical_and_technical_traps.

16. Branka Vuleta, *How Much Data is Created Every Day? [27 Staggering Stats]*, SEEDSCIENTIFIC: BLOG (Oct. 28, 2021), https://seedscientific.com/how-much-data-is-created-every-day.

17. Vasquez-Santos v. Mathew, 168 A.D.3d 587, 588 (N.Y. App. Div. 2019) (highlighting where the defendant was permitted access to social media posts in which the plaintiff was tagged that showed the plaintiff playing basketball post-accident and could be used to rebut claims that he could no longer play).

18. *See, e.g.*, Moroccanoil, Inc. v. Marc Anthony Cosmetics, Inc., 57 F.Supp.3d 1203, 1213 (C.D. Cal. 2014) (considering the admissibility of Facebook posts in which customers discussed Moroccanoil's products).

19. Ben Meyerson & Andrew Wang, *Tweet Lawsuit: Chicago Landlord Sues Ex-Tenant Over Tweet Complaining About Apartment*, CHI. TRIB. (July 29, 2009), https://www.chicagotribune.com/news/chi-twitter-suit-29-jul29-story.html.

20. Alien Tort Claims Act, 28 U.S.C. § 1350 (2012).

21. *See, e.g.*, Prosecutor v. Al-Werfalli, ICC-01/11-01/17-13, Warrant of Arrest ¶¶ 3, 12–16, 22, 29 (Aug. 15, 2017) (showing that the videos posted to social media provided the basis for an international arrest warrant).

22. *See generally* Alexa Koenig, Emma Irving, Yvonne McDermott & Daragh Murray, *New Technologies and the Investigation of International Crimes: An Introduction*, 19 J. OF INT'L CRIM. JUST. 1 (2021); BERKELEY PROTOCOL, *supra* note 11.

23. Rebecca J. Hamilton, *User-Generated Evidence*, 57 COLUM. J. OF TRANSNAT'L L. 1, 58 n.241 (2018).

24. The risk of wrongfully convicting an innocent person while the real perpetrator goes free.

25. *Mistaken Identity: FBI Probe into Jan 6 Rioters Sees Challenges*, AL JAZEERA (May 6, 2021), https://www.aljazeera.com/news/2021/5/6/mistaken-identity-fbi-probe-into-jan-6-rioters-sees-challenges.

and limitations endemic to the field of image comparison and interpretation—a practice that includes matching visual clues in videos, photographs, and satellite imagery—raising the possibility that individuals could be misidentified and mistakenly implicated in wrongdoing.[26]

While several scholars have now written about the advantages of relying on digital open source information from a plaintiff or prosecution perspective (including us),[27] few have addressed the vulnerabilities of these methods from a defense perspective.[28] Potential shortcomings, which are common to both civil and criminal law should be acknowledged for a number of reasons—not only because it is important for attorneys and investigators to be prepared for them as digital open source investigations increase in popularity, but to ensure that the integration of these materials into legal processes does not harm due process or impede the truth.

In this Article, we present a roadmap of how such user-generated evidence could be challenged in both civil and criminal cases. In particular, we discuss the components of a digital open source investigation that are most vulnerable to cross-examination—six lines of inquiry that attorneys can use to probe the quality of such investigations, and for which both proffering attorneys and expert witnesses should be prepared. These include challenges to (1) the investigator's qualifications and experience, (2) the investigative process, (3) the evidence itself, (4) the analytical conclusions, (5) the witnesses' testimony about the evidence and analysis, and (6) the in-court presentation of the evidence.

We take a critical perspective to exploring the use of these digital methods with the goal of strengthening the quality and professionalism of the results. Our critical perspective is not meant to hinder the growing use of digital open source information, but rather to ensure that such information is introduced into courtrooms responsibly and effectively. Lowering the bar, only to introduce mistaken findings in cases and produce bad verdicts, can do significant damage to the long-term credibility of these techniques and the legitimacy of our legal system.

---

26. *Id.*

27. *See, e.g.*, Emma Irving, *And So It Begins. . . Social Media Evidence in an ICC Arrest Warrant*, OPINIOJURIS: BLOG (Aug. 17, 2017), http://opiniojuris.org/2017/08/17/and-so-it-begins-social-media-evidence-in-an-icc-arrest-warrant; Koenig & Freeman, *supra* note 10; Lindsay Freeman, *Digital Evidence and War Crimes Prosecutions: The Impact of Digital Technologies on International Criminal Investigations and Trials*, 41 FORDHAM INT'L L.J. 283, 307–27 (2018); Alexa Koenig, Felim McMahon, Nikita Mehandru & Shikha Silliman Bhattacharjee, *Open Source Fact-Finding in Preliminary Examinations*, *in* 2 QUALITY CONTROL IN PRELIMINARY EXAMINATIONS 681 (Morten Bergsmo & Carsten Stahn eds., 2018); DIGITAL WITNESS: USING OPEN SOURCE INFORMATION FOR HUMAN RIGHTS INVESTIGATION, DOCUMENTATION AND ACCOUNTABILITY (Sam Dubberley eds., 2020) (multiple sources).

28. *See* Yvonne McDermott, Alexa Koenig & Daragh Murray, *Open Source Information's Blind Spot: Human and Machine Bias in International Criminal Investigations*, 19 J. INT'L CRIM JUST. 85, 88 (2021) for initial critiques.

I. CHALLENGING THE INVESTIGATOR

Due to the relative newness of social media and smartphones, many of the lay and professional investigators conducting open source investigations have received little or no formal training for the job. Instead, digital open source investigators are often self-taught, drawing on skills from a number of different fields of practice and educational resources.[29] Of course, the fact that there is not yet a well-established field of practitioners does not preclude their participation in justice processes.[30] However, the absence of standard procedures grounded in proper testing and objectively verifiable successes makes such investigators particularly vulnerable to challenge, providing the defense with multiple lines to attack their credibility on *voir dire*.

Open source investigators do not necessarily have the academic credentials or certifications that are typical for expert witnesses who attest to the validity of scientific evidence in court.[31] Defense attorneys may argue that without formal scientific training the investigator or analyst may not adhere to the same standards or take the same precautions against bias as those who have that training. Defense attorneys may also argue that such lay investigators do not understand the methodological value of safeguards such as peer review, working with multiple hypotheses, and clear documentation of the investigative process.

By contrast, professional investigators and forensic experts have usually received formal training.[32] The typical corollary to formal training would be years of experience, ideally under the supervision of a more experienced investigator. Today, in many large organizations, open source investigations are conducted by relatively young or junior members of a team because of their facility with digital technologies.[33] However, even in this newer information

29. A number of digital open source investigation "toolkits" and trainings support the learning of new tools and methods. *See, e.g.*, Justin Nirdine, *OSINT Framework*, OSINT FRAMEWORK, https://osintframework.com (last visited July 1, 2022); *OSINT Tools*, OSINT TECHNIQUES, https://www.osinttechniques.com/osint-tools.html (last visited July 1, 2022); *Open Source Intelligence (OSINT) Tool and Resources,* OSINT.LINK, https://osint.link/?__cf_chl_jschl_tk__=pmd_zYiy28FklWuQb6uaf8lM6iDHwxZzDCfOeCRx1TR1v9w-1629768034-0-gqNtZGzNAdCjcnBszQh9 (last visited July 1, 2022). See *Trainings and Workshops*, U.C. BERKELEY HUM. RTS. CTR., https://humanrights.berkeley.edu/resources/trainings-and-workshops (last visited July 1, 2022) for examples of training courses; *Open Source Investigations for Human Rights*, AMNESTY INT'L, https://advocacyassembly.org/en/partners/amnesty (last visited July 1, 2022); *Training*, BELLINGCAT, https://www.bellingcat.com/tag/training (last visited July 1, 2022); Michael Bazzell, *Online OSINT Video Training*, INTEL TECHNIQUES, https://inteltechniques.com/training.html (last visited July 1, 2022).

30. *See, e.g.*, *About*, BELLINGCAT, https://www.bellingcat.com/about (last visited July 1, 2022) (discussing the lay investigation team's contributions to justice efforts).

31. For example, the founder of Bellingcat, Eliot Higgins, is self-taught. *See* ELIOT HIGGINS, WE ARE BELLINGCAT: GLOBAL CRIME, ONLINE SLEUTHS, AND THE BOLD FUTURE OF NEWS 67 (2021).

32. *See, e.g.*, *UT National Forensic Academy*, UT L. ENF'T INNOVATION CTR., https://leic.tennessee.edu/home/training/forensic-training/national-forensic-academy/#:~:text=The%20National%20Forensic%20Academy%20(NFA,from%20across%20the%20United%20States (last visited July 1, 2022); *Upcoming Regular Courses*, INST. FOR INT'L CRIM. INVESTIGATIONS, https://iici.global/courses (last visited July 1, 2022) for an overview of investigation training topics.

33. *See, e.g.*, Gretchen Kell, *Doctor, Lawyer, Open Source Investigator? New Field Plucks Berkeley Grads*, BERKELEY NEWS (May 1, 2019), https://news.berkeley.edu/2019/05/01/this-one-doctor-lawyer-open-

environment, basic principles of research and analysis apply—and thus having a deep understanding of the building blocks of quality research is critical.

Open source investigators are often generalists: people who understand how and where individuals communicate online and have a sense of the broad array of tools and platforms that can assist in finding and evaluating that information.[34] They are jacks of all trades, often masters of none. While they draw on a vast repository of online data and software, experimenting with tactics affiliated with well-established disciplines, including digital forensics, data science, geospatial analysis, forensic image and video comparison, and forensic image and video interpretation, they are rarely experts in any of these fields. While the ability to self-educate is an important one, such breadth of expertise—while invaluable—should complement and not supplant the depth of understanding that comes with expertise in established fields of practice.

Disciplinary canons exist for a reason: to ensure quality work, minimize blind spots, and ideally protect the legitimacy of those disciplines. In the international open source investigation context, the *Berkeley Protocol on Digital Open Source Investigations* offers professional, methodological, and ethical principles that speak to the necessary ability of the investigator to credibly perform various open source investigation tasks, and to assess when an expert in a subfield should be called in to either supplant or supplement a generalist's process and analysis.[35] The principles require indicia of accountability, competency, accuracy, objectivity, legality, humility, independence, transparency, and security awareness, among others.[36]

Other critical considerations include whether the investigator has pre-existing biases that may prejudice the investigation, and if so, the kinds of pre-existing biases; whether the investigator may have been biased or prejudiced by the information environment; and whether the investigator is susceptible to external influence that may call into question the quality of the conclusions, either because of funding sources, reputational concerns, or otherwise.[37]

II. CHALLENGING THE INVESTIGATION PROCESS

After examining the investigator's qualifications and competencies, lawyers should assess whether there are any vulnerabilities to the quality of the investigation itself—in particular, the process by which the investigator identified, collected, and preserved online information. Such an examination

source-investigator-new-field-seeks-berkeley-grads (discussing the newness of open source investigations and the training of college students to do this work).

34. Examples of the wide array of tools that an open source investigator might use are available on open source investigation "dashboards," which are essentially websites that aggregate hyperlinks to those tools. Nirdine, *supra* note 29; Bruno Mortier, OSINT START.ME, https://start.me/p/ZME8nR/osint?locale=en (last visited July 1, 2022).

35. BERKELEY PROTOCOL, *supra* note 11, at 9–15.

36. *Id.* at 11–15.

37. *See, e.g.*, McDermott et al., *supra* note 28, at 88.

should evaluate the thoroughness and objectivity of online inquiries to ensure that the investigator has not intentionally or inadvertently cherry-picked open source information, the criteria that was used to determine what to collect, and the forensic soundness of the tool used to capture and preserve the information.

One of the biggest opportunities *and* challenges for online investigators is the sheer scale of information flooding the Internet. As of 2021, approximately 6,000 tweets were being sent out every second;[38] by late 2018 more than 400 hours of stories were being uploaded to Instagram each day; and 500 hours of videos were being uploaded to YouTube each minute.[39] This volume makes it impossible for human investigators to thoroughly review all potentially relevant information online. The vast and dynamic nature of the Internet and the significant volume of information it holds make digital investigations vulnerable to multiple challenges regarding the investigator's decision-making process, especially their determination of what is and is not relevant and what they choose to review and collect. This becomes especially critical in cases where relevant information has been removed from the Internet between the time of the event and the start of a formal legal investigation, for example, due to social media content moderation or the user regretting their post. In such cases, the item collected by a lay investigator may be the only version available by the time the case gets to trial.

The average person may not be familiar with the many biases that influence online investigations, but professionals and lay investigators who document information from the Internet should be aware of such biases and should take proactive measures to counter them.[40] Three categories of bias are especially important to consider. The first is *access bias*, which relates to who has access to digital tools and who does not, and thus whose perspectives and experiences are and are not represented online.[41] The second is *algorithmic bias*, which refers to a search engine's programming and how it determines which search results should be shown to its users.[42] Search results from Google, Yahoo, Bing, Yandex, Baidu and DuckDuckGo differ between users, and between each other. Their search algorithms use several data points—including, but not limited to, the user's location, device, browser, and prior search history—to prioritize and

38. *See "Twitter by the Numbers: Stats, Demographics & Fun Facts*," OMNICORE (Feb. 22, 2022), https://www.omnicoreagency.com/twitter-statistics.

39. *See generally* Lindsay Freeman & Raquel Vazquez Llorente*, Finding the Signal in the Noise: International Criminal Evidence and Procedure in the Digital Age*, 19 J. OF INT'L CRIM. JUST. 163 (2021); *see also* Anmar Frangoul, *With Over 1 Billion Users, Here's How YouTube is Keeping Pace with Change*, CNBC, https://www.cnbc.com/2018/03/14/with-over-1-billion-users-heres-how-youtube-is-keeping-pace-with-change.html (Mar. 14, 2018).

40. See generally RICHARDS J. HEUER JR., PSYCHOLOGY OF INTELLIGENCE ANALYSIS (1999), for an overview of biases especially relevant to digital open source investigations; *see also* McDermott et al., *supra* note 28, at 100.

41. *See* McDermott et al., *supra* note 28, at 89.

42. *Id.*

customize which websites are displayed.[43] Therefore, digital investigators must take proactive measures to maximize the neutrality of their search results. The specific keywords and languages they use along with the Boolean operators deployed to connect sources and keywords, can also radically influence results.[44] The third category of bias consists of the investigator's *cognitive biases*, which may influence not only how and where the investigator searches for information, but how they interpret results, what they choose to collect and preserve, and what they disregard.[45]

Another issue is documentation: one of the biggest differences between amateur and professional open source investigators is often how they document their investigation—or fail to do so. Accepted standards require that all technical processes, such as the extraction and processing of images, must be documented in detailed contemporaneous notes.[46] These notes should include details of any hardware or software used to process the information, and the parameters applied. Such notes may be disclosed in court. Investigators without legal or other formal training may be unaware of documentation expectations, which can have significant downstream effects. As one example, poor documentation notably led to the exclusion of Facebook evidence in *Regina v. Hamdan*[47]—a Canadian terrorism case that relied on the defendant's posts, which were excluded because the various agencies involved in the investigation did not document their collection, collected information inconsistently, and failed to use forensic software that would have automatically created a record of the process.[48]

In examining the investigation process, lawyers should scrutinize whether the open source investigator can provide clear, thorough, and, to the extent possible, contemporaneous documentation of the online inquiry and collection processes; whether those processes can be audited; and whether the investigator can account for any documentation gaps or procedural violations.

---

43. *How Search Algorithms Work*, GOOGLE, https://www.google.com/search/howsearchworks/algorithms/ (last visited July 1, 2022).

44. *See* McDermott et al., *supra* note 28, at 92.

45. *See, e.g.*, Alexa Koenig & Ulic Egan, *Power and Privilege: Investigating Sexual Violence with Digital Open Source Information*, 19 J. INT'L CRIM. JUST. 55, 62–63 (2021); Alexa Koenig & Ulic Egan, *Hiding in Plain Site: Using Online Open-Source Information to Investigate Sexual Violence and Gender-Based Crimes*, *in* TECHNOLOGIES OF HUMAN RIGHTS REPRESENTATION (Alexandra Moore & James Dawes, eds., forthcoming 2022) (both describing how cognitive biases can interview with the effective investigation of international crimes).

46. *See, e.g.,* 2 NAT'L CRIME AGENCY, CPS, METRO. POLICE & FORENSIC SCI. REGULATOR, FORENSIC IMAGE COMPARISON AND INTERPRETATION EVIDENCE: GUIDANCE FOR PROSECUTORS AND INVESTIGATORS 5 (2016), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/912880/Image_Comparison_and_Interpretation_Guidance_Issue_2.pdf.

47. *See generally* Regina v. Othman Ayed Hamdan, [2017] S.C.R. 1770 (Can.).

48. *See generally id.*

III. CHALLENGING THE EVIDENCE

After examining the investigation process, the opposing party should look for vulnerabilities in the evidence—the digital information on which the analyst or expert bases their conclusions. As with other types of evidence, lawyers must authenticate digital files—establish that an item is what the proffering party says it is—in order to have it admitted at trial. The malleability and ephemerality of digital material, however, make such evidence especially vulnerable to authenticity challenges.[49] Proving the authenticity of online evidence is further complicated by anonymous and pseudonymous sources, lack of provenance, and widespread mis- and disinformation. The Internet is replete with inauthentic or misleading sources—bots and botnets,[50] cyborgs,[51] imitator accounts,[52] sock puppets,[53] and pseudonyms—complicating the ability to determine who created and who posted the item.[54] Authentication generally requires establishing an item's author/creator, provenance, chain of custody, and integrity. There is a quickly growing series of cases in which courts have refused to consider social media-derived evidence because of a lack of appropriate authentication (for example, finding that authenticity cannot be established by the person who finds the information on the Internet, but that more is needed).[55]

Author/creator and provenance can be assessed together in digital open source investigations, as the term *source* is often used broadly to encompass both the platform on which a digital item was found and the user who uploaded it to that platform. If the content's origins are unknown, then reverse image searches and websites like the Wayback Machine that offer a historical timeline and

---

49. For an overview of various authentication techniques for digital evidence, see *Image Alteration (Forgery) Detection: An Overview of Passive Techniques*, JONATHAN W. HAK, Q.C.: BLOG (Oct. 26, 2021), https://www.jonathanhak.com/2021/10/26/image-alteration-forgery-detection-an-overview-of-passive-techniques/.

50. A bot is an automated social media account. A botnet is a network of private computers taken over by malicious software to run multiple social media accounts. Botnets can be used to amplify a message and give the impression that many people are interested in something, when it really comes from one source. *See* DFRLab, *Human, Bot or Cyborg*, MEDIUM (Dec. 23, 2016), https://medium.com/@DFRLab/human-bot-or-cyborg-41273cdb1e17.

51. A cyborg is a social media account that is sometimes run by a human and other times automated like a bot. *See id*.

52. An imitator account is a social media account created in the name of someone else, usually a celebrity, in order to pretend to be that person online.

53. A sock puppet is a social media account with a fake identity. *See* Technisette, Sector035, Micah Hoffman & Dutch_OsintGuy, *The OSINT Puppeteer*, OSINTCURIO.US (Dec. 27, 2018), https://osintcurio.us/2018/12/27/the-puppeteer/.

54. Carlotta Dotto & Seb Cubbon, *How to Spot a Bot (or Not): The Main Indicators of Online Automation, Coordination and Inauthentic Activity*, FIRST DRAFT NEWS (Nov. 28, 2019), https://firstdraftnews.org/articles/how-to-spot-a-bot-or-not-the-main-indicators-of-online-automation-co-ordination-and-inauthentic-activity/.

55. One notable example is a trademark dispute in which the complainant argued that Facebook comments that had simply been screen-shotted should not be admissible since the posts had not been appropriately authenticated. The court agreed, citing Internet Specialties W., Inc. v. ISPWest, 05-cv-3296-FMC-AJWX, 2006 WL 4568796 (C.D. Cal. Sept 19, 2006) for the idea that the burden of authentication cannot be properly met "by the person who went to the website and printed out the home page" since "anyone can put anything on the internet." Moroccanoil, Inc. v. Marc Anthony Cosmetics, Inc., 57 F. Supp. 3d 1203, 1213 n.5 (C.D. Cal. 2014).

record of online data may be used to help the analyst establish the content's provenance and author.[56] Investigators should start by establishing whether the content can be attributed to a particular person or organization, and whether it is primary (first-hand knowledge) or secondary (including second-hand information or commentary about primary content).

The investigator will also have to address the information's chain of custody. There are two chains of custody to consider when dealing with digital open source evidence: the first refers to possession or control of an item from when it is created to its collection by the investigator (this is often referred to as provenance); the second refers to custody, possession, or control of the item from the time of its collection to its presentation in court.[57] Establishing that the digital item has not been altered once in the investigator's possession is important for confirming its authenticity.[58] This includes detailing how the digital item has been preserved and stored, such as whether the item was hashed and reliably time stamped at the point of collection. Opposing lawyers will want to ask several questions to determine whether the digital evidence has been properly preserved. For example, was the evidence collected in a manner that maintains its integrity using a collection tool that has been recognized as forensically-sound by a court or other practitioners? Lawyers will also want to assess whether the collection was complete. In other words, was enough accompanying information captured in conjunction with the digital item to understand the context that surrounded it?

The integrity of digital information may be difficult to establish but is especially challenging (and especially important) if investigators are unable to establish the author or provenance of the item, or critical links in the chain of custody. The Internet is rife with intentionally inauthentic information—including visual or audio disinformation that is misleadingly presented, edited, distorted, or computer-generated.[59] The Internet is also saturated with information that is unintentionally inauthentic—misinformation that includes everything from distorted imagery to misguided reporting.[60] The increased accessibility and rapidly falling costs of sophisticated photo and video-editing software has lowered the bar for amateurs to enter the digital manipulation game

56. *See generally Internet Archive*, WAYBACK MACHINE, https://archive.org/web/ (last visited July 1, 2022).

57. *See generally* Aida Ashouri, Caleb Bowers & Cherrie Warden, *The 2013 Salzburg Workshop on Cyber Investigations: An Overview of the Use of Digital Evidence in International Criminal Courts*, 11 DIGITAL EVIDENCE & ELEC. SIGNATURE L. REV. 115 (2014).

58. *See generally Record Integrity and Authenticity*, *in* BUILDING AN ELECTRONIC RECORDS ARCHIVE AT THE NATIONAL ARCHIVES AND RECORDS ADMINISTRATIONS: RECOMMENDATIONS FOR LONG-TERM STRATEGY (Robert F. Sproull & Jon Eisenberg eds., 2005) (discussing that an authentic record is one that is what it purports to be and that has been preserved without alteration).

59. Sophia Ignatidou, *Deepfakes, Shallowfakes and Speech Synthesis: Tackling Audiovisual Manipulation*, EUR. SCI.-MEDIA HUB (Dec. 4, 2019), https://sciencemediahub.eu/2019/12/04/deepfakes-shallowfakes-and-speech-synthesis-tackling-audiovisual-manipulation/.

60. Claire Wardle, *Understanding Information Disorder*, FIRST DRAFT NEWS (Sept. 22, 2020), https://firstdraftnews.org/long-form-article/understanding-information-disorder/.

and is aided by the fact that humans are notoriously bad at detecting image manipulations.[61]

In addition to authenticity, the proffering party will need to establish the information's reliability. Reliability is considered at different stages in different jurisdictions, sometimes as part of the admissibility assessment and sometimes as part of the weight assessment. In the absence of traditional indicia of reliability—such as a witness who can testify to the origins and accuracy of evidence—investigators must be creative about collecting enough contextual and corroborating information for a judge or jury to be comfortable relying on it. In the open source investigation context, reliability can be established through a three-pronged verification process that includes source analysis, content analysis and technical analysis.[62] According to the Scientific Working Group on Digital Evidence (SWGDE), image authentication involves an examination of visual information within an image (the content) and non-visual information about the image itself (technical aspects like the structure of pixels and any metadata).[63] When it comes to information found on the Internet rather than directly provided by a witness, authentication also involves an examination of the source of the image.[64] Source analysis differs from the attribution analysis mentioned above. Attribution analysis is the process of identifying the original source, if possible,[65] whereas source analysis occurs once the source is identified and needs to be evaluated. Thus, source analysis really means examining the *credibility* of the source.[66]

Content analysis refers to an analysis of the information contained within the "four corners"[67] of the digital item, whether it be an image, video, audio file, document, spreadsheet, social media post, or something else. In the United States, the most important factor for the court is usually whether the item is a fair and accurate portrayal of facts at issue in the case. Using a video pulled from social media as an example, analysis may include a review of the built or natural environment depicted in the video to determine if what is seen is consistent with the purported place and time; scrutiny of human features to determine if known perpetrators, victims, or witnesses are depicted; evidence of staging; the

61. Sophie J. Nightingale Kimberley Wade, Hany Farid & Derrick Watson, *Can People Detect Errors in Shadows and Reflections?*, 81 ATTENTION PERCEPTION & PSYCHOPHYSICS 2917, 2917 (2019). Seven experiments tested people's ability to use shadows to determine whether an image has been manipulated. Overall, detection rates were poor. *Id.*

62. BERKELEY PROTOCOL, *supra* note 11, at 62–65.

63. SCI. WORKING GRP. ON DIGIT. EVIDENCE, SWGDE BEST PRACTICES FOR IMAGE AUTHENTICATION 4 (2018), https://drive.google.com/file/d/1Z0DsJMa6aDZlFJ9kRfOL8cow5VjhVT0t/view.

64. *Id.* at 10.

65. BERKELEY PROTOCOL, *supra* note 11, at 63.

66. *Id.*

67. "Four corners" is a legal term used in contract law to refer to what is written in a document itself. *See* Legal Info. Inst., *Four Corners of an Instrument*, CORNELL L. SCH. (Sept. 2021), https://www.law.cornell.edu/wex/four_corners_of_an_instrument#. Here, it is used more broadly to encompass what is seen in an image or video.

photographic conditions, such as the quality of lighting, which may affect what can be perceived; and more.[68]

Technical analysis includes an examination of any metadata or Exif (Exchangeable Image File) data that may be attached to the item.[69] In order for forensic video analysts to interrogate imagery, they collect data on "image size, pixels, type of compression, frame rate, aspect ratio, GOP structure, file format, etc."[70] Relevant metadata may include the make, model, serial number, and settings of the device used to capture the image; the date and time of creation, as well as image resolution and size; any GPS coordinates or elevation data; information about frame rate, lens or flash; and thumbnails.[71] Lay, open source investigators often use free online tools like InVid or FotoForensics to review and extract the metadata.[72] However, this may differ from the commercial tools that are used by professional digital forensic analysts working for private law firms or prosecutors' offices.

One especially concerning consideration for digital videos is compression—a process that reduces and removes redundant information so that the digital video file can more easily be streamed over the Internet or transferred across a network.[73] Most digital videos that are emailed or uploaded online are subject to something called lossy compression, which means the loss of original information.[74] When there is a loss of data during compression, a computer later fills in the areas that were lost when compressed.[75] This results in artifacting—a noticeable distortion in the video's quality.[76] Such distortion can be a problem for lay investigators, who may not be aware of the ways in which such distortion can be spotted and how it can affect an analysis.

One form of distortion affects perception of movement due to variability in frame rates. The frame rate of a video is the number of frames in one second of video. There are standards for real time frame rate, which can vary.[77] When a video is played at that standard rate or higher, the human eye perceives motion

68. *See id.*

69. BERKELEY PROTOCOL, *supra* note 11, at 64–65.

70. *Interpreting Video Images: Can You "Say What You See"?*, JONATHAN W. HAK, Q.C.: BLOG (May 12, 2020), https://www.jonathanhak.com/2020/05/12/interpreting-video-images-can-you-say-what-you-see/.

71. SCI. WORKING GRP. ON DIGIT. EVIDENCE, *supra* note 62, at 7.

72. *See Invid Verification Plugin*, INVID, https://www.invid-project.eu/tools-and-services/invid-verification-plugin/ (last visited July 1, 2022); FOTOFORENSICS, http://fotoforensics.com/ (last visited July 1, 2022).

73. *Digital Video and Compression*, RGB SPECTRUM, https://www.rgb.com/digital-video-and-compression#:~:text=Video%20compression%20is%20a%20process,transmission%2C%20recording%2C%20or%20storage (last visited July 1, 2022).

74. *Id.*

75. *Id.*

76. *See, e.g.*, *Video Artifact*, TECHOPEDIA, https://www.techopedia.com/definition/31896/video-artifact (last visited July 1, 2022).

77. The real time frame rate for analog video is 30 frames per second under the North American standard, and 25 frames per second under the European standard. Levi Tijerina, *What is Frame Rate, and Why Does it Matter? (24fps vs. 30fps)*, GAMUT (Feb. 22, 2021), https://gamut.io/why-frame-rate-matters-24fps-vs-30fps.

accurately.[78] However, if the frame rate falls below the standard because of compression, the human brain may misinterpret the motion captured in the video.[79] Such human misperception can have disastrous consequences.[80] One example is a case from Florida in which a woman who worked as a nanny was accused of abusing an infant in her care and charged with eight counts of child abuse.[81] Key evidence was footage captured by a "nanny cam," a home surveillance system used to monitor and document caregivers' behavior, often without their knowledge.[82] At thirty frames per second, the video captured by the system is consistent with what the "human eye is capable of seeing, so the image appears fluid."[83] However, this rate is achieved when only one of the four cameras in the system is operating at a given time; with each additional camera the resolution drops, all the way down to 7.5 frames per second if all four cameras are operating—a rate far below what the human eye expects, and one that produces a "choppy" image.[84]

On the day the parents reviewed the nanny cam recording, their relatively choppy video appeared to show the nanny shaking the baby.[85] The parents took the baby to the emergency room and called the police. Although the baby had no visible injuries, the nanny was arrested for child abuse after a detective viewed the footage. Because the father allowed the nanny cam to keep running (police forgot to tell him to turn it off to preserve evidence) the original video was overwritten and all that was left of fourteen days' worth of recordings—which would have provided helpful context and data—was a two-hour long copy preserved by a law enforcement technician. The nanny spent two years in jail, refusing to take a plea deal, before the case was ultimately thrown out. While the parents continued to believe that the video was clear evidence of abuse, "prosecutors acknowledged that the video evidence was worthless" due to compression and framerate, which "could make [even] gentle motions appear violent."[86]

In addition to compression and framerate, video and image analysis can be tainted by a distortion in the aspect ratio, namely the relationship of width to height and the number of lines of information in the video.[87] Reliable comparison analysis cannot be done on a video with an incorrect aspect ratio;

78. *Id.*

79. *See id.*

80. *See generally* State v. Muro, 909 So.2d 448 (Fla. Dist. Ct. App. 2005), https://caselaw.findlaw.com/fl-district-court-of-appeal/1429591.html.

81. *Id.* at 449.

82. *Id.*

83. *Id.*

84. *Id.*

85. *Id.* at 450.

86. *Nanny Cleared of Violently Shaking Baby*, ABC NEWS (Mar. 21, 2006, 7:03 AM), https://abcnews.go.com/GMA/LegalCenter/story?id=1749672.

87. HANY FARID, FAKE PHOTOS 87 (2019).

therefore, the aspect ratio must be corrected before analysis.[88] Similarly, reliable comparison analysis cannot be done on a video or image that is too low quality. A video or image's quality—or resolution—might also interfere with analysis.[89] Low resolution results in a blurring effect. Resolution degrades with every copy, and most online videos and images are copies that further degrade when uploaded to social media platforms and again when downloaded by the investigator.[90] Unlike aspect ratio, poor image quality cannot be corrected.[91] Thus, while image comparison can provide compelling evidence in the courtroom, the strength of the comparison depends in part on the quality of the imagery. Insufficient quality imagery results in unreliable findings.[92]

Ultimately, digital open source investigators must properly verify the digital information on which they rely. False, forged, manipulated, degraded, and other problematic data can lead to erroneous findings. A significant volume of digital information, even when seemingly corroborative, cannot compensate for these underlying weaknesses. Once the digital evidence itself has been subjected to scrutiny, attorneys should look to the methods used to interpret, analyze, and draw conclusions about the raw information.

IV. CHALLENGING THE ANALYTICAL FINDINGS

In addition to challenging the imagery on which a conclusion was reached, defense lawyers may also challenge the method of analysis, particularly if it does not comply with a set standard in the relevant jurisdiction. Investigators and analysts should verify digital information before analyzing and reaching conclusions based on the information. Open source investigators' lack of validated analytical methods creates vulnerabilities, which opposing parties can attack on cross examination. At the same time, open source investigations' perceived newness may blind lawyers to the similarities between open source investigation methods like geolocation and well-established analytical methods like geospatial analysis. For example, an untrained open source investigator might compare two images to corroborate or disprove the location depicted in one of the images, failing to realize that there is an established discipline of

88. SCI. WORKING GRP. ON DIGIT. EVIDENCE, SWGDE TECHNICAL OVERVIEW FOR FORENSIC IMAGE COMPARISON 7 (2019), https://drive.google.com/file/d/11Pa7DOJmSJ00AieJcjNEGu7BJkCnvQkF/view; SCI. WORKING GRP. ON DIGIT. EVIDENCE, *supra* note 62, at 10.

89. SCI. WORKING GRP. ON DIGIT. EVIDENCE, *supra* note 62.

90. *Id*. at 8.

91. *Id*. at 10.

92. *See Best Practices for Forensic Image Analysis, Scientific Working Group on Imaging Technology,* FBI (Mar. 14, 2005), https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/oct2005/standards/2005_10_standards01.htm; *Webinar on Image Quality and Clarity: The Key to Forensic Digital Image Processing*, NAT'L INST. JUST. & FOREIGN TECH. CTR. OF EXCELLENCE (Aug. 3, 2021), https://nij.ojp.gov/events/image-quality-and-clarity-keys-forensic-digital-image-processing; *see also* 2 FORENSIC IMAGE COMPARISON AND INTERPRETATION EVIDENCE, *supra* note 45, at 7; Jonathan W. Hak, *Evaluation of the Forensic Science Regulator's Recommendations Regarding Image Comparison Evidence*, 1 FORENSIC SCI. INT'L: SYNERGY 294, 294 (2019).

forensic image comparison that has its own best practices. Similarly, open source investigators might try to confirm the location of a particular incident by comparing a video or photograph of the event to publicly available satellite imagery, while ignorant of the potential distortions, edits and other pitfalls of which geospatial analysts are well aware.

The scientific process, including the deployment of protocols, is a means of combating bias when examining and drawing conclusions from data.[93] With video comparison forensics, professional video analysts employ a methodology known as ACE-VR, an acronym for Analyze Compare Evaluate Verify and Report.[94] The process includes peer review, which is an essential prerequisite for scientific reliability.[95]

One of the most common open source investigation techniques is geolocation, which involves matching built and natural objects in videos and photographs to satellite imagery in order to determine the physical location where the images were shot.[96] The process of identifying visual clues in a video or photograph and matching those clues to known imagery is frequently used in professional investigations. While many claim this practice is new, it is not. What these investigators are doing falls within a well-defined subset of forensics: forensic image comparison, as discussed above. Forensic image comparison is defined as "an assessment of the correspondence between features in questioned items depicted in images and either questioned or known objects or images for the purpose of rendering an expert opinion regarding identification or elimination (as opposed to a demonstrative exhibit)."[97] Given this overlap, the analytical technique used by open source investigators should be considered and judged based on the standards that have already been established and vetted. When it comes to the ACE-VR method, according to SWGDE's best practices, "In order to accurately interpret the content of an image…it is imperative that the examiner recognize the conditions and limitations that occurred during image capture, processing or editing."[98] These conditions and limitations may include resolution, optical or sensor defects, lighting conditions, and motion

93. NAT'L CTR. FOR ST. CTS, HELPING COURTS AVOID IMPLICIT BIAS: STRATEGIES TO REDUCE THE INFLUENCE OF IMPLICIT BIAS 14–15, https://horsley.yale.edu/sites/default/files/files/IB_Strategies_033012.pdf (last visited July 1, 2022).

94. FORENSIC SCI. REGULATOR, CODES OF PRACTICE AND CONDUCT, APPENDIX: DIGITAL FORENSICS – VIDEO ANALYSIS 30–31 (2020), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/912390/FSR-C-119_Video_analysis_Issue_2.pdf; *Ace-V (Analysis, Comparison, Evaluation and Verification)*, EVISCAN, https://www.eviscan.com/en/ace-v-method-for-the-examination-anddocumentation-of-latent-fingerprints/ (last visited July 1, 2022).

95. *See generally* SHAUNA BRITTANI BREWER, ACE-V EXAMINATION METHOD TRAINING MANUAL (2014), https://scholarworks.calstate.edu/downloads/d791sg30j.

96. *See* Aric Toler, *"How to Verify and Authenticate User-Generated Content"*, *in* DIGITAL WITNESS: USING OPEN SOURCE INFORMATION FOR HUMAN RIGHTS INVESTIGATION, DOCUMENTATION AND ACCOUNTABILITY 198–216 (Sam Dubberley, Alexa Koenig & Daragh Murray, eds., 2020).

97. SCI. WORKING GRP. ON DIGIT. EVIDENCE, *supra* note 87, at 4.

98. *Id*. at 7.

and/or focal blur.[99] Therefore, defense attorneys should look at these factors and the quality of the image upon which the analysis was based in order to determine whether the analytical conclusions can be challenged on this basis.

V. CHALLENGING THE TESTIMONY

When it comes to testifying in court, there is the threshold issue of whether a digital open source investigator should be considered a lay witness or an expert witness. In most jurisdictions, witnesses can only testify to what they did or observed, while witnesses qualified as experts can provide opinions or conclusions.[100] For certain types of evidence, lawyers have different options for how to introduce such evidence at trial.[101] For example, graphology, otherwise known as hand-writing analysis, is an example of a comparative method that can be testified to by a lay witness or an expert.[102] Such evidence can be introduced in three ways: (1) a person who is familiar with the handwriting in question can testify to its validity; (2) a graphology expert can compare an unknown sample to a known sample; or (3) the two samples may be presented to the fact-finder (either judge or jury) to make their own determination about whether the two are likely from the same person.[103]

Comparisons between two digital open source images could likely be introduced to the fact-finder by expert witnesses with their conclusions or by lay witnesses leaving it to the fact-finders to draw their own conclusions. In most jurisdictions, case law or statutory rules establish who can be considered an expert in legal proceedings and thus provide opinions related to the evidence. In the United States, the standards were established by the *Frye* (1923)[104] and *Daubert* (1993)[105] cases and were codified in Federal Rule of Evidence 702. Per that rule:

> A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:
>
> (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;

---

99. *Id*.

100. Dani Alexis Ryskamp, *Lay Witness v. Expert Witness: What's the Difference?*, EXPERT INST. (June 25, 2020), https://www.expertinstitute.com/resources/insights/the-differences-between-expert-witness-and-lay-witness-testimony/.

101. Melissa Taylor, Ron Cowen, Katherine Fuller, Christina Frank, MacKenzie Robertson & Katherine Ritterhoff, Forensic Handwriting Examination and Human Factors: Improving the Practice Through a Systems Approach 79–108 (2020), https://nvlpubs.nist.gov/nistpubs/ir/2020/NIST.IR.8282.pdf.

102. An expert is a witness that is qualified to testify about a certain area of expertise, and who can provide conclusions about that area of expertise. A lay witness is a witness that is not testifying as an expert and therefore does not need to be qualified. A lay witness can only testify to their personal knowledge. *See Expert Witness*, CORNELL L. SCH.: LEGAL INFO. INST., https://www.law.cornell.edu/wex/expert_witness (last visited July 1, 2022).

103. *See generally* TAYLOR, *supra* note 100.

104. Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).

105. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 582 (1993).

> (b) the testimony is based on sufficient facts or data;
>
> (c) the testimony is the product of reliable principles and methods; and
>
> (d) the expert has reliably applied the principles and methods to the facts of the case.[106]

U.S. case law establishes that the burden is on the proffering party to establish the admissibility of expert testimony to a preponderance of the evidence.[107] With regard to the practices they might talk about, *Daubert* lays out a nonexclusive, non-dispositive checklist that courts can use when assessing reliability that includes whether the theory or technique has been tested or subjected to peer review and publication, and whether it has attracted "widespread acceptance within the relevant scientific community."[108]

Since digital open source investigators are often generalists, however, rather than masters of any particular forensic practice, a third issue concerns the potential scope of their expertise. Even if the investigator qualifies as an expert and can therefore provide opinions, it remains unclear as to what they can opine *on*. For example, a single video of the aftermath of an explosion might contain injured people and damaged buildings. A medical forensic expert could provide testimony with conclusions about the injuries in the video but could not speak to the structural damage to the buildings. A forensic architect could explain, for example, the likely cause of a building's damage, but could not opine about the cause of injuries to people. Rarely will one person qualify in all of the areas of expertise about which many open source investigators might draw conclusions. Thus, defense attorneys can challenge expert witnesses who step outside their areas of competence.

As established above, in U.S. Federal Rule 702, expert testimony can be provided by someone "who is qualified as an expert by knowledge, skill, experience, training, or education."[109] The Federal Rules also provide that an individual may testify if his or her "scientific, technical or other specialized knowledge will help the trier of fact to understand the evidence," the "testimony is based on sufficient facts or data," "the testimony is the product of reliable principles and methods" and "the expert has reliably applied the principles and methods to the facts of the case."[110] Interpretation of all of these prongs is relatively unsettled as applied to digital open source investigators and their methods. This leaves openings for attack—attacks that should be anticipated by the person testifying and others involved in the case's progression.

---

106. Testimony by Expert Witnesses, 28 U.S.C. 702 (2012).

107. *Id.*

108. Christine Funk, Daubert v. Frye*: A National Look at Expert Evidentiary Standards*, EXPERT INST. (Aug. 9, 2021), https://www.expertinstitute.com/resources/insights/daubert-versus-frye-a-national-look-at-expert-evidentiary-standards/#:~:text=Under%20Frye%2C%20the%20scientific%20community,courts%20consider%20the%20issue%20once.

109. 28 U.S.C. 702 (2011).

110. *Id.*

VI. CHALLENGING THE PRESENTATION

Finally, how digital open source evidence is presented in court, for example, via data visualizations that are packaged as demonstrative evidence, may raise serious concerns.[111] Demonstrative evidence—visual information that is aggregated and designed to aid the fact-finder(s) in understanding the geographic layout or other features of a location, such as a 3D digital reconstruction of a crime scene—is not evidence *per se*, but a visual aid designed to assist the judge or jury in their assessment of the actual evidence.[112] Such demonstrative evidence could, for example, consist of charts, maps, and graphs to help the fact-finder better understand multiple data points. In the open source investigation context, they are often compilations of many types of data, ranging from videos to photographs to satellite imagery, and may include 3D reconstructions of crime scenes, including built and natural layouts.[113]

Since this "aid to visually link evidence" to the underlying facts of a case is merely supposed to assist a court's understanding of the evidence, it is not supposed to be treated as evidence itself.[114] However, there's a risk that such visual aids may be overly compelling, resulting in prejudice about how the facts of the case come together that even judges cannot undo—or of which they may be unaware.[115] This can be particularly damaging if, for example, a digital reconstruction or other compilation is based on faulty underlying information, or digital items that are poorly interpreted or constructed.

While the underlying data—the individual videos and photographs that *are* evidence—must be introduced separately, the compilation may present and yet simultaneously obscure the underlying data. Not all data is equally reliable, so the individual data that comprise the whole might mislead, as might the compilation. This may become especially acute as new forms of reconstruction are introduced into courtrooms, such as virtual reality-based reconstructions that immerse witnesses, lawyers, or judges at the scene of a particular incident or

111. *See generally* Sarah Zarmsky, *Why Seeing Should Not Always Be Believing: Considerations Regarding the Use of Digital Reconstruction Technology in International Law*, 19 J. INT'L CRIM. L. 213 (2021) (discussing the strengths and weaknesses of introducing demonstrative evidence in international courtrooms). See *SITU Research Merges Data and Design to Create New Pathways of Justice,* SITU RSCH. https://situ.nyc/research, (last visited July 1, 2022), as one example of such visualizations; see also *Investigations*, FORENSIC ARCHITECTURE, https://forensic-architecture.org/ (last visited July 1, 2022), as a second example of such visualizations; Koenig, *supra* note 10 for a brief history of the use of such visualizations in international criminal trials.

112. *See, e.g.*, *The Complex Case of Cerro de Pasco Explained through an Interactive Platform*, SOURCE INT'L, https://www.source-international.org/news/discover-the-case-of-cerro-de-pasco-through-the-new-platform (last visited July 1, 2022); *see also* Working Draft: Practitioner Guidelines on the Use of Digitally Derived Evidence in International Accountability Mechanisms 43–49 (Leiden U. 2021).

113. See, e.g., SITU Research Launches SPEA Project, SITU (June 12, 2015), http://www.situstudio.com/blog/category/human-rights/.

114. Working Draft, *supra* note 111, at 43.

115. Waltraud Baier, Jason Warnett, Mark Payne & Mark A. Williams, *Introducing 3D Printed Models as Demonstrative Evidence at Criminal Trials*, 63 J. FORENSIC SCI. 1298, 1302 (2018); Rachael M. Carew, Ruth M. Morgan & Carolyn Rando, *A Preliminary Investigation into the Accuracy of 3D Modeling and 3D Printing in Forensic* Anthropology Evidence Reconstruction, 64 J. FORENSIC SCI. 342, 342 (2018).

event.[116] Ultimately, the judge and jury may be unable to interrogate each part of a potentially persuasive or even prejudicial whole.

In addition, these reconstructions can be expensive, ultimately exacerbating equality of arms considerations between prosecution and defense, and magnifying disparities between wealthy and less well-resourced parties. Finally, the demonstrative evidence may be constructed in such a way that it not only represents underlying facts but tells a story: it might fill in gaps or unknowns to make a narrative of events flow better or enable the viewer to fill in gaps in their head that may take on a sense of being the truth.

CONCLUSION

At its core, the introduction of technical, scientific, or other expert evidence into legal cases raises several overarching concerns, including the difficulty of distinguishing experts from convincing frauds; confusion caused by disagreements between experts; a lack of judicial training in the underlying methods and those methods' vulnerabilities; and an exacerbation of equality of resources issues.[117] While many of the digital methodologies that feed into civil and criminal investigations can be promising additions to investigative toolkits, enthusiasm for these techniques should be tempered with healthy skepticism—and knowledge of the most helpful questions to ask about any digital open source investigative process.

This task is not easy, nor straightforward. The complexity of the online information environment—riddled with botnets, sock puppets, trolls, deepfakes and shallowfakes, and operating at an almost unfathomable speed and scale—is simultaneously overwhelming and yet invaluable for identifying information that can contribute to justice and accountability. Our goal with this article has been to break down the very real vulnerabilities in digital open source investigations and encourage careful analysis of each component in order to make the risks more manageable.

Given the issues detailed above, we recommend that digital investigators preserve content according to emerging forensic standards and carefully document their investigative process. Witnesses should refrain from giving opinions on matters to which they do not have the proper expertise, while lawyers and judges need to be equipped to adequately ascertain the reliability and validity of digital open source information and the quality of its analysis. In

116. For more on virtual reality advances and pioneering uses of virtual reality in legal processes, see, e.g., *Methodology → Virtual Reality*, FORENSIC ARCHITECTURE, https://forensic-architecture.org/methodology/virtualreality (last visited July 1, 2022). Groups like Forensic Architecture are increasingly using virtual reality to aid witnesses' recollections of events and to help judges understand the layout of sites that are relevant to cases.

117. In addition to the resource differentials, junk science has often been found to benefit prosecutors or plaintiffs more than the defense. *See generally* Aviva A. Orenstein, *Debunked, Discredited, but Still Defended: Why Prosecutors Resist Challenges to Bad Science and Some Suggestions for Crafting Remedies for Wrongful Conviction Based on Changed Science*, 48 SETON HALL L. REV. 1139 (2018).

the meantime, digital open source information should be triangulated with physical, testimonial, or other documentary evidence whenever possible. If conducted carefully and professionally, digital open source investigations can offer tremendous value for both civil and criminal proceedings, internationally and at home.[118]

118. *See generally* BERKELEY PROTOCOL, *supra* note 11, at 12–13.