Il documento digitale come fonte di prova: produzione, gestione e conservazione

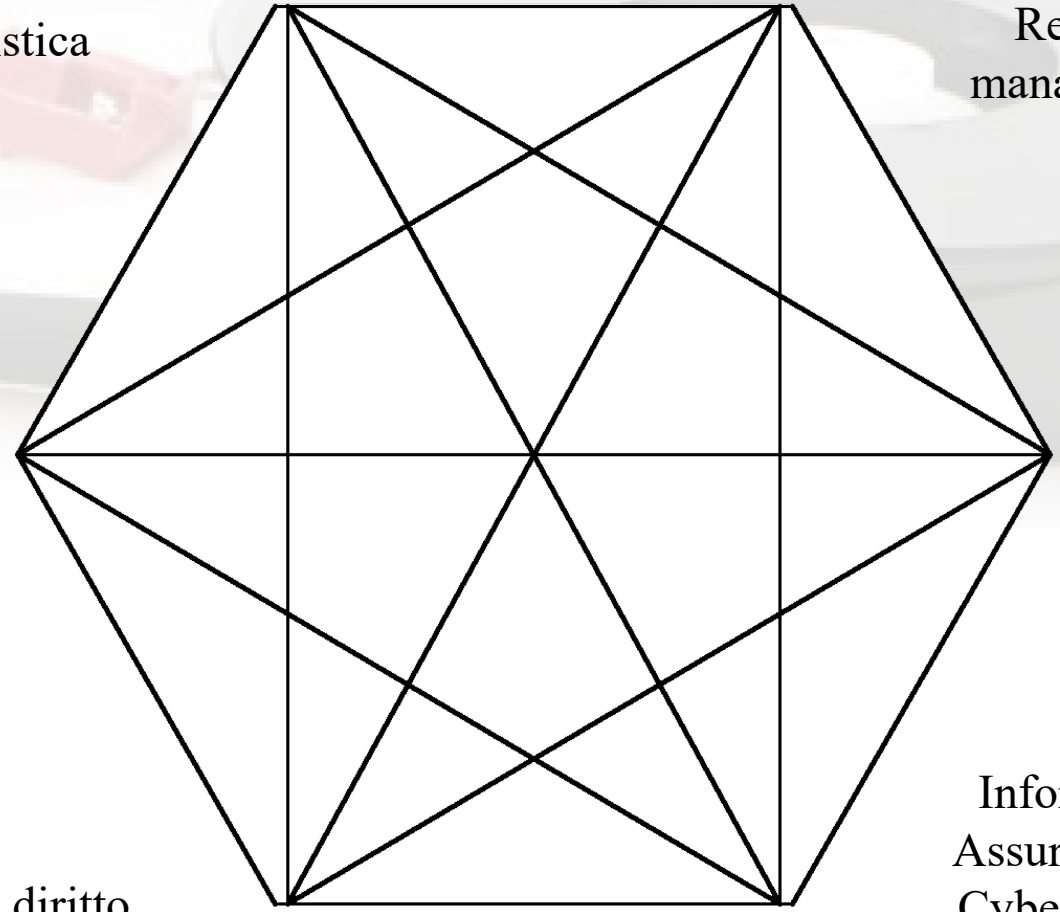
Luciana Duranti
Director, InterPARES & DRF Projects
Bologna, 16 Dicembre 2010

Digital Records Forensics Project

La natura della sfida

- Sistemi documentari ibridi
- Ambiente digitale che supporta manipolazione e riuso dei dati
- Indifferenza del produttore dei documenti al problema della loro autenticità, dovuta a fiducia nella tecnologia
- Natura proprietaria e idiosincratica delle applicazioni
- Obsolescenza di supporti e di sistemi
- Assenza di un originale
- Caratteristiche del documento digitale

Quadro concettuale per l'analisi



Quadro concettuale per l'analisi

Diplomatica/Archivistica Digitale: lo sviluppo e l'applicazione delle nostre conoscenze disciplinari ai documenti digitali

Digital Forensics: l'uso di metodi scientifici per l'acquisizione, la validazione, l'identificazione, l'analisi, l'interpretazione, la documentazione e la presentazione di fonti digitali di prova allo scopo di facilitare la ricostruzione di eventi criminali o anticipare atti non autorizzati che possono danneggiare operazioni pianificate

In particolare, abbiamo bisogno di conoscenze interdisciplinari che, una volta integrate, potrebbero essere definite **"Digital Records Forensics"**.

Digital Records Forensics

Contributi della diplomatica/archivistica

- Concetto di documento digitale
- Concetti di affidabilità, accuratezza, autenticità e autenticazione

Contributi di digital forensics:

- Processo affidabile di estrazione o riproduzione
- Categorizzazione dei documenti digitali
- Distinzione tra integrità di documenti e di riproduzioni
- Regole per la determinazione dell'integrità di sistemi
- Principi di non-interferenza e interferenza identificabile
- Principi per determinare autenticità
- Basi per l'autenticazione

Importanza del concetto di documento archivistico

L'identificazione di documenti archivistici nell'ambiente digitale costituisce un problema sia in tribunale che nei dibattiti politici.

- British Columbia Rail case: il giudice fece presente che la legislazione richiede la conservazione dei documenti archivistici del parlamento; il deputato liberale Ralph Sultan chiese "What is the definition of a record?"facendo riferimento alla controversia sulla natura di documento archivistico della posta elettronica
- La Supreme Court del Canada deve decidere se hyperlinks in un documento siano simili a note o rendano il materiale a cui lo connettono una componente del documento stesso

Definizione di documento in Computer Science

- L'aggregazione di ogni tipo di dati che siano esclusivamente leggibili per mezzo di un computer
- Costituita di Dati Binari – immagazinati in Base2, un sistema numerico che ha solo due simboli
- Ogni coppia di questi simboli è un BInary digiT, o bit

0 = 0
1 = 1
2= 10
3 = 11
4 = 100
5 = 101
6 = ????
110

Bits e Bytes

- Bits si aggregano in gruppi di 8-bit
- Ci sono 256 valori possibili da 0 a 255

  0= 00000000
  1= 00000001
  2= 00000010
  255= 11111110
  256= 11111111

- ASCII ha valori 1-127
- Esempio:

  F o u r a n d s e v e n
  70 111 117 114 32 97 110 100 32 115 101 118 101 110 32 = 00100000

Grandi Bytes

Nome	Abbr	Misura
Kilo	K	2^10 = 1,024
Mega	M	2^20 = 1,048,576
Giga	G	2^30 = 1,073,741,824
Tera	T	2^40 = 1,099,511,627,776
Peta	P	2^50 = 1,125,899,906,842,624
Exa	E	2^60 = 1,152,921,504,606,846,976
Zetta	Z	2^70 = 1,180,591,620,717,411,303,424
Yotta	Y	2^80 = 1,208,925,819,614,629,174,706,176

Il concetto di documento archivistico in archivistica

Un documento archivistico è un documento prodotto, cioè generato o ricevuto e archiviato, da una persona fisica o giuridica nel corso di un'attività pratica come suo strumento e residuo

documento è informazione affissa ad un supporto in una forma determinata

informazione è un messaggio comunicato attraverso lo spazio o il tempo e composto di dati

dato è il più piccolo pezzo di informazione che abbia significato

Un documento digitale è un documento il cui contenuto e la cui forma sono codificati usando valori numerici distinti (i valori binari 0 e 1) piuttosto che uno spettro continuo di valori (come quelli generati da un sistema analogico).

Un documento elettronico è un documento analogico o digitale che viene trasportato da un conduttore elettrico e richiede l'uso di tecnologie per essere reso intellegibile a una persona

Caratteristiche di un documento archivistico digitale in diplomatica/archivistica

- **Vincolo** esplicito con gli altri documenti interni o esterni al sistema digitale per mezzo di un codice di classificazione o di un altro identificatore unico
- Un **contesto** amministrativo identificabile
- Un **autore**, un **destinatario**, uno **scrittore**, un **produttore**, un **originatore**
- Un **atto** in cui il documento partecipa o a cui il documento fornisce supporto o proceduralmente o come parte di un processo decisionale
- **Contenuto stabile**
- **Forma fissa**

Contenuto stabile

- I dati e il messaggio nel documento sono immutati dal momento in cui sono stati scritti, e inalterabili
- Non è possibile scrivere sopra dati esistenti, alterarli, o cancellarli
- Non è possibile aggiungere dati alla prima manifestazione del documento

Forma fissa

- Il contenuto binario è affisso al supporto in modo da rimanere completo e inalterato e il messaggio può essere reso con la stessa forma documentaria che aveva quando salvato per la prima volta, anche se la presentazione digitale cambia (e.g. Word to.pdf)
- Se il contenuto presentato ogni volta è selezionato da un contenuto fisso nel sistema e le regole che governano la selezione non cambiano, ogni presentazione è una vista diversa dello stesso documento immagazzinato (e.g. dati statistici)
- "Variabilità limitata": se le variazioni nella forma sono causate dalla tecnologia o dovute all'intenzione dell'autore e ciò che le permette o causa è anche ciò che le limita

Caratteristiche del documento digitale (cont.)

- **Elementi formali**: le caratteristiche che sono visibili sulla faccia del documento, come l'intestazione, il saluto, la sottoscrizione (elementi intrinseci), o il colore, la punteggiatura, il sigillo (elementi estrinseci)
- **Attributi**: le caratteristiche, come il nome dell'autore, la data o la materia, che gli forniscono un'identità unica. Possono manifestarsi come elementi di forma o come metadati connessi al documento, o possono essere impliciti nei suoi vari contesti (documentario, procedurale, tecnologico, di provenienza, o giuridico-amministrativo)
- **Componenti digitali**: un oggetto digitale che contiene tutto o parte del contenuto di un documento e/o i dati o i metadati necessari a ordinare, strutturare o manifestare il contenuto, e che richiede un metodo specifico di conservazione. Quando il documento viene immagazzinato si scinde nelle sue componenti digitali, che sono perciò unità di conservazione. Il documento digitale non esiste come un'entità fisica dopo essere stato chiuso per la prima volta.

Documenti immagazzinati e manifesti

- **Documento immagazzinato**: le componenti digitali usate nel riprodurre un documento o più di uno, compresi i dati che devono essere elaborati per riprodurre il documento manifesto (dati di contenuto e dati di forma) e le regole per processare i dati, incluse quelle che abilitano le variazioni (dati di composizione). A volte non c'è documento manifesto che gli corrisponda.
- **Documento manifesto**: la visualizzazione o materializzazione del documento in una forma appropriata per essere presentato a una persona o un sistema. A volte non c'è un documento immagazzinato che gli corrisponda, ma viene ricreato da dati fissi di contenuto quando l'atto di un utente li associa con dati specifici di forma e composizione (e.g. un documento prodotto da una banca dati relazionale)

Tipologia di documenti digitali

Documento statico: non esiste la possibilità di cambiarne il contenuto o la forma manifestati sul monitor e ne è permessa solo l'apertura, la chiusura e la navigazione interna.

Appena un documento statico è reperito e manifestato sul monitor, il suo intero contenuto è disponibile all'utente e la sua struttura è invariabile.

L'interazione dell'utente con il sistema non cambia il contenuto o la forma del documento.

Richieste identiche di ogni utente che eserciti l'opzione di navigare all'interno del documento o di vedere il documento manifestato in modi diversi ottengono gli stessi risultati

Tipi di documenti statici

Documenti che costituiscono gli equivalenti digitali di documenti tradizionali.

Esempi

Lettere in forma di e-mail o come attachments a e-mail; relazioni su esperimenti scientifici o su osservazioni di fenomeni naturali prodotte da sistemi dinamici; registrazioni digitali di pezzi musicali; film digitale; fotografie digitali.

Tipi di documenti statici (cont.)

Documenti che non trovano un esatto equivalente tra i documenti tradizionali ma hanno forma documentaria fissa e contenuto inalterabile.

Esempi

Presentazioni di pagine web, e registrazioni di esecuzioni di opere d'arte che presentano caratteristiche che possono esistere solo in ambiente digitale; i risultati dell'atto di congelare e di catturare l'output di un sistema che modifica le sue proprie istruzioni per manipolare o presentare contenuti.

Tipologia di documenti digitali (cont.)

Documenti interattivi:

Documenti che presentano contenuto e/o forma variabile ma per i quali le regole che governano il contenuto e la forma della presentazione possono essere o fisse o variabili

Tipi di documenti interattivi

Documenti interattivi che *non sono* dinamici:

Documenti per i quali le regole che governano il contenuto e la forma della presentazione *non* variano, e per i quali il contenuto presentato in ciascun caso è selezionato tra i dati contenuti in un un deposito fisso di dati entro il sistema (=variabilità limitata).

Esempi

- *Cataloghi di vendita online, inventari archivistici online, pagine web interattive, e documenti che permettono l'esecuzione di musica e altre opere d'arte, come computer patches.*

Tipi di documenti interattivi (cont.)

Documenti interattivi che ***sono*** dinamici:

Documenti per i quali le regole che governano il contenuto e la forma della presentazione possono variare

Sottotipi:

1. Documenti per i quali le regole che governano il contenuto della presentazione variano perchè essi includono o sono influenzati da dati che cambiano frequentemente, come

   Documenti in sistemi disegnati in un modo che permette l'aggiornamento, la sostituzione o l'alterazione dei dati ma non il mantenimento dei dati precedenti, e siti web che acquisiscono dati dagli utenti o riguardanti le interazioni degli utenti con il sito o iI loro interventi sul sito, e usano quei dati per generare o per determinare le presentazioni successive.

Documenti interattivi dinamici (cont.)

2. Documenti il cui contenuto varia perchè include dati ricevuti da fonti esterne e non immagazzinati nel sistema, come *Siti web che presentano informazione su soggetti come il tempo o il tasso di cambio della valuta; opere d'arte interattive*
3. Documenti prodotti in applicazioni di "dynamic computing", come Geographic Information Systems, che selezionano gruppi diversi di regole per produrre i documenti sulla base delle variazioni nell'input dell'utente, nelle fonti dei dati che formano il contenuto, e nelle caratteristiche del contenuto stesso

Tipi di documenti interattivi (cont.)

4. Documenti prodotti da “adaptive or evolutionary computing applications”, dove il software che genera i documenti può cambiare autonomamente, come

   Siti web che includono la schedatura e modellatura dei mercati finanziari e alcuni tipi di siti per intrattenimento.

Funzione del documento digitale

- ***Ad substantiam*** and ***ad probationem*** (dispositivi e probativi=documenti legali)
- **Di supporto**: generati per essere usati nel corso di varie attività come fonte di informazione (e.g., GIS)
- **Narrativi**: generati come strumento di comunicazione ma la loro produzione non è richiesta dal sistema giuridico (e.g., e-mails, rapporti, web sites)

Nuove funzioni

- **Istruttivi**: indicano la forma di presentazione di contenuto esterno al documento in questione (e.g., spartiti, copioni, regole manuali di procedura, istruzioni per riempire moduli)
- **Abilitanti**: abilitano esecuzioni artistiche (software patches), transazioni (interacting business applications), la condotta di esperimenti (un workflow prodotto e usato per fare un esperimento, di cui è strumento e residuo), l'analisi di dati di osservazione (interpreting software), etc.

Definizione di documento in Digital Forensics

- Una combinazione delle definizioni di computer science e archivistica: un oggetto digitale costituito di bits e bytes e prodotto o ricevuto nel corso ordinario degli affari
- Digital Forensics identifica quattro categorie di documenti, tra le quali tre possono essere usati come fonte di prova per il loro contenuto; una può solo essere usata come prova materiale.

Definizione di documento in Digital Forensics (cont.)

Categorizzazione dei documenti digitali:

1. Documenti prodotti e tenuti in un computer
2. Documenti prodotti da un computer o dall'interazione di sistemi
3. Documenti prodotti da una persona fisica e un computer
4. Oggetti dinamici: nell'Internet

I primi sono esaminati e valutati come documenti archivistici (inerentemente affidabili)

I secondi sono esaminati e valutati come ogni prova materiale

I terzi e i quarti sono esaminati e valutati da entrambi i punti di vista.

Un documento degno di fede: il punto di vista dell'archivistica contemporanea

Un documento affidabile, accurato e autentico

(*in contrasto con la diplomatica classica, che fa coincidere i tre concetti, presumendo che l'ultimo implichi gli altri*)

Affidabilità

La capacità di un documento di rappresentare i fatti di cui tratta

(*è la responsabilità del produttore ed è stabilita sulla base della completezza del documento e dei controlli stabiliti sulla procedura che lo produce*)

Accuratezza

- si riferisce all'esattezza e correttezza del contenuto
- è la responsabilità dell'autore e dell'archivista
- dipende dal controllo sui processi che registrano i dati e che li trasferiscono tra sistemi e nel tempo

Autenticità

Si riferisce al fatto che un documento sia ciò che dichiara di essere e non sia stato falsificato o corrotto

(*è a rischio durante la trasmissione e la conservazione, è la responsabilità sia del produttore che dell'archivio, e si stabilisce sulla base del rispetto dei requisiti stabiliti per presumere, verificare o mantenere l'autenticità*)

Autenticità è una proprietà del documento che lo accompagna per tutto il tempo che il documento esiste. Si stabilisce sulla base **dell'identità** e **dell'integrità** del documento.

Identità di un documento

è costituita dagli attributi di un documento che, nel loro insieme, lo caratterizzano in modo unico e lo distinguono da altri documenti.

Questi attributi includono:

i nomi delle persone che concorrono alla sua formazione,

le date di produzione e trasmissione,

la materia o l'atto a cui si riferisce,

la sua forma documentaria e digitale,

l'espressione della sua relazione con gli altri documenti,

l'indicazione di allegati,

il nome dell'ufficio competente,

esistenza di firma digitale.

Integrità di un documento

- *La sua interezza e perfezione*. Un documento ha integrità se è intatto e non corrotto, cioè se il messaggio che intendeva comunicare per raggiungere il suo scopo è inalterato
- L'integrità fisica di un documento, come per esempio il numero appropriato di bit strings, può essere compromessa, purchè l'articolazione del contenuto e i necessari elementi formali rimangano gli stessi.
- Integrità può essere dimostrata o da evidenza che appare sul documento o da attributi, espressi come metadati, relativi al documento, o in uno o più contesti
- I metadati che la dimostrano sono relativi alla responsabilità per il documento e alle sue trasformazioni technologiche

Attributi di integrità

nome della persona competente per la pratica
nome della persona responsabile per il documento
esistenza di annotazioni
indicazione di cambiamenti tecnici
indicazione di firme digitali aggiunte o rimosse
data della rimozione pianificata dal sistema
data di trasferimento al custode designato
data di distruzione pianificata
esistenza e collocazione di duplicati

Quadro concettuale per i requisiti per l'autenticità

- Con i sistemi elettronici, la presunzione di autenticità deve essere basata su prova che un documento è ciò che dichiara di essere e che non è stato modificato o corrotto in modo sostanziale.
- Per stabilire l'autenticità di un documento, la persona responsabile per la sua conservazione deve poter stabilire la sua identità e dimostrare la sua integrità durante il processo di valutazione per la selezione
- Tale persona assume il ruolo di **custode designato affidabile**

Un documento degno di fede: il punto di vista di Digital Forensics

Affidabilità: un documento è affidabile se la sua *fonte* è affidabile, cioè se è stato prodotto o da una persona affidabile o da un software affidabile

Un software affidabile è un open source software, perchè i processi di produzione e tenuta dei documenti possono solo essere autenticati o 1) descrivendo il processo o il sistema usato per generare un risultato, o 2) mostrando che il processo o il sistema hanno prodotto un risultato accurato. In entrambi i casi il codice del software deve essere conosciuto.

Un documento degno di fede: il punto di vista di Digital Forensics

Accuratezza è una componente di autenticità e, in modo specifico, di integrità. Oggetti digitali sono garantiti accurati se sono ripetibili.

Ripetitibilità, che è uno dei principi fondamentali di digital forensics, si basa sulla documentazione di tutte le azioni a cui l'evidenza digitale è stata soggetta.

Open source software è la scelta migliore anche per stabilire l'accuratezza, specialmente quando si fanno conversioni o migrazioni, perchè permette una dimostrazione pratica del fatto che niente avrebbe potuto essere alterato, perso, inserito o distrutto nel corso di tali processi.

Un documento degno di fede: il punto di vista di Digital Forensics

Autenticità significa che i dati o il contenuto del documento sono ciò che dichiarano di essere e sono stati prodotti oppure sono pervenuti dalla fonte da cui si dichiara che siano stati prodotti o che siano pervenuti. Il termine "fonte" si riferisce o a una persona (fisica o giuridica), o a un sistema, software o hardware.

Come in diplomatica, **autenticità implica integrità**, ma il contrario non è vero, cioè **integrità non implica authenticità**.

Integrità: il punto di vista di Digital Forensics

Integrità dei dati: il fatto che i dati non siano stati modificati o accidentalmente o intenzionalmente "senza appropriata autorizzazione."

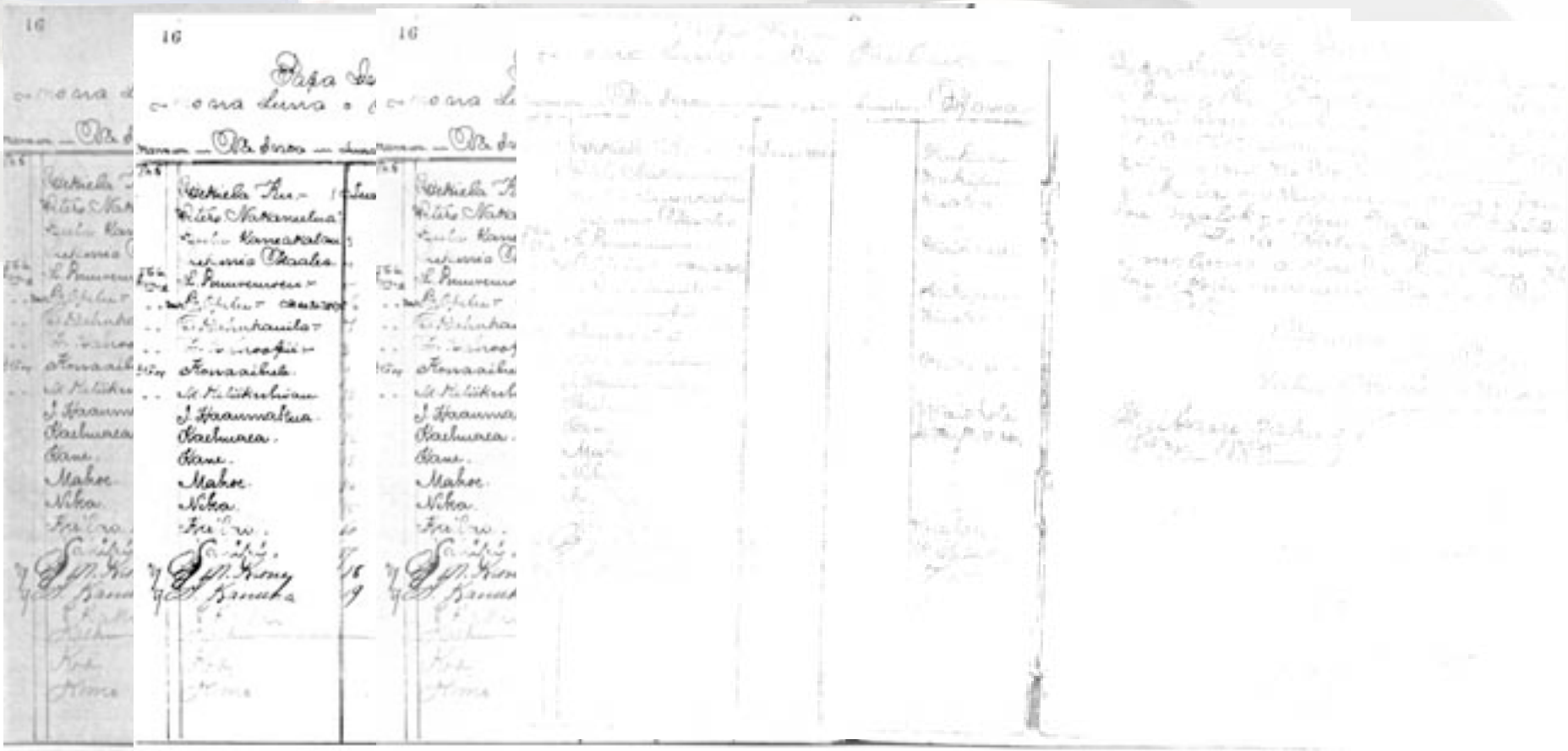
Si basa su **Bitwise Integrity**

Integrità: il punto di vista di Digital Forensics (cont.)

Bitwise Integrity

- I bits originali sono completi e inalterati dal momento della loro “capture”
- I bits originali sono nello stesso ordine e hanno lo stesso valore
- Un cambiamento minimo in un bit costituisce un valore molto diverso sia sul monitor che in termini di azione avvenuta in un programma o un database.

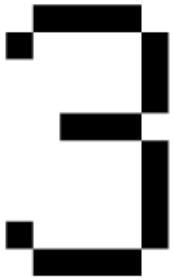
Perdita di fedeltà:
Analogico verso Digitale



Digital Records Forensics Project

Perdita di fedeltà: Analogico verso Digitale (cont.)

- Bits originali 101
- Cambia cosi': 110
- O cosi' 011



- Stessi bits, but valore differente

Perdita di fedeltà

- Si può prevenire per mezzo di permessi di accesso e controlli sull'accesso
- Richiede metodi per determinare se l'oggetto digitale è stato alterato
- Non può basarsi sulle dimenzioni del file, sulle date o su altri attributi
- Richiede audit logs e metodi forti

Checksum

- Forma di autenticazione di dati
  - Se la checksum non rimane identica nella trasmissione attraverso lo spazio o nel tempo, i dati sono corrotti o incompleti
- Somma il valore dei bits in un gruppo
  - Se meno di 255 il valore reale viene usato
  - Se più di 255, il totale viene diviso per 256
- Esempio:

Byte1	Byte2	Byte3	Byte4	Byte5	Byte6	Byte7	Byte8	Total	Checksum
212	232	54	135	244	15	179	80	1151	127

1,151 / 256 = 4.496 (arrotondato a 4)
4 x 256 = 1,024

1,151 - 1,024 = 127

Algoritmo HASH

- Computato dal numero di base usando un algoritmo
- Quasi impossibile derivarlo senza i dati originali
- Normalmente usa un algoritmo di 128bit o più alto, cioè 2^{128}
- Esempio:

Input	Hash	Value
10,667	Input x143	1,525,381

Valore HASH

- Comprime i bits di un oggetto in un valore di grandezza fissa
- Estremamente difficile ricostruire l'oggetto sulla base del valore hash
- Hash comuni
  - SHA-1 160 bit
  - RIPEMD-160 160bit
  - MD5 – 128 bit

Integrità: il punto di vista di Digital Forensics (cont.)

Integrità della riproduzione: il fatto che, dato un gruppo di dati, il processo di produrre un duplicato non modifichi i dati (intenzionalmente o accidentalmente) e che il duplicato sia una copia "bitwise" identica del gruppo originale di dati.

Gli esperti di digital forensics connettono l'integrità del duplicato al tempo e usano *time stamps* a questo scopo.

Integrità: il punto di vista di Digital Forensics (cont.)

Integrità del Computer: il computer produce risultati accurati quando è usato e operato in modo appropriato e fu cosi' usato e operto quando generò l'oggetto digitale in questione.

Integrità del sistema: il sistema esegue le sue funzioni in modo indisturbato, senza manipolazione non autorizzata, intenzionale o accidentale

Entrambe implicano **integrità di hardware and software**

Integrità di computer o di sistema

Si deduce da:

- Misure di sicurezza sufficienti a prevenire accesso o non autorizzato o non tracciabile a computers, networks, devices, o storage.
  - Data
  - Users/permissions
  - Passwords
  - Logs
  - Firewalls

System e Auditing Logs

Un gruppo di files *automaticamente* generate per tracciare le azioni fatte, i servizi usati, o files consultate o modificate, quando, da chi e da dove

- Web logs
- Access logs
- Transaction logs

Web Log tipico

- Client IP Address
- Request Date/Time
- Page Requested
- HTTP Code
- Bytes Sent
- Browser Type
- OS Type
- Referrer

Access Log tipico

- User account ID
- User IP address
- File Descriptor
- Bind record results
- Actions taken upon record
- Unbind record
- Closed connection

Transaction Log tipico

- Storia delle azioni condotte su un sistema per assicurare ACID in caso di crash (Atomicity, Consistency, Isolation, Durability)
- Sequence number
- Link to previous log
- Transaction ID
- Type
- Updates, commits, aborts, completes

Auditing Logs

- Sempre più richiesti dalla legge per dimostrare l'integrità del sistema
- Se ben configurati, e protetti, sono efficaci
- Determinano l'efficacia delle misure di sicurezza
- Identificano gli errori
- Forniscono notifica istantanea di ciò che occorreevents

Auditing Logs (cont.)

- Permettono di determinare le responsabilità
- Forniscono uno snapshot per ricostruire gli eventi dopo il fatto (‘black-box’)
- Rispondono alle domande: Chi, Che Cosa, Dove, Quando
- Solo se conservati per un tempo sufficiente (spazio vs. osti vs. rischio vs. trasparenza)

Integrità: il punto di vista di Digital Forensics (cont.)

Integrità del processo: Procedura formalizzata per l'acquisizione, il recupero, l'interpretazione e la presentazione di evidenza.

Esempio: UK ACPO:

- No action taken by law enforcement agencies or their agents should change data held on a computer or storage media which may subsequently be relied upon in court.
- In exceptional circumstances, where a person finds it necessary to access original data held on a computer or on storage media, that person must be competent to do so and be able to give evidence explaining the relevance and the implications of their actions.
- An audit trail or other record of all processes applied to computer based electronic evidence should be created and preserved. An independent third party should be able to examine those processes and achieve the same result.
- The person in charge of the investigation (the case officer) has overall responsibility for ensuring that the law and these principles are adhered to.

Regole per stabilire l'integrità di computer e sistema

La teoria, la procedura o il processo per produrre o gestire e mantenere il documento

- Devono essere stati testati e non possono essere stati manomessi
- Devono essere stati sottoposti alla valutazione di esperti e/o sono risultati in pubblicazione (e.g. standards)
- Devono essere generalmente accettati dalla comunità scientifica competente, e
- La ratio di errore conosciuta o potenziale che offrono deve essere accettabile

Il sistema migliore

- Gli attributi di tale sistema sono **ripetitibilità delle operazioni, verificabilità, oggettività e trasparenza,** che richiedono la documentazione accurata di qualunque operazione sul sistema e all'interno del sistema.

- Come già stabilito, **open source software** è la scelta migliore per valutare integrità, specialmente in caso di upgrade, conversione e migrazione.

Principi relativi all'integrità

Non-interferenza: il metodo usato per fare upgrade, conversione o migrazione non cambia il contenuto e la forma documentaria del documento nativo

Interferenza identificabile: il metodo usato per fare upgrade, conversione o migrazione altera il documento nativo ma i cambiamenti sono identificabili

Questi principi, che incorporano la posizione etica e professionale dell'archivista, caratterizzano anche il suo ruolo istituzionale di custode affidabile che esercita il controllo sul sistema di produzione e gestione dei documenti

Principi relativi all'autenticità

- **Autenticità** è la certezza dell'identità della fonte, persona o sistema. Come detto, autenticità implica integrità ma non viceversa.
- Autenticità si può basare sulla dichiarazione di un esperto che il sistema di gestione e tenuta dei documenti e le procedure che lo controllano sono affidabili, sulla base di
    1. uno schema di metadati di identità e integrità
    2. uno schema di classificazione (titolario)
    3. regole di selezione e scarto connesse al titolario
    4. un sistema di registrazione di protocollo
    5. un sistema di reperimento
    6. previlegi di accesso a documenti archiviati

Autenticazione: il punto di vista diplomatico/archivivistico

- Una dichiarazione di autenticità che risulta o dall'inserimento o dall'aggiunta di un elemento o di un'affermazione al documento, secondo norme legislative
- Un metodo per provare che un documento è quello che dichiara di essere in un momento determinato (sigilli, firme digitali)

Autenticazione

- Certe tecniche matematiche si dice che forniscano un meccanismo **incontrovertibile** per assicurare l'autenticità di oggetti digitali (e.g., firme digitali crittografiche)
- A tali tecnologie si è dato valore legale (e.g., European Directive on electronic signatures, Security and Exchange Commission on hash functions).
- La firma digitale è abilitata da una infrastruttura complessa a chiave pubblica che è molto costosa (PKI)
- La firma digitale è basata sulla stessa tecnica matematica usata dalla cifratura, ma **non** dà confidenzialità

Autenticazione

- La firma digitale è uno strumento valido per garantire l'autenticità dei documenti attraverso lo **spazio……ma non nel tempo**!
- I sistemi giuridici nordamericani non la considerano il miglior metodo di autenticazione
- La firma digitale è soggetta a obsolescenza e quindi complica il problema della conservazione digitale
- Il metodo prevalente di autenticazione è **una catena di custodia legittima** in sistemi che passino i test di **ripetitibilità, verificabilità, oggettività e trasparenza.**

Autenticazione: il punto di vista di Digital Forensics

Prova di autenticità può essere fornita da un **testimone** che può attestare l'esistenza o la sostanza del documento sulla base della sua familiarità col documento stesso, o **da un computer programmer** che mostri che *il computer process o il sistema produce risultati accurati quando viene usato o operato in modo giusto e che ciò avvenne quando il documento fu prodotto.*

La forza della prova circostanziale può essere aumentata da metadati che mostrino (1) la data e l'ora esatte della produzione, (2) quale computer ha prodotto il documento e (3) quale computer lo ha ricevuto.

Altri mezzi di autenticazione

Sistemi biometrici di identificazione e crittografia **non sono** considerati i metodi migliori di autenticazione.

Inferenza di integrità del sistema: Evidenza circostanziale che un sistema esegue le sue funzioni in modo corretto

Una dichiarazione fatta da un esperto sulla base dell'**affidabilità del sistema di tenuta dei documenti e delle procedure che lo controllano** (quality assurance).

Una catena di custodia legittima è la base principale per dedurre l'autenticità e autenticare un documento.

E in assenza di catena di custodia legittima?

Questa **Catena Ininterrotta di Custodia** (vedi InterPARES COP model) è possibile—almeno in teoria– per i documenti digitali di istituti e enti pubblici e privati per cui esiste un'entità archivio designata esterna (ministero, tribunale) o interna (università, banca).

Come conservare i documenti digitali per cui non esiste un archivio storico designato?

Parliamo di archivi di professionisti, persone, famiglie, studi (architetti, avvocati, dentisti), ditte, ecc.

Due situazioni possibili

1. La sovrintendenza o l'archivio storico interviene quando i documenti sono ancora correnti e
   - il produttore desidera collaborazione
   - Il produttore accetta consigli anche se senza interferenza
2. La sovrintendenza o l'archivio storico si trova di fronte a documenti non correnti, spesso su supporti esterni al sistema in cui sono stati prodotti

Situazione di collaborazione

a) Creare un' infrastruttura

- stabilire la portata e gli obbiettivi
- acquisire risorse
- focus sui documenti digitali
- dare consiglio su tecnologia e formati
- fornire esempi
- sviluppare policy e procedure, assegnare responsabilità
- sviluppare strategie di mantenimento

Collaborazione (cont.)

b) Valutare i documenti

- Identificare i documenti tra gli oggetti digitali prodotti
- Identificare co-autori e proprietari multipli
- Determinare l'autenticità e documentarla
- Determinare problemi di privacy
- Monitorare i documenti da conservare
- Identificare tutte le componenti digitali
- Determinare la fattibilità della conservazione
- Sviluppare un piano di versamento

Collaborazione (cont.)

c) Versamento

- Migrare i documenti all'ambiente tecnologico dell'archivio
- Conservare il formato logico più vecchio ancora disponibile
- Evitare l'acquisizione di duplicati
- Documentare ogni attività a cui i documenti sono sottoposti

Consigli senza interferenza

Si preparano linee guida chiare e comprensibili

- Un opuscolo generale
  - Come scegliere software e formati (standards)
  - Come organizzare i documenti
  - Come mantenere i documenti accessibili nel tempo (back-ups, system upgrade, conversione, dispersione)
  - Come prevenire la perdita di documenti

Consigli senza interferenza (cont.)

- **Un opuscolo specifico sulla gestione e tenuta dell'e-mail**
  - Come organizzarla, anche in relazione ad altri documenti
  - Come trattare gli attachments
  - Come trattare i threads
  - Come conservarla in altri formati
  - Come selezionarla
- **Un opuscolo su come donare i propri documenti digitali**
  - Perchè donare, cosa donare, come, quando
  - Considerazioni sui diritti intellettuali, di privacy, sicurezza, accesso
  - Lista di persone/enti a cui rivolgersi per consigli e aiuto
- **Lezioni e workshops su come proteggere le proprie foto, ecc**.

Assenza di contatto

Materiali che ci potrebbero essere presentati:

- **Documenti generati da word processing**
- **E-mail con word processing attachments**
- **Foto, video e registrazioni musicali**
- **Agende o calendari**
- **Web portals, blogs & wikis**
- **Registrazioni di videoconferenze & webcasting**
- **Databases**
- **Flash drives & altre storage devices con contenuti vari**
- **Remote PDAs, Blackberrys, etc. etc. etc.**

Assenza di contatto

Siamo di fronte a oggetti digitali non correnti, spesso su supporti esterni al sistema in cui sono stati prodotti

Regole da seguire:

- Si crea una copia o un'immagine (non sono la stessa cosa)
- Si analizza la copia o l'immagine e si determina se gli oggetti digitali sono documenti archivistici
- Se lo sono, si determina la loro autenticità
- Si determina la fattibilità della conservazione
- Si procede al versamento

Il processo di Digital Forensics

1. Produzione della copia delle entità logiche o dell'immagine dell'hard drive o supporto esterno
2. Identificazione degli oggetti di interesse potenziale
3. Analisi degli oggetti identificati
4. Valutazione e interpretazione dei risultati
5. Presentazione dei risultati in un rapporto che descrive in dettaglio le caratteristiche degli oggetti, l'interpretazione dei fatti e le opinioni di esperti
6. La revisione tecnica e amministrativa da parte di un soggetto neutrale

Trasferimento all'archivio storico

Le procedure e i sistemi usati per trasferire i documenti all'archivio storico, mantenerli e riprodurli devono incorporare controlli adeguati e efficaci per garantire l'identità e l'integrità dei documenti, e specificatamente devono assicurare che:

- Sia mantenuta la custodia ininterrotta dei documenti;
- Siano rispettate regole relative a metadati, privilegi di accesso
- Siano messe in atto e monitorate procedure di protezione, sicurezza e controllo; e
- Il contenuto dei documenti rimanga inalterato dopo la riproduzione

Conservazione

- Non è possible conservare un documento elettronico, ma solo la capacità di riprodurlo. Perciò dobbiamo proteggere le componenti digitali che contengono la sostanza delle parti costitutive e dell'identità del documento.
- Dobbiamo accettare che è impossibile mantenere letteralmente inalterato un documento elettronico
- L'unico modo di provare che un documento elettronico è autentico è produrre una copia autentica dei documenti del produttore presunti o verificati autentici

Documentazione della riproduzione

L'attività di riproduzione deve essere documentata e la documentazione deve includere:

- La data della riproduzione e il nome della persona responsabile;
- Una descrizione della relazione tra i documenti acquisiti dal produttore e le copie prodotte dall'archivista;
- Una descrizione dell'impatto del processo di riproduzione sulla forma, il contenuto, l'accessibilità e l'uso dei documenti; e
- Nei casi in cui la copia non riproduca pienamente e fedelmente gli elementi che esprimono l'identità e integrità del documento, tale informazione deve essere documentata dall'archivista e questa documentazione deve essere facilmente accessibile all'utente.

Mantenere sempre il formato nativo

Descrizione archivistica

La descrizione archivistica del fondo che contiene i documenti elettronici deve includere—oltre all'informazione sui contesti giuridico-amministrativo, di provenienza, procedurale e documentario—anche informazione sui cambiamenti che i documenti hanno subito fin da quando sono stati prodotti

Conservazione affidabile

- La fonte più importante per stabilire l'autenticità dei documenti è la descrizione archivistica
- Descrizione come attestazione collettiva dell'autenticità dei documenti in un fondo e di tutte le loro relazioni
- Descrizione come prospettiva storica sui documenti e sulle loro trasformazioni

Altri concetti rilevanti

- **Catena di custodia legittima vs. Catena di documentazione** (le condizioni di acquisizione dei documenti, l'identità di coloro che li hanno trattati, il tipo, la durata e le conseguenze di ogni attività, le condizioni di sicurezza in cui i documenti sono stati trattati o conservati, il modo in cui i documenti sono stati trasferiti al custode successivo ecc.)
- **Prevenzione vs. Preparazione** (per scoperta e risposta)
- **Identificazione e Acquisizione vs. Ricerca e Sequestro**
- **Criptografia vs. Steganografia** (una forma nascosta di protezione dell'informazione)
- **Copia vs. Immagine**

Il punto di vista di Digital Forensics: Immagine

Immagine: una riproduzione bit per bit dei dati su un supporto (hard drive, disk, nastro, ecc.) fatta prima di esaminarne il contenuto.

Produrre l'immagine di un disco è importante per digital forensics per:

- garantire che l'informazione sul disco non venga cambiata inavvertitamente.
- riprodurre i risultati dei test forensici sull'evidenza originale.
- catturare informazioni normalmente invisibili al sistema operativo quando in uso.

Il punto di vista di Digital Forensics: Copia

Copia: reproduzione selettiva di files

- Si può copiare solo ciò che si vede
- Raramente l'atto di copiare include conferma di completezza
- Copiare consiste nel muovere files su un altro supporto individualmente
- Fornisce una visione incompleta del supporto digitale

Perchè Digital Forensics?

Nel contesto dell'**archivio corrente**: per garantire la capacità dei propri documenti di servire come prova

Nel contesto dell'**archivio storico,** per

- estrarre oggetti digitali da software e hardware obsoleti
- autenticare oggetti digitali di provenienza incerta
- documentare il contesto tecnologico dei documenti
- proteggere il materiale digitale da conservare permanentemente, e
- anche qui, garantire la capacita dei documenti di continuare a servire come prova attraverso conversioni e migrazioni

Web Sites

www.digitalrecordsforensics.org

www.interpares.org