CORSO AVANZATO DI DIGITAL FORENSICS – WEB FORENSICS

Acquisizione forense di siti web, pagine web e servizi online

Metodo, quadro normativo e laboratorio con strumenti open source

Relatore: Vincenzo Calabrò

Programma e obiettivi della mezza giornata

Parte	Contenuto	Durata
1	La prova online: volatilità e requisiti probatori	30 min
2	Quadro normativo e standard	40 min
	Pausa	15 min
3	Metodo, strumenti e loro limiti	50 min
4	Laboratorio guidato con strumenti open source	70 min
5	Caso di studio, quiz di verifica, bibliografia	35 min

Quattro ore complessive. Il laboratorio richiede una virtual machine Linux per ogni coppia di studenti.

Al termine lo studente è in grado di

- distinguere ciò che un'acquisizione web dimostra da ciò che si limita a rendere verosimile;
- collegare ogni artefatto (packet capture, key log, video, timestamp) al requisito normativo che soddisfa;
- eseguire e documentare l'acquisizione di una pagina e di un intero sito con strumenti open source;
- esaminare un'acquisizione eseguita da altri e individuarne i punti contestabili.

PARTE 1 · 30 MINUTI

La prova online

Che cosa la distingue da un supporto sequestrato e che cosa si deve poter dimostrare davanti al giudice.

Supporto fisico e contenuto online a confronto

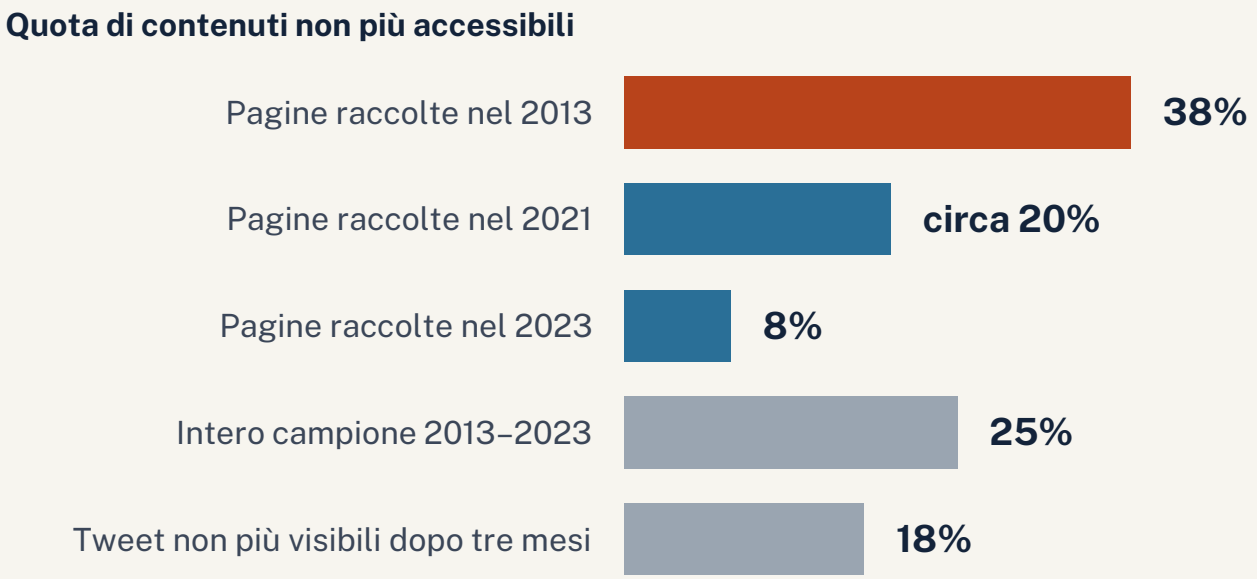
Aspetto	Acquisizione da supporto (dead-box)	Acquisizione di contenuto online
Accesso alla fonte	Il supporto è nella disponibilità di chi acquisisce	Il server è di terzi e si raggiunge soltanto attraverso la rete
Stato della fonte	Ferma: write blocker e copia bit a bit	Viva: il contenuto può cambiare mentre lo si osserva
Ripetibilità	L'immagine si rifà e si confronta tramite hash	La stessa richiesta, ripetuta, può restituire un risultato diverso
Oggetto della copia	L'intero supporto	Ciò che il server invia a quel client, da quell'indirizzo, in quel momento
Formati	EWF, AFF, immagini raw	Nessuno standard condiviso: WARC e WACZ, pcap, pacchetti proprietari
Verifica successiva	Hash della copia contro l'originale	Coerenza interna fra gli artefatti: l'originale può non esistere più

Fonte: A. Cantelli-Forti, G. Longo, F. Lupia, E. Russo, Int. J. Inf. Secur. 24:81 (2025).

La volatilità dei contenuti online

38%

delle pagine web esistenti nel 2013 non era più raggiungibile a ottobre 2023.

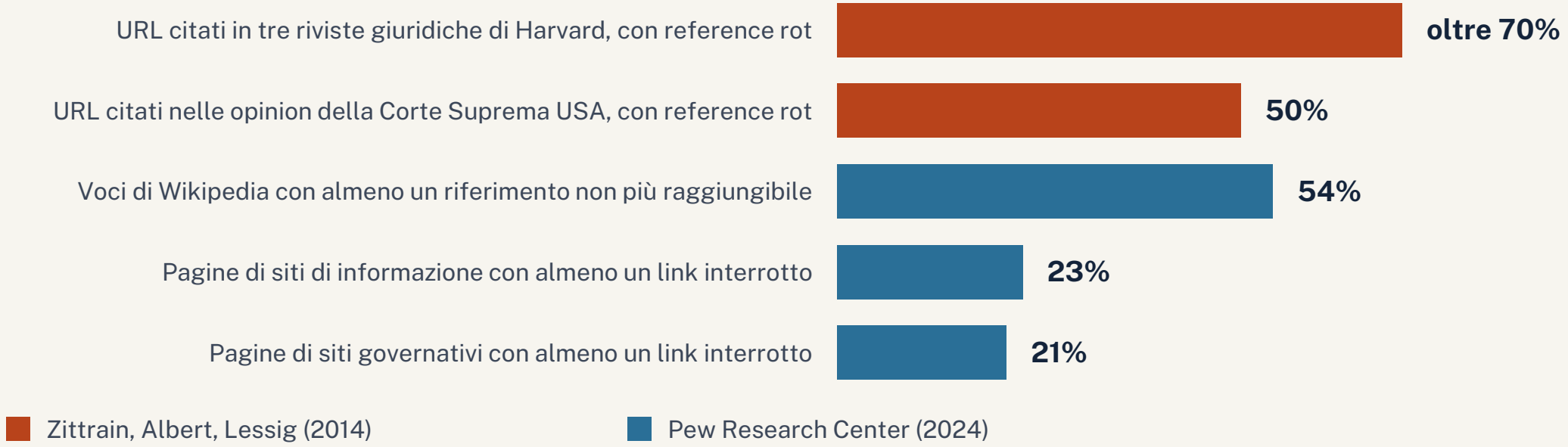


Conseguenza operativa. L'acquisizione precede ogni altra attività sul contenuto, compresa la diffida alla rimozione: una volta rimosso, resta soltanto ciò che è stato cristallizzato.

Fonte: Pew Research Center, When Online Content Disappears, 17 maggio 2024 (circa un milione di pagine da Common Crawl; quasi cinque milioni di tweet).

Link rot e reference rot nelle fonti giuridiche

Link rot: l'URL non restituisce più alcun contenuto. **Reference rot**: l'URL risponde ancora, ma l'informazione citata è cambiata o non c'è più.



Se nemmeno le citazioni di una corte suprema restano stabili, l'URL indicato in un atto di parte non può valere come rinvio durevole al contenuto.

Fonti: J. Zittrain, K. Albert, L. Lessig, 127 Harv. L. Rev. F. 176 (2014); Pew Research Center, When Online Content Disappears (2024).

Che cosa deve poter dimostrare un’acquisizione

Che cosa

Quale contenuto era visibile

Risorse HTTP salvate, pagina renderizzata dal browser, registrazione video dello schermo.

Da dove

Da quale server proveniva

Traffico di rete con chiavi TLS, risoluzione DNS, traceroute, catena dei certificati, whois.

Quando

In quale momento

Orologio sincronizzato via NTP, intestazioni Date dei server, marca temporale sul pacchetto finale.

Come

Con quale procedura, e che cosa è accaduto dopo

Log dei comandi, descrizione dell’ambiente, hash, chain of custody.

Ogni artefatto copre una parte del quadro. Il valore dell’acquisizione dipende da quanto gli artefatti concordano fra loro.

Elaborazione su ISO/IEC 27037:2012.

I limiti dello screenshot

Sul piano tecnico

- È un'immagine: non conserva il codice della pagina né il traffico che l'ha prodotta.
- Il DOM si modifica in pochi secondi con gli strumenti per sviluppatori del browser, e nell'immagine non ne resta traccia.
- Data e ora sono quelle dell'orologio locale, che l'operatore controlla.
- Non separa ciò che il sito ha servito da ciò che estensioni, proxy o cache hanno aggiunto.

Sul piano giuridico

- **Cass. civ., sez. lav., 16 febbraio 2004, n. 2912**: la copia di una pagina web ha valore probatorio solo se raccolta con garanzie di rispondenza all'originale e di riferibilità a un momento individuato.
- **Cass. pen., sez. V, n. 8736/2018**: lo screenshot è un documento ai sensi dell'art. 234 c.p.p. e si acquisisce anche senza attestazione di conformità.
- **Art. 2712 c.c.**: la riproduzione informatica fa piena prova se la parte contro cui è prodotta non la disconosce.

Lo screenshot è ammissibile. Quanto pesa dipende da ciò che la controparte sceglie di non contestare.

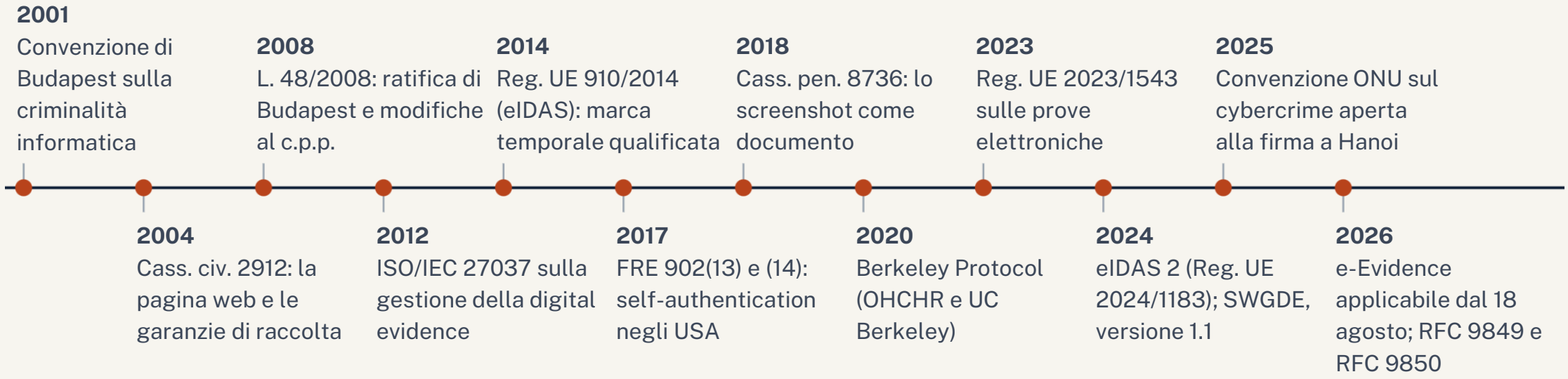
Cass. 2912/2004 (Rv. 570140-01); Cass. pen., sez. V, ud. 16 gennaio 2018, dep. 22 febbraio 2018, n. 8736.

PARTE 2 · 40 MINUTI

Quadro normativo e standard

Le norme fissano il risultato da garantire e lasciano al tecnico la scelta degli strumenti. Gli standard indicano come renderne conto.

Norme, standard e sentenze dal 2001 al 2026



Quattro tappe su dodici cadono negli ultimi quattro anni: dal 2023 il quadro si muove più in fretta che nel ventennio precedente.

I riferimenti completi sono nelle slide di bibliografia.

Italia: il processo penale dopo la L. 48/2008

La legge 18 marzo 2008, n. 48 ratifica la Convenzione di Budapest e riscrive, per i dati informatici, le norme sui mezzi di ricerca della prova.

Attività	Norma del c.p.p.
Ispezione	art. 244, co. 2
Perquisizione	artt. 247, co. 1-bis, e 352, co. 1-bis
Sequestro presso i fornitori di servizi	art. 254-bis
Accertamenti urgenti della polizia giudiziaria	art. 354, co. 2

Per Cass. pen. n. 8736/2018 l'estrazione di dati non è un accertamento tecnico irripetibile: della genuinità si discute nel merito.

Conservazione

«misure tecniche dirette ad assicurare la conservazione dei dati originali e ad impedirne l'alterazione»

Conformità

copia eseguita «mediante una procedura che assicuri la conformità della copia all'originale e la sua immodificabilità»

Il legislatore indica due esiti e non nomina strumenti: nessuna norma impone un software o un algoritmo di hash. La scelta, e l'onere di motivarla, restano a chi acquisisce.

Fonti: L. 18 marzo 2008, n. 48, artt. 8 e 9; artt. 244, 247, 254-bis, 352 e 354 c.p.p.

Italia: giurisprudenza, processo civile e prassi notarile

Fonte	Principio	Ricaduta per chi acquisisce
Cass. civ., sez. lav., n. 2912/2004	Le informazioni tratte dalla rete sono volatili: la copia vale se raccolta con garanzie di rispondenza all'originale e di riferibilità temporale	Servono un riferimento temporale e una prova di corrispondenza che non poggino sulla sola parola dell'operatore
Cass. pen., sez. III, n. 48178/2017	I fotogrammi scaricati da Google Earth sono prove documentali utilizzabili ai sensi dell'art. 234 c.p.p.	Il contenuto di un servizio online entra nel processo come documento
Cass. pen., sez. V, n. 8736/2018	Lo screenshot è un documento; non occorre un'attestazione di conformità	L'ammissione è agevole; il confronto si sposta sull'attendibilità
Art. 2712 c.c.; art. 20, co. 1-bis, CAD	La riproduzione informatica fa piena prova se non disconosciuta; il valore del documento informatico si valuta in base a sicurezza, integrità e immodificabilità	Una documentazione tecnica completa restringe lo spazio del disconoscimento
Consiglio Nazionale del Notariato, Studio 7-2007/IG	La copia conforme di una pagina web indica indirizzo, protocollo e browser utilizzato	Certifica ciò che il pubblico ufficiale vede a schermo; non documenta il traffico di rete

Unione europea: eIDAS e prove elettroniche

Reg. (UE) 910/2014

eIDAS, artt. 41 e 42

A una marca temporale elettronica non si negano effetti giuridici perché elettronica o non qualificata.

La marca temporale qualificata gode della presunzione di accuratezza di data e ora e di integrità dei dati associati, ed è riconosciuta in tutti gli Stati membri.

Reg. (UE) 2024/1183

eIDAS 2: nuovi servizi fiduciari

Archiviazione elettronica e registri elettronici (electronic ledgers) diventano servizi fiduciari.

I record di un registro qualificato godono della presunzione di ordinamento cronologico univoco e accurato e di integrità. Standard nei regolamenti di esecuzione 2025/1929, 2025/2531 e 2025/2532.

Reg. (UE) 2023/1543

e-Evidence, dal 18 agosto 2026

Ordine europeo di produzione (dieci giorni, otto ore in emergenza) e di conservazione (sessanta giorni, prorogabili di trenta), rivolti al service provider.

Riguarda i dati detenuti dal provider. L'acquisizione di contenuti pubblicamente accessibili resta fuori dal suo perimetro.

OpenTimestamps è una marca temporale elettronica non qualificata: si produce in giudizio, ma senza presunzione. Affiancarle una marca temporale qualificata (RFC 3161) toglie la data dal terreno della contestazione.

Fonti: Reg. (UE) 910/2014 e 2024/1183; Reg. (UE) 2023/1543; Commissione europea, DG HOME, pagina e-evidence.

Il piano internazionale: convenzioni e diritto statunitense

Convenzioni

- **Convenzione di Budapest (2001).** Artt. 16 e 17: conservazione rapida dei dati. Art. 32, lett. a): accesso transfrontaliero ai dati pubblicamente disponibili senza l'autorizzazione dello Stato in cui si trovano.
- **Convenzione ONU contro la criminalità informatica.** Adottata dall'Assemblea generale il 24 dicembre 2024 e aperta alla firma a Hanoi il 25 ottobre 2025. Entra in vigore novanta giorni dopo la quarantesima ratifica.

Stati Uniti

- **Federal Rules of Evidence 902(13) e 902(14)**, dal 1° dicembre 2017: i record prodotti da un processo elettronico e le copie di dati si autenticano con la certificazione di una persona qualificata. Per le copie il riscontro passa di regola dall'hash value.
- **Wayback Machine.** Ammessa in United States v. Bansal (3d Cir. 2011) e United States v. Gasperini (2d Cir. 2018) con la testimonianza di personale dell'Internet Archive. In Weinhoffer v. Davie Shoring (5th Cir. 2022) la corte ha escluso il judicial notice di una pagina archiviata non autenticata.

Il tratto comune: l'archivio di un terzo entra nel processo quando qualcuno ne spiega il funzionamento e ne risponde.

Fonti: Convenzione del Consiglio d'Europa ETS n. 185; UNODC; Fed. R. Evid. 902; sentenze citate.

ISO/IEC 27037:2012 applicata al web

1. Identification → 2. Collection → 3. Acquisition → 4. Preservation

Nell’acquisizione web la fase di collection manca: non c’è un supporto da prelevare, e identification e acquisition si svolgono nella stessa sessione.

Auditability

Un terzo deve poter ripercorrere ogni passo. Sul web: log dei terminali, video dello schermo, verbale.

Repeatability

Stessi strumenti e stesse condizioni, stesso risultato. Sul web vale per l’analisi del pacchetto, non per una nuova visita alla fonte.

Reproducibility

Strumenti diversi, stesso risultato. pcap e key log si rileggono con qualunque analizzatore.

Justifiability

Ogni scelta va motivata: quel browser, quello user agent, quel punto di uscita sulla rete.

La fonte non si può congelare. Ciò che manca in ripetibilità si recupera con la ricchezza della documentazione.

Fonte: ISO/IEC 27037:2012, Guidelines for identification, collection, acquisition and preservation of digital evidence, §§ 5.3 e 5.4.

Berkeley Protocol e SWGDE

OHCHR e UC Berkeley, 2020

Berkeley Protocol on Digital Open Source Investigations

Fissa gli standard minimi per identificare, raccogliere, conservare, verificare e analizzare informazioni open source nelle indagini su crimini internazionali e violazioni dei diritti umani.

Principi professionali: accountability, competency, objectivity, legality, security awareness.

Principi metodologici: accuracy, data minimization, preservation, security by design.

Principi etici: dignity, humility, inclusivity, independence, transparency.

SWGDE 21-F-001, versione 1.1, 2024

Best Practices for Acquiring Online Content

Riguarda i contenuti rivolti al pubblico: siti, servizi di streaming, piattaforme di comunicazione. Restano fuori le ricerche su mandato e l'acquisizione presso i cloud provider, cui è dedicato un documento a parte.

È indipendente dagli strumenti e distingue tre categorie di sito.

Static

Lo stesso contenuto per ogni visitatore.

Dynamic

Generato a richiesta: dipende da sessione, posizione, script.

Ephemeral

Destinato a sparire: storie, dirette, messaggi a tempo.

Fonti: OHCHR, HR/PUB/20/2 (2020); SWGDE 21-F-001-1.1 (15 marzo 2024). Le tre definizioni sono una sintesi didattica.

Dal requisito all’artefatto

Requisito	Fonte	Artefatto tecnico	Limite da conoscere
Integrità della copia	Art. 354, co. 2, c.p.p.; ISO/IEC 27037	Hash SHA-256 di ogni file e del pacchetto finale	Dimostra che nulla è cambiato dopo, non che il contenuto fosse genuino prima
Riferimento temporale	Cass. 2912/2004; art. 41 eIDAS	NTP, marca temporale qualificata (RFC 3161), OpenTimestamps	La marca fissa il «non dopo»; il «non prima» va ancorato in altro modo
Provenienza	Cass. 2912/2004, rispondenza all’originale	pcap con chiavi TLS, DNS, traceroute, certificati	TLS non garantisce la non-repudiation
Documentazione	ISO/IEC 27037, auditability; Berkeley Protocol	Log dei terminali, video dello schermo, verbale	Si produce durante l’acquisizione; ricostruita dopo perde valore
Verificabilità da terzi	ISO/IEC 27037, reproducibility; FRE 902(14)	Formati aperti, virtual machine conservata	Presuppone che la controparte riceva il pacchetto completo

Letta per righe, la tabella è già l’indice di un verbale di acquisizione.

Tabella di sintesi elaborata per il corso sulle fonti delle slide precedenti.

PARTE 3 · 50 MINUTI

Metodo, strumenti e limiti

Dalla pratica operativa alla ricerca più recente: che cosa registrare, in quali formati, e fin dove arriva ciò che ogni strumento riesce a provare.

Il metodo: documentare e incrociare

Rendere un’acquisizione web difendibile in giudizio quanto l’immagine forense di un disco.

Per i dischi esistono EWF e AFF, per i dispositivi mobili i formati dei rispettivi strumenti. Per una visita a una pagina il pacchetto va costruito: una virtual machine dedicata in cui ogni passaggio viene registrato da più fonti indipendenti.

La VM, chiusa e archiviata, fa parte del pacchetto.

Che cosa si registra, dall’esterno verso il sigillo

1. Video dello schermo	Ciò che l’operatore ha visto e fatto
2. Traffico e chiavi TLS	Ciò che client e server si sono scambiati
3. DNS, traceroute, whois	Dove si trovava il server e a chi faceva capo
4. NTP e ancore esterne	Quando: orologio sincronizzato, siti di informazione visitati in apertura
5. Log dei comandi	Come: ogni terminale sotto script, il browser sotto strace
6. Hash e timestamp	Il sigillo sul pacchetto e sulla VM compressa

Due scenari di acquisizione

Manuale, via browser

Una persona naviga, l’ambiente registra

Adatta a pagine costruite nel browser, streaming, contenuti che compaiono solo dopo un’interazione, profili social.

Limite: non scala a un sito intero e dipende dalla disciplina dell’operatore.

Automatica, via crawler

Un programma segue i link e salva le risposte

Adatta a siti interi, anche estesi, con un solo comando.

Limite: ciò che il crawler riceve può differire da ciò che vede un utente. Le pagine rese via JavaScript richiedono un crawler che piloti un browser vero.

La scelta discende da ciò che va dimostrato: l’aspetto di una pagina per chi la visita, oppure l’estensione di un contenuto. Nei casi seri servono entrambe.

Strumenti locali: lo stato a novembre 2026

Strumento	Licenza	Che cosa produce	Stato
wget, HTTrack	Open source	Copia dei file del sito	Disponibili ovunque; non nascono come strumenti forensi
ArchiveWeb.page, pywb	Open source	WARC e WACZ da navigazione manuale; replay	Progetti Webrecorder, alla base del formato WACZ
Browsertrix Crawler	AGPL-3.0	Crawl con Chromium pilotato da Puppeteer; WARC e WACZ	Ramo 1.x in sviluppo; versione 1.14 a luglio 2026
FIT, Freezing Internet Tool	MIT	Pagina singola con video, pcap e firma	Nato come tesi di master; termine di confronto nello studio WEFT
MAGNET Web Page Saver	Gratuito, Windows	Snapshot a scorrimento con hash	Ultima versione 3.3, settembre 2020
OSIRT	Gratuito	Diario della navigazione	Strumento da OSINT più che da forensics
FAW, Hunchly, X1 Social Discovery, Paliscope, WebPreserver	Commerciali	Pacchetti proprietari con report	Verificare presso il produttore formato del pacchetto e possibilità di riesame da parte di terzi

Fonti: pagine dei progetti e dei produttori, consultate a novembre 2026.

Servizi online e il caso Archive.today

Gratuiti

Internet Archive (Wayback Machine), Archive.today, Conifer, Perma.cc, Legalizer.

Commerciali

LegalEye, Cliens Prova Digitale, CRIO Solutions, Kopjra Web Acquisition, TrueScreen, Web Freeze di eWitness, VeriFact.

I servizi commerciali affiancano di norma all’acquisizione un report e una marca temporale; cambiano il formato del pacchetto e ciò che un terzo può riesaminare.

FEBBRAIO 2026

La Wikipedia in lingua inglese depreca Archive.today

La comunità decide di rimuovere circa 695.000 link: il servizio aveva sfruttato i browser dei visitatori per un attacco DDoS contro un blog e, secondo quanto emerso nella discussione, aveva alterato il contenuto di alcune pagine archiviate.

Chi affida la prova a un servizio terzo ne eredita l’affidabilità.

Fonti: Ars Technica, 10 e 20 febbraio 2026; TechCrunch, 21 febbraio 2026; Wikipedia, Requests for comment/Archive.is RFC 5.

Che cosa copre ciascun metodo

	Contenuto	Resa visiva	Provenienza di rete	Data opponibile	Riesame di terzi
Screenshot	In parte	**Sì**	No	No	No
«Salva pagina» o wget	**Sì**	In parte	No	No	In parte
WARC o WACZ da browser	**Sì**	**Sì**	In parte	In parte	**Sì**
Web archive pubblico	**Sì**	In parte	No	In parte	In parte
Servizio commerciale con report	**Sì**	**Sì**	In parte	**Sì**	In parte
VM con pcap, key log, video e timestamp	**Sì**	**Sì**	**Sì**	**Sì**	**Sì**

Nessun metodo offre una prova crittografica della provenienza. Nella colonna corrispondente «sì» significa che il traffico è documentato per intero e che un terzo può riesaminarlo.

Valutazione qualitativa proposta per la discussione in aula; non deriva da una misurazione.

I web archive come fonte di prova

8 su 8

servizi di archiviazione esaminati risultano vulnerabili ad almeno una classe di attacco.

7 su 8

sono esposti ad attacchi «anacronistici», che cambiano l'aspetto di uno snapshot dopo la cattura.

Lerner, Kohno, Roesner (ACM CCS 2017)

Il modo in cui la Wayback Machine archivia e riproduce le pagine consente a chi controlla determinati domini di alterare ciò che l'utente vede di uno snapshot, senza violare l'archivio.

Kirchner, Tsoukaladelis, Johns, Nikiforakis (ACM CCS 2025)

Il gestore di un sito può riconoscere i crawler degli archivi da indirizzi IP, ASN e fingerprint TLS e servire loro un contenuto diverso. Può anche predisporre riferimenti che, popolati in seguito, modificano snapshot già acquisiti.

In pratica. Lo snapshot di un terzo vale come riscontro. Finché la pagina è online la si riacquisisce nel proprio ambiente, con data e traffico documentati.

Servizi esaminati nel 2025: Wayback Machine, Archive.today, Perma.cc, Megalodon, Ghost Archive, Arquivo.pt, FreezePage, Conifer.

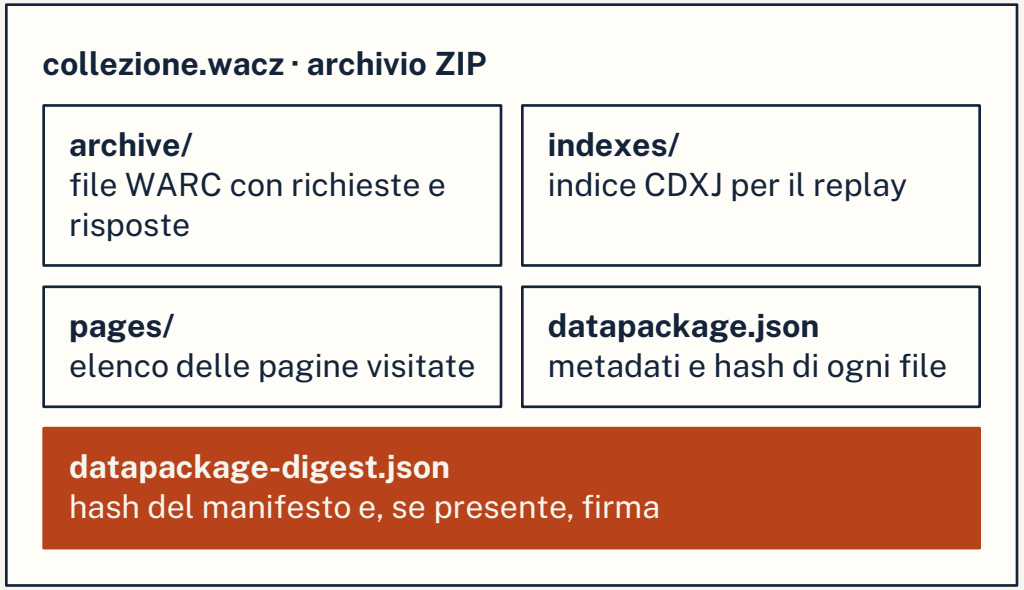
WARC e WACZ

WARC (ISO 28500:2017) è il formato contenitore degli archivi web: conserva richieste e risposte HTTP così come sono state scambiate, con i metadati della cattura.

WACZ (Webrecorder, 2021) riunisce in un unico ZIP i file WARC, gli indici, l'elenco delle pagine e un manifesto con gli hash. Si riapre con ReplayWeb.page senza un server.

WACZ Signing and Verification (versione 0.1.0) aggiunge la firma del pacchetto e la marca temporale di un terzo.

La stessa specifica avverte che, con gli standard web attuali, non si può dimostrare che un server abbia servito un dato URL in un dato momento.



Un WACZ documenta bene che cosa è stato scaricato e come si presentava. Su provenienza e data vale quanto la firma e la marca che porta con sé.

Fonti: ISO 28500:2017; Webrecorder, WACZ 1.1.1 e WACZ Signing and Verification 0.1.0; Library of Congress, FDD 000586.

UN LIMITE STRUTTURALE

TLS autentica il server davanti al client. Non consegna al client una prova da esibire a un terzo.

Le chiavi di sessione sono simmetriche e note a entrambe le parti: chi possiede il key log potrebbe, in linea di principio, costruire un traffico coerente.

Per questo pcap e chiavi valgono insieme a video, DNS, intestazioni dei server e marche temporali di terzi. Presi da soli non bastano.

Fonte: Webrecorder, WACZ Signing and Verification 0.1.0, sezione «Proof of Authenticity».

SSLKEYLOGFILE: rileggere in chiaro il traffico acquisito

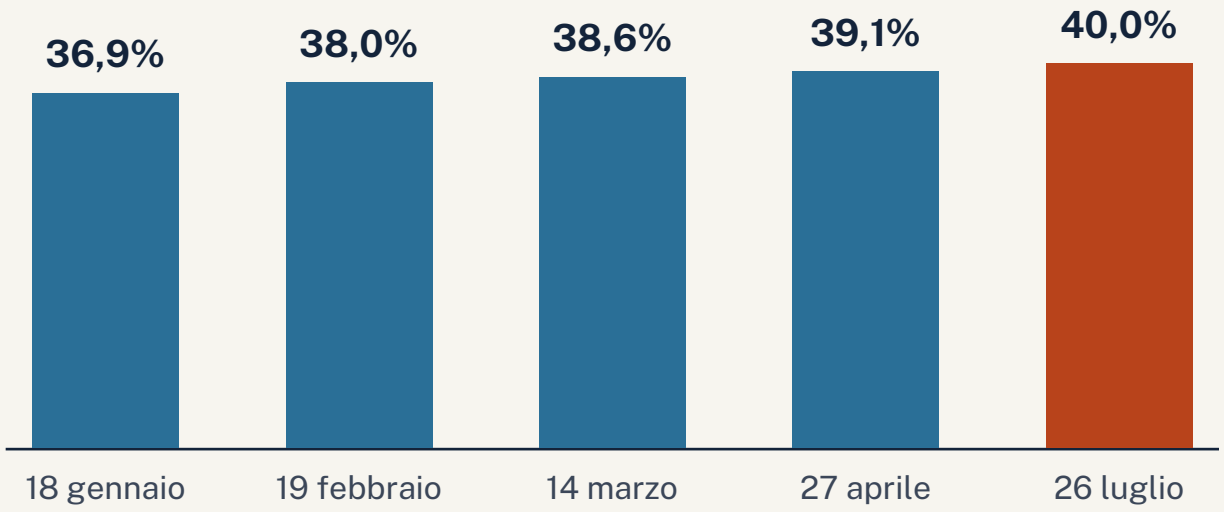


- **RFC 9850** (luglio 2026, informational) descrive il formato: secret per TLS 1.2 e 1.3, valido anche per QUIC, con etichette dedicate a Encrypted Client Hello.
- Lo stesso RFC vieta l'uso del meccanismo nei sistemi di produzione: una ragione in più per lavorare in una virtual machine dedicata.
- In Wireshark il file si indica in Preferences › Protocols › TLS, alla voce «(Pre)-Master-Secret log filename». I secret si possono anche incorporare nel pcapng.

Fonti: RFC 9850, The SSLKEYLOGFILE Format for TLS (2026); Wireshark Wiki, Transport Layer Security.

HTTPS e HTTP/3: che cosa finisce nella cattura

Siti web che usano HTTP/3, rilevazioni del 2026



Asse verticale da 0 a 50%.

Circa il 95% delle navigazioni in Chrome avviene su HTTPS (Google, ottobre 2025).

- Senza le chiavi di sessione il pcap di una visita documenta indirizzi e tempi, non contenuti.
- HTTP/3 viaggia su QUIC, quindi su UDP porta 443: un filtro di cattura limitato a TCP perde una quota crescente del traffico.
- Il key log copre anche QUIC e Wireshark lo usa allo stesso modo.

Fonti: W3Techs, Usage statistics of HTTP/3 for websites (2026); Google, «HTTPS by default», 28 ottobre 2025.

Il riferimento temporale: «non prima» e «non dopo»

Non prima di

Ancore esterne

- la home page di un sito di informazione, visitata in apertura;
- l'intestazione Date nelle risposte HTTP;
- l'istante dell'handshake TLS.

→

SESSIONE

Acquisizione

Video, pcap e log dei terminali condividono l'orologio della virtual machine, sincronizzato via NTP prima di cominciare. L'esito della sincronizzazione si salva nel fascicolo.

→

Non dopo

Marca temporale sul pacchetto

- marca qualificata RFC 3161 di un prestatore di servizi fiduciari;
- OpenTimestamps, con ancoraggio a un blocco Bitcoin.

Una sola marca apposta in chiusura certifica un istante e lascia scoperta la durata della sessione. WEFT ne appone una all'apertura e una alla chiusura.

Fonti: Cantelli-Forti et al. (2025); RFC 3161; RFC 5905.

OpenTimestamps

1. Pacchetto dell’acquisizione → 2. SHA-256 con nonce casuale → 3. Calendar server: Merkle tree → 4. Transazione Bitcoin → 5. Block header: l’attestazione

```
$ ots stamp acquisition.tar.gz # crea acquisition.tar.gz.ots, prova ancora incompleta
$ ots upgrade acquisition.tar.gz.ots # qualche ora dopo: aggiunge l’attestazione del blocco
$ ots verify acquisition.tar.gz.ots # controllo indipendente, con il file originale accanto
```

Che cosa prova

Che il file esisteva prima di quel blocco. Tace su chi lo ha creato e sulla verità del contenuto.

Riservatezza

Ai calendar server arriva soltanto un hash: il pacchetto non lascia la macchina.

Valore legale

Marca temporale non qualificata ai sensi di eIDAS. Si affianca a una marca qualificata, non la sostituisce.

Fonti: P. Todd, «OpenTimestamps: Scalable, Trust-Minimized, Distributed Timestamping with Bitcoin» (2016); opentimestamps.org.

Lo stato della ricerca: WEFT (2025)

Quattro limiti dei metodi correnti secondo gli autori

C1 · Evidence tampering
La chain of custody si firma alla fine: fino a quel momento gli artefatti restano modificabili.

C2 · Timestamp tampering
Una marca apposta in chiusura certifica un istante, non la sessione.

C3 · Single source of truth
File separati, senza una fonte che li leghi: allineare video e traffico è laborioso.

C4 · Self-verification
Manca una funzione che verifichi e metta in relazione gli artefatti.

LA PROPOSTA

Un unico file pcapng che contiene traffico, video, eventi di input e secret TLS, in blocchi concatenati con HMAC-SHA256. Marca RFC 3161 all'apertura e alla chiusura. Verifica automatica.

Prova su 100 siti fra i più visitati, a confronto con FIT



Valori medi per sito. Finalizzazione del pacchetto: sotto 1 secondo per WEFT, 25,8 secondi di mediana per FIT.

Fonte: A. Cantelli-Forti, G. Longo, F. Lupia, E. Russo, Int. J. Inf. Secur. 24:81 (2025), doi:10.1007/s10207-025-00991-8.

PARTE 4 · 70 MINUTI

Laboratorio guidato

L’acquisizione di una pagina e di un sito con strumenti open source, provati sulle distribuzioni Linux del 2026.

Obiettivo e regole del laboratorio

Acquisire una pagina pubblica indicata dal docente e consegnare un pacchetto che un’altra coppia riesca a verificare senza chiedere chiarimenti.

Si lavora in coppia

Operatore: esegue i comandi e naviga.

Verbalizzante: annota ora, comando ed esito di ogni passo; non tocca la tastiera.

Regola del video. Le quattro finestre restano sempre visibili. Una finestra coperta è un passaggio non documentato.

Prerequisiti

- virtual machine Linux (Debian, Ubuntu o Lubuntu) con sessione Xorg;
- snapshot della VM preso prima di cominciare;
- utente con sudo e rete senza proxy che intercetti TLS.

Fase	Tempo
Preparazione della VM e dei terminali	15 min
Acquisizione	25 min
Chiusura e sigillo	10 min
Verifica del pacchetto di un’altra coppia	20 min

Passo 0: preparare la virtual machine

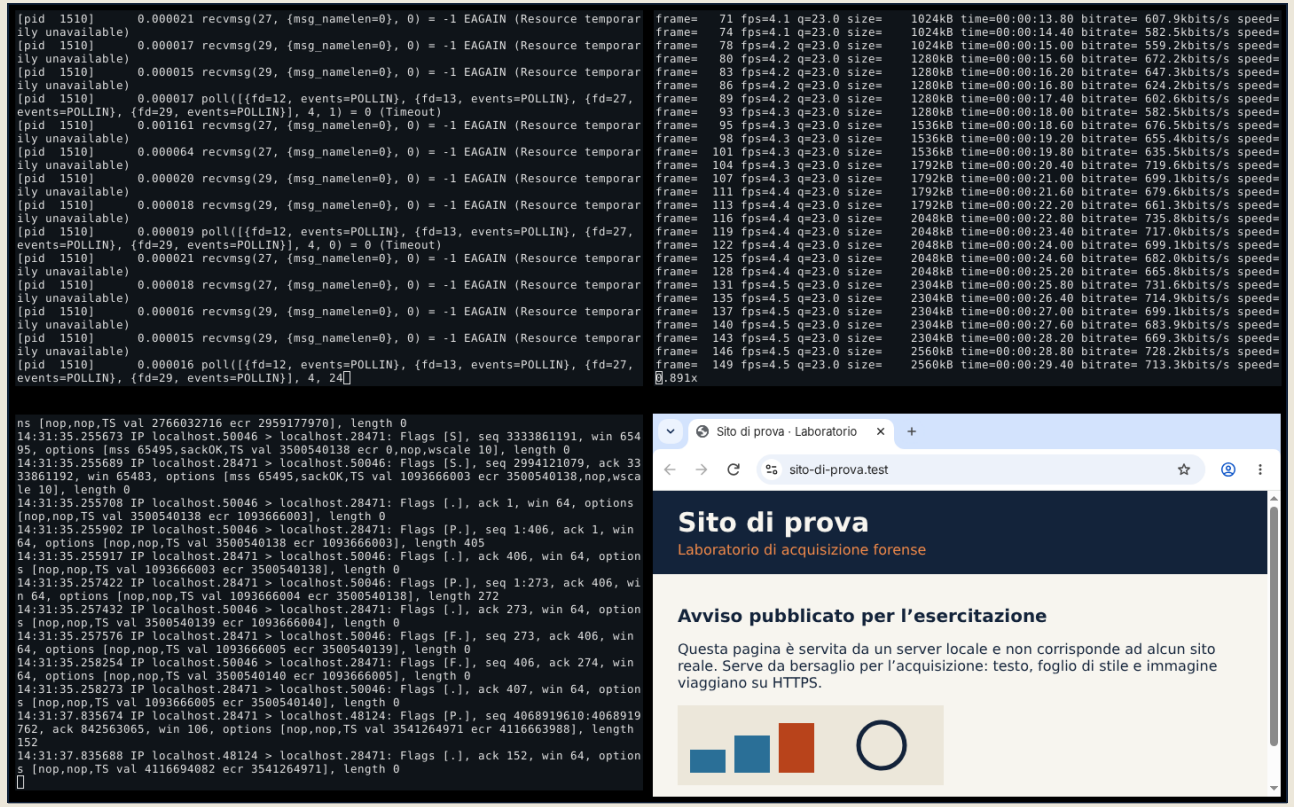
INSTALLAZIONE

```
$ sudo apt update
$ sudo apt install ffmpeg tcpdump strace traceroute dnsutils whois wireshark pipx
$ pipx install opentimestamps-client
$ ots --version && wget --version | head -4 && curl --version | head -1
# il browser: google-chrome-stable dal repository di Google, oppure chromium
```

- Sulle distribuzioni recenti pip rifiuta l'installazione di sistema (PEP 668): opentimestamps-client si installa con pipx.
- ffmpeg con x11grab richiede una sessione Xorg. Sotto Wayland lo schermo si registra con OBS Studio.
- ntpdate spesso non è preinstallato: se manca, si usa il servizio di sincronizzazione presente e se ne salva l'output.
- L'ultima riga del blocco fissa nel log le versioni degli strumenti.

Comandi provati a novembre 2026 su Ubuntu 24.04 LTS; opentimestamps-client 0.8.0 da PyPI.

Il desktop durante l’acquisizione



Schermata del laboratorio eseguito su un sito di prova locale, Ubuntu 24.04, novembre 2026.

In alto a sinistra · Terminale 1

Variabile SSLKEYLOGFILE, comandi di contesto, browser lanciato sotto strace.

In alto a destra · Terminale 2

ffmpeg registra lo schermo; il contatore dei frame scorre.

In basso a sinistra · Terminale 3

tcpdump salva il traffico e lo mostra pacchetto per pacchetto.

In basso a destra · Browser

La navigazione oggetto dell’acquisizione.

Passi 1–3: i tre terminali

TERMINALE 1 · IN ALTO A SINISTRA

```
$ mkdir ~/caso01 && cd ~/caso01
$ script w1.txt
$ export SSLKEYLOGFILE="$HOME/caso01/sslkeylogfile.log"
# script registra tutto ciò che passa dal terminale; la variabile va esportata qui, prima del browser
```

TERMINALE 2 · IN ALTO A DESTRA

```
$ cd ~/caso01 && script w2.txt
$ ffmpeg -f x11grab -y -r 5 -s 1440x900 -i :0.0 -c:v libx264 video.avi
# -s deve coincidere con la risoluzione dello schermo; 5 fotogrammi al secondo bastano
```

TERMINALE 3 · IN BASSO A SINISTRA

```
$ cd ~/caso01 && script w3.txt
$ sudo tcpdump -n -U -s 65535 -w - | tee traffic.pcap | tcpdump -r -
# -w - scrive su standard output, tee salva il file, il secondo tcpdump mostra i pacchetti a video
```

Passo 4: contesto di rete e orario

TERMINALE 1 · SOSTITUIRE EXAMPLE.ORG CON IL DOMINIO ASSEGNATO

```
$ sudo ntpdate pool.ntp.org | tee ntpdate.txt
$ cp /etc/hosts ./hosts
$ dig www.example.org | tee dig.txt
$ traceroute www.example.org | tee traceroute.txt
$ whois example.org | tee whois.txt
```

File	Che cosa documenta
ntpdate.txt	Lo scarto dell'orologio della VM rispetto a una fonte di tempo pubblica
hosts	Che il nome non è stato risolto da una voce locale
dig.txt	Gli indirizzi restituiti dal resolver in quel momento, da confrontare con quelli del pcap
traceroute.txt	Il percorso di rete verso il server
whois.txt	Intestatario e registrar del dominio alla data dell'acquisizione

Passo 5: il browser sotto strace

TERMINALE 1

```
$ strace -r -f google-chrome-stable 2>&1 | tee strace.txt
# -r tempi relativi fra le system call, -f segue i processi figli del browser
```

La navigazione

1. Portare il browser in basso a destra, senza coprire i terminali.
2. Aprire per prima una pagina con data e ora e la home di un sito di informazione.
3. Digitare l'URL per esteso: niente segnalibri, niente cronologia.
4. Scorrere fino in fondo, aprire i contenuti a comparsa, attendere le immagini.

Che cosa aggiunge strace

Nel log restano le letture, le scritture e le connessioni del browser: un riscontro indipendente dal pcap su quali risorse siano state davvero caricate.

Migliorie possibili

Screenshot a pagina intera con un'estensione, salvataggio della pagina, audio registrato con OBS, mitmproxy per rendere leggibili le richieste HTTP.

Passo 6: chiusura e sigillo

```
# chiudere il browser, poi Ctrl+C su tcpdump, q su ffmpeg, Ctrl+D in ciascun terminale
$ cd ~/caso01 && sha256sum * > ../caso01.sha256
$ cd .. && tar -czvf acquisition.tar.gz caso01 caso01.sha256
$ sha256sum acquisition.tar.gz | tee acquisition.sha256
$ ots stamp acquisition.tar.gz
```

Poi

La virtual machine

Si spegne la VM, si comprime l'intera cartella e si appone una marca temporale anche sul file compresso.

Accanto

La marca qualificata

Sullo stesso pacchetto, una marca RFC 3161 di un prestatore qualificato. In aula è facoltativa, in un incarico reale no.

Infine

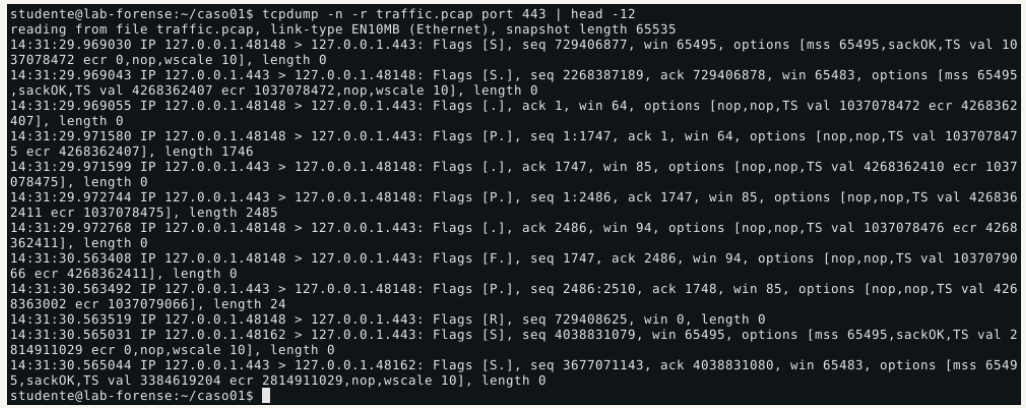
Il verbale

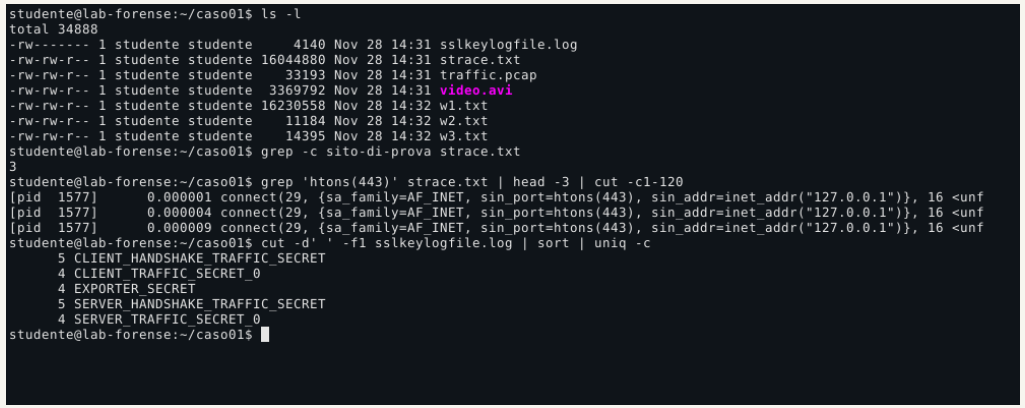
Ora di inizio e di fine, operatore, strumenti con le rispettive versioni, hash del pacchetto, anomalie osservate.

Sequenza provata a novembre 2026 su Ubuntu 24.04 LTS.

Passo 7: la verifica incrociata

Domanda	Dove guardare
La pagina viene dal server dichiarato?	Indirizzi in dig.txt contro la destinazione nel pcap; certificato presentato nell'handshake
Il video mostra ciò che è passato in rete?	Wireshark con il key log: richiesta e risposta della pagina, a confronto con i fotogrammi
I tempi coincidono?	ntpdate.txt, orari dei pacchetti, orologio visibile nel video
Il browser si è connesso a quel server?	Chiamate connect verso la porta 443 in strace.txt, a confronto con il pcap
Il pacchetto è integro e datato?	Da dentro caso01: sha256sum -c ../caso01.sha256; poi ots verify sul pacchetto





A sinistra il pcap riletto con tcpdump; a destra file del caso, connect in strace.txt, secret nel key log.

Variante: l’intero sito con un crawler

WGET · AL POSTO DEL BROWSER

```
$ strace -r -f wget -m --keep-session-cookies --save-cookies=cookies.txt --limit-rate=500k \
-w 1 --random-wait --no-check-certificate -e robots=off -o log.txt -vv -S \
--preserve-permissions -np -E -k -K -p --show-progress https://www.example.org 2>&1 | tee strace.txt
$ wget https://www.example.org/robots.txt -P robots --no-check-certificate -o wget_log_robots.txt
$ openssl s_client -connect www.example.org:443 -showcerts < /dev/null > ssl.txt
```

BROWSERTRIX CRAWLER · PER I SITI RESI VIA JAVASCRIPT

```
$ docker run -v $PWD/crawls:/crawls/ -it webrecorder/browsertrix-crawler crawl \
--url https://www.example.org/ --generateWACZ --collection caso01
```

- wget non salva né robots.txt né il certificato del server: si acquisiscono a parte.
- wget scrive le chiavi TLS solo se compilato con GnuTLS. Su Ubuntu 24.04 la versione 1.21.4 usa OpenSSL e ignora SSLKEYLOGFILE; curl 8.5.0 la rispetta.
- robots=off e --no-check-certificate sono scelte: vanno motivate nel verbale.

Che cosa è cambiato dal 2022

Prassi del 2022	A novembre 2026	Che cosa fare
PhantomJS per l’automazione	Sviluppo sospeso dal marzo 2018, repository archiviato	Playwright o Puppeteer; Browsertrix per i crawl
wget ricompilato con GnuTLS	Su Ubuntu 24.04 wget 1.21.4 ignora SSLKEYLOGFILE, curl 8.5.0 la rispetta	Controllare wget --version; usare curl o un crawler basato su browser
pip3 install opentimestamps-client	pip di sistema bloccato (PEP 668); client alla versione 0.8.0	pipx install opentimestamps-client
ffmpeg con x11grab	Sessione Wayland predefinita su molte distribuzioni	Accedere con Xorg oppure registrare con OBS Studio
Cattura con tcpdump	HTTP/3 su QUIC per circa il 40% dei siti	Nessun filtro limitato a TCP; il key log vale anche per QUIC
SSLKEYLOGFILE come convenzione	RFC 9850 (luglio 2026); Encrypted Client Hello in RFC 9849 (marzo 2026)	Citare gli RFC nel verbale
Archive.today fra i servizi gratuiti	Deprecato dalla Wikipedia in lingua inglese (febbraio 2026)	Solo come riscontro, mai come unica copia

Le righe su wget, curl e pip derivano da prove eseguite a novembre 2026 su Ubuntu 24.04 LTS; le altre dalle fonti in bibliografia.

Errori ricorrenti e contestazioni prevedibili

Errore	Che cosa ne farà il consulente della controparte
Orologio non sincronizzato, o sincronizzazione non salvata	Sosterrà che la data è quella dichiarata dall'operatore e nient'altro
Key log assente, o variabile esportata dopo l'avvio del browser	Farà notare che il pcap è illeggibile e che resta soltanto il video
Finestre sovrapposte durante la registrazione	Indicherà il minuto in cui un passaggio non è documentato
Profilo del browser già in uso, cookie o sessione autenticata	Obietterà che il contenuto era personalizzato e diverso da quello pubblico
Contenuti a caricamento differito non scorsi	Mostrerà che nel traffico mancano proprio le parti contestate
Hash calcolato dopo il trasferimento del pacchetto	Individuerà un tratto della chain of custody privo di copertura
Prova OpenTimestamps mai completata con upgrade	Rileverà che la marca dipende ancora dai calendar server
Paese di uscita sulla rete non dichiarato	Produrrà la stessa pagina vista da un altro Paese, diversa

Ciascuna di queste righe si previene con una voce in più nella lista di controllo del verbalizzante.

Limiti legali ed etici dell’acquisizione

Soltanto ciò che è pubblico

Si acquisisce ciò che è raggiungibile senza superare misure di sicurezza. Entrare in un’area riservata con credenziali altrui, o aggirando un controllo, integra l’accesso abusivo a sistema informatico (art. 615-ter c.p.).

Data minimization

Si acquisisce ciò che serve alla tesi da provare. Nel pcap finiscono anche dati di terzi e, se si è usato un account, le proprie credenziali in chiaro: il pacchetto va protetto e la sua circolazione limitata.

Scelte da dichiarare

Ignorare robots.txt, limitare la banda, cambiare user agent, uscire da una VPN o da Tor per vedere la pagina da un altro Paese: tutto legittimo, a condizione che il verbale lo riporti con la ragione.

Osservare, non interagire

Le operazioni sotto copertura sono riservate alla polizia giudiziaria nei casi previsti dalla legge (art. 9, l. 146/2006). Il consulente di parte si ferma all’osservazione di ciò che chiunque può vedere.

Fonti: art. 615-ter c.p.; l. 16 marzo 2006, n. 146, art. 9; Reg. (UE) 2016/679; Berkeley Protocol, principi di legality e data minimization.

PARTE 5 · 35 MINUTI

Caso di studio e verifica

Un incarico verosimile da impostare in gruppo, otto domande per controllare che cosa è rimasto, le fonti per approfondire.

Caso di studio: l’incarico

Scenario didattico di fantasia

Venerdì, ore 17:40. Uno studio legale assiste un’azienda alimentare del Nord-Est.

Da tre giorni circolano contenuti che l’azienda ritiene diffamatori. L’avvocato chiede di «bloccare tutto» prima di lunedì, quando partirà la diffida.

Vincoli. I contenuti possono sparire in qualunque momento. La controparte contesterà la data. Il compenso copre mezza giornata di lavoro.

Che cosa c’è da acquisire

1 · Un post pubblico su un social network

Una quarantina di commenti, alcuni compressi, e un video incorporato di due minuti.

2 · La scheda dell’azienda su un portale di recensioni

Le recensioni si caricano man mano che si scorre la pagina; l’ordine cambia a ogni visita.

3 · Un blog di una trentina di pagine

Dominio registrato da poche settimane; ripubblica gli stessi contenuti con immagini proprie.

Caso di studio: consegne e traccia di discussione

Ai gruppi, venti minuti

1. Per ciascun contenuto scegliere scenario e strumenti, motivando.
2. Elencare gli artefatti e l’ordine delle operazioni.
3. Scrivere l’indice del verbale di acquisizione.
4. Prevedere tre contestazioni e la risposta a ciascuna.

Contenuto	Traccia per la discussione
Post social	Browser nella VM, senza login se il post è pubblico. I commenti si espandono uno a uno; il video va riprodotto per intero, altrimenti non entra nel traffico.
Portale di recensioni	Browser con scorrimento completo. In aggiunta un crawl con Browsertrix in WACZ, utile per il replay in udienza.
Blog	Crawler. robots.txt, certificato e whois a parte: qui la data di registrazione del dominio è un fatto rilevante.
Per tutti	Si acquisisce prima della diffida. Marca qualificata accanto a OpenTimestamps. Uno snapshot su Internet Archive vale come riscontro.

Quiz di verifica: domande 1–4

1. Per Cass. civ. n. 2912/2004 la copia di una pagina web ha valore probatorio se

A. è stampata a colori e sottoscritta dalla parte

B. è raccolta con garanzie di rispondenza all'originale e di riferibilità a un momento individuato

C. è sempre autenticata da un notaio

D. è depositata entro trenta giorni dalla pubblicazione

2. Il file indicato dalla variabile SSLKEYLOGFILE contiene

A. la chiave privata del server

B. i certificati della catena di fiducia

C. i secret delle sessioni TLS negoziate dal client

D. le password salvate nel browser

3. Una prova OpenTimestamps completa dimostra che

A. il file è stato creato dall'operatore indicato nel verbale

B. il file esisteva prima di un determinato blocco Bitcoin

C. il contenuto del file corrisponde al vero

D. il file reca una firma elettronica qualificata

4. Da una cattura limitata al traffico TCP di un sito servito in HTTP/3 resta fuori

A. la risoluzione DNS

B. il traffico QUIC su UDP porta 443

C. l'handshake TLS 1.2

D. nulla, perché HTTP/3 usa TCP

Quiz di verifica: domande 5–8

5. Per l’art. 41 del regolamento eIDAS una marca temporale non qualificata

A. è priva di qualunque effetto giuridico

B. gode della presunzione di accuratezza di data e ora

C. non può essere respinta come prova per il solo fatto di non essere qualificata

D. vale soltanto nello Stato membro in cui è stata emessa

6. Lo studio di Kirchner e colleghi (ACM CCS 2025) mostra che

A. i web archive non sono manipolabili

B. soltanto Archive.today è vulnerabile

C. tutti gli otto servizi esaminati cedono ad almeno una classe di attacco

D. la Wayback Machine firma ogni snapshot

7. Fra i requisiti di ISO/IEC 27037, rispetto alla fonte un’acquisizione web fatica soprattutto a garantire

A. auditability

B. repeatability

C. justifiability

D. nessuno dei tre

8. Su Ubuntu 24.04, wget 1.21.4 avviato con SSLKEYLOGFILE impostata

A. scrive i secret come farebbe Chrome

B. ignora la variabile, perché è collegato a OpenSSL

C. si rifiuta di partire

D. salva le chiavi dentro il pcap

Quiz di verifica: soluzioni

N.	Risposta	Motivazione
1	B	La Corte richiede rispondenza all’originale e riferibilità temporale; il notaio è una delle vie possibili, non l’unica.
2	C	Sono i secret di sessione scritti dal client. La chiave privata del server non lascia mai il server.
3	B	La marca attesta l’esistenza del file prima di un dato momento. Tace sull’autore e sul contenuto.
4	B	HTTP/3 viaggia su QUIC, cioè su UDP. Un filtro limitato a TCP lo esclude.
5	C	Art. 41, par. 1. La presunzione del par. 2 spetta soltanto alla marca qualificata.
6	C	Otto servizi su otto; sette su otto risultano esposti anche agli attacchi anacronistici.
7	B	La fonte cambia: una nuova visita può restituire un contenuto diverso. La ripetibilità resta per l’analisi del pacchetto.
8	B	Riscontro di laboratorio di novembre 2026. Sulla stessa macchina curl 8.5.0 scrive i secret.

In sintesi

1. Prima si acquisisce, poi si analizza. Il contenuto online non aspetta la diffida.

2. Nessun artefatto regge da solo. Conta la concordanza fra video, traffico, contesto di rete e riferimenti temporali.

3. Le norme chiedono conformità e immodificabilità, e non indicano uno strumento. Ogni scelta tecnica si motiva e si mette a verbale.

4. Gli strumenti invecchiano più in fretta del metodo. I comandi si riprovano ogni volta sulla macchina che si userà.

Bibliografia: norme e standard

Consiglio d’Europa, Convenzione sulla criminalità informatica, Budapest, 23 novembre 2001 (ETS n. 185).

Legge 18 marzo 2008, n. 48, di ratifica della Convenzione di Budapest.

Codice di procedura penale, artt. 234, 244, 247, 254-bis, 352, 354. Codice civile, art. 2712. D.lgs. 7 marzo 2005, n. 82 (CAD), art. 20.

Regolamento (UE) n. 910/2014 (eIDAS), artt. 41 e 42; regolamento (UE) 2024/1183; regolamenti di esecuzione (UE) 2025/1929, 2025/2531 e 2025/2532.

Regolamento (UE) 2023/1543 e direttiva (UE) 2023/1544 sulle prove elettroniche.

Nazioni Unite, Convenzione contro la criminalità informatica, risoluzione 79/243 dell’Assemblea generale, 24 dicembre 2024.

Federal Rules of Evidence (USA), Rule 902(13) e 902(14), in vigore dal 1° dicembre 2017.

ISO/IEC 27037:2012, Guidelines for identification, collection, acquisition and preservation of digital evidence.

ISO 28500:2017, WARC file format.

OHCHR e Human Rights Center, UC Berkeley, Berkeley Protocol on Digital Open Source Investigations, HR/PUB/20/2, 2020.

SWGDE, Best Practices for Acquiring Online Content, 21-F-001-1.1, 15 marzo 2024.

IETF: RFC 3161 (Time-Stamp Protocol); RFC 5905 (NTPv4); RFC 9849 (TLS Encrypted Client Hello, 2026); RFC 9850 (The SSLKEYLOGFILE Format for TLS, 2026).

Webrecorder, WACZ 1.1.1 e WACZ Signing and Verification 0.1.0, specs.webrecorder.net.

Bibliografia: letteratura scientifica e tecnica

Cantelli-Forti A., Longo G., Lupia F., Russo E., «WEFT: a consistent and tamper-proof methodology for acquisition of automatically verifiable forensic web evidence», International Journal of Information Security, 24:81, 2025. doi:10.1007/s10207-025-00991-8

Kirchner R., Tsoukaladelis C., Johns M., Nikiforakis N., «The Power to Never Be Wrong: Evasions and Anachronistic Attacks Against Web Archives», ACM CCS 2025.

Lerner A., Kohno T., Roesner F., «Rewriting History: Changing the Archived Web from the Present», ACM CCS 2017. doi:10.1145/3133956.3134042

Zittrain J., Albert K., Lessig L., «Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations», 127 Harvard Law Review Forum 176, 2014.

Chapekis A., Bestvater S., Remy E., Rivero G., When Online Content Disappears, Pew Research Center, 17 maggio 2024.

Todd P., «OpenTimestamps: Scalable, Trust-Minimized, Distributed Timestamping with Bitcoin», 2016, petertodd.org.

Consiglio Nazionale del Notariato, Studio n. 7-2007/IG, «La copia conforme di una pagina web: natura giuridica e modalità operative».

Wireshark Wiki, «Transport Layer Security (TLS)»; Webrecorder, documentazione di Browsertrix Crawler.

Calabrò V., «Web Forensics: Acquisizione delle fonti di prova online»; 2022, vincenzocalabro.it.

Bibliografia: giurisprudenza, dati e immagini

Giurisprudenza

Cass. civ., sez. lav., 16 febbraio 2004, n. 2912 (Rv. 570140-01).

Cass. pen., sez. III, 15 settembre 2017, n. 48178 (Rv. 271313).

Cass. pen., sez. V, ud. 16 gennaio 2018, dep. 22 febbraio 2018, n. 8736.

United States v. Bansal, 663 F.3d 634 (3d Cir. 2011).

United States v. Gasperini, 894 F.3d 482 (2d Cir. 2018).

Weinhoffer v. Davie Shoring, Inc., No. 20-30568 (5th Cir., 20 gennaio 2022).

Dati e cronaca

W3Techs, Usage statistics of HTTP/3 for websites, rilevazioni da gennaio a luglio 2026.

Google, «HTTPS by default», 28 ottobre 2025.

Brodkin J., Ars Technica, 10 e 20 febbraio 2026; TechCrunch, 21 febbraio 2026 (Archive.today).

Immagini e prove di laboratorio

Le schermate alle slide 35 e 40 sono state prodotte eseguendo il laboratorio su un sito di prova locale. Diagrammi e grafici sono stati elaborati per il corso sui dati delle fonti citate.

Il laboratorio e i riscontri su wget, curl, pip e opentimestamps-client sono stati eseguiti a novembre 2026 su Ubuntu 24.04 LTS.

CORSO AVANZATO DI DIGITAL FORENSICS

Grazie per l'attenzione

Vincenzo Calabrò

www.vincenzocalabro.it



Inquadra per aprire il sito