

UN FUTURO PER IL PRESENTE

POLITICHE STRATEGIE E STRUMENTI DELLA CONSERVAZIONE DIGITALE

La conservazione digitale tra aspetti giuridici e archivistici nel panorama internazionale

Luciana Duranti, University of British Columbia, Vancouver, Canada

Bologna 11-12 aprile 2011

A brave new world & i suoi problemi

Il 28 novembre 2010 WikiLeaks cominciò a pubblicare la maggior quantità di documenti segreti mai rilasciata fornendo un esempio plateale di come, nell'ambiente digitale, i diritti più importanti—anche se spesso opposti—su cui le moderne democrazie si basano siano in pericolo: il diritto

- alla trasparenza e alla confidenzialità
- all'accesso e alla privacy
- alla memoria e all'oblio
- alla conoscenza e al vantaggio economico
- alla creatività dell'uso e all'integrità dell'opera

Social Media & research data

L'uso dei social networks e la contestata acquisizione da parte della Library of Congress dell'archivio di Twitter hanno elevato il livello di consapevolezza della sfida che ci viene presentata.

Una sfida che si estende al **ri-uso di data sets** per scopi diversi da quelli per cui furono originalmente sviluppati, e alla conseguente acquisizione di banche dati per la conservazione permanente, ma anche alla **perdita di documenti vitali** a causa della fragilità e obsolescenza della tecnologia

Nessuno di questi fenomeni è nuovo. La **dimensione** che essi assumono lo è.

Una crisi informativa di enormi proporzioni

In **Canada**, in 2009, l' Information Commissioner, in un rapporto intitolato *A Dire Diagnosis for Access to Information in Canada*, scrisse: "The poor performance shown by institutions is symptomatic of what has become a **major information management crisis**. A crisis that is only exacerbated with the pace of technological developments. Access to information has become hostage to this crisis and is about to become its victim. There is currently no universal and horizontal approach to managing or accessing information within government. Some institutions don't even know exactly what information they are holding."

La risposta non è legislazione, ma policy

La legislazione deve fornire un quadro stabile di principi. Nel contesto della legislazione, **politiche, strategie** e **procedure** devono assicurare l'approccio orizzontale coerente e continuo, sia all'interno di ciascuna istituzione o ente che tra di loro, che l'Information Commissioner del Canada considera necessario

Di conseguenza, **Library and Archives Canada** ha pubblicato una serie di policies e linee guida:
<http://www.lac-bac.gc.ca/digital-initiatives/012018-2000-e.html>

Approcci simili

The National Archives of the UK ha rilasciato in marzo un documento in bozza che guida archivi e organizzazioni che intendono istituire un proprio archivio nello sviluppare una politica di conservazione:

<http://www.nationalarchives.gov.uk/information-management/projects-and-work/policy-guidance.htm>

Questo documento fa riferimento a molte linee guida per politiche di conservazione, tra cui quelle di InterPARES.

L'American Council for Technology e l'Industry Advisory Council hanno rilasciato in aprile una bozza di policy per l'uso di social media che, diversamente dalle molte policies di piccole e grandi organizzazioni, come la National Library of Australia, discute la conservazione dei documenti prodotti usando social media:

[http://www.actgov.org/knowledgebank/whitepapers/Documents/Shared%20Interest%20Groups/Collaboration%20and%20Transformation%20SIG/Best%20Practices%20of%20Social%20Media%20Records%20Policies%20-%20CT%20SIG%20-%202003-31-11%20\(3\).pdf](http://www.actgov.org/knowledgebank/whitepapers/Documents/Shared%20Interest%20Groups/Collaboration%20and%20Transformation%20SIG/Best%20Practices%20of%20Social%20Media%20Records%20Policies%20-%20CT%20SIG%20-%202003-31-11%20(3).pdf)

Linee guida per i produttori

Altre organizzazioni, come il National Archives degli Stati Uniti (NARA), preferiscono guidare i produttori nell'uso delle varie tecnologie:

Web Publishing: Piattaforme usate per creare, pubblicare, e reusare contenuti.

Microblogging (Twitter, Plurk); Blogs (WordPress, Blogger); Wikis (Wikispaces, PBWiki); Mashups (Google Maps, popurls)

Social Networking: Piattadorme usate per fornire interazione tra utenti.

Social Networking tools (Facebook, LinkedIn); Social Bookmarks (Delicious, Digg); Virtual Worlds (Second Life, OpenSim); Crowdsourcing/Social Voting (IdeaScale, Chaordix)

File Sharing/Storage: Piattaforme per condividere contenuti.

Photo Libraries (Flickr, Picasa); Video Sharing (YouTube, Vimeo); Storage (Google Docs, Drop.io); Content Management (SharePoint, Drupal)

Ma il problema più serio è rappresentato dal **Cloud Computing, la Nuvola**

Che cosa è la Nuvola?

National Institute of Standards and Technology

“La Nuvola è un modello che permette un accesso conveniente su network, su domanda, a un insieme condiviso di risorse tecnologiche configurabili (e.g., networks, servers, spazio per storage, applicazioni, e servizi) che può essere rapidamente preparato e fornito con minimo sforzo di gestione e minima interazione con chi fornisce il servizio.”

National Archives and Records Administration

“La Nuvola è una tecnologia che permette agli utenti di avere accesso a dati condivisi e a usare servizi di computing via Internet o via un Network Privato Virtuale (VPN). Dà agli utenti accesso a risorse senza dover costruire un’infrastruttura che supporti tali risorse all’interno del proprio ambiente o network.”

Ampio accesso al network:

Funzionalità sono disponibili sul network e accessibili attraverso meccanismi standard (e.g., laptops, mobile devices, etc.)

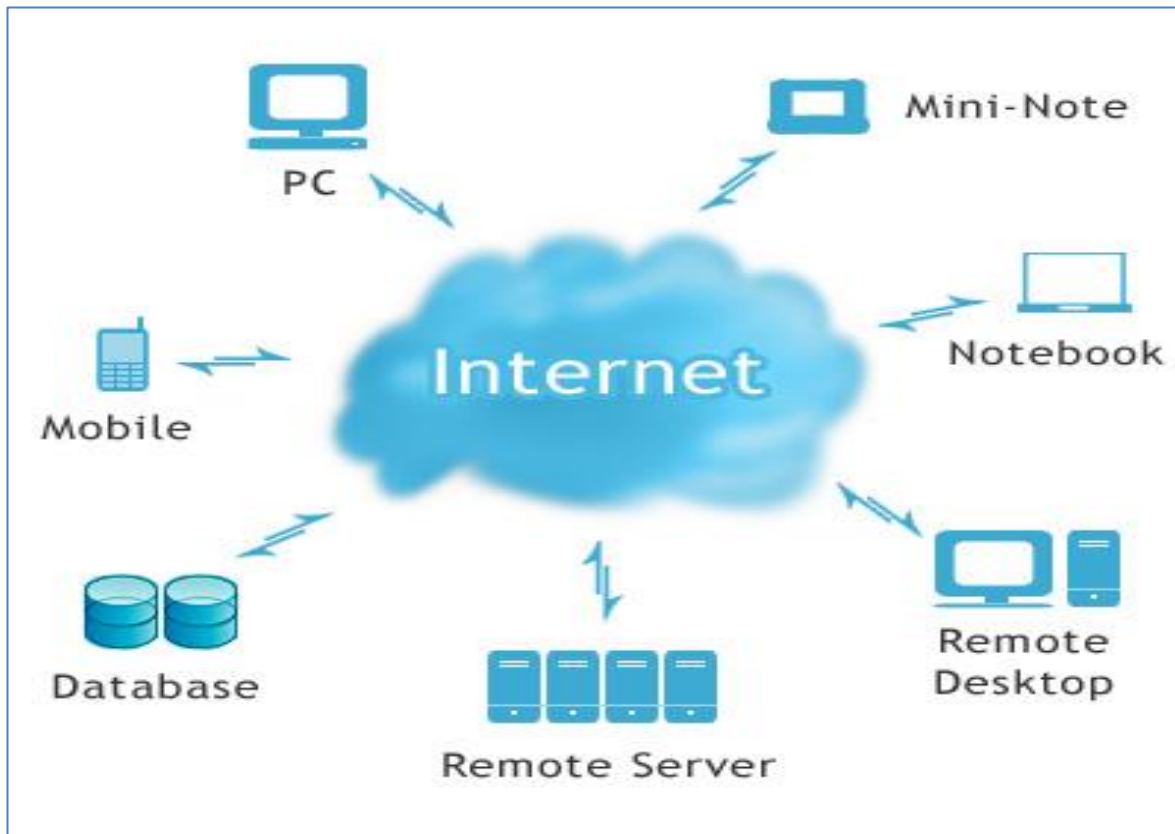
Rapida flessibilità : Funzionalità possono essere fornite rapidamente e con flessibilità, in alcuni casi automaticamente, per diminuire velocemente o aumentare i servizi



Servizio controllato: Cloud systems controllano automaticamente e ottimizzano l'uso delle risorse per mezzo di una capacità di misurazione al livello di astrazione appropriato al tipo di servizio (e.g., immagazzinamento, reperimento, bandwidth, e gestione di conto e-mail)

Self-service on-demand: Un utente può unilateralmente scegliere funzionalità di computing, come tempo di utilizzazione del server o quantità di immagazzinamento sul network, come vuole, automaticamente senza interagire con ciascun fornitore di servizi

Uso condiviso di risorse: Le risorse di computing del fornitore sono messe insieme per servire clienti multipli usando un modello di inquilini multipli, in cui risorse diverse, fisiche e virtuali, vengono dinamicamente assegnate e riassegnate a seconda delle richieste dei clienti



è possibile avere accesso attraverso Internet a molti servizi,
including Software, Piattaforma, or Infrastruttura

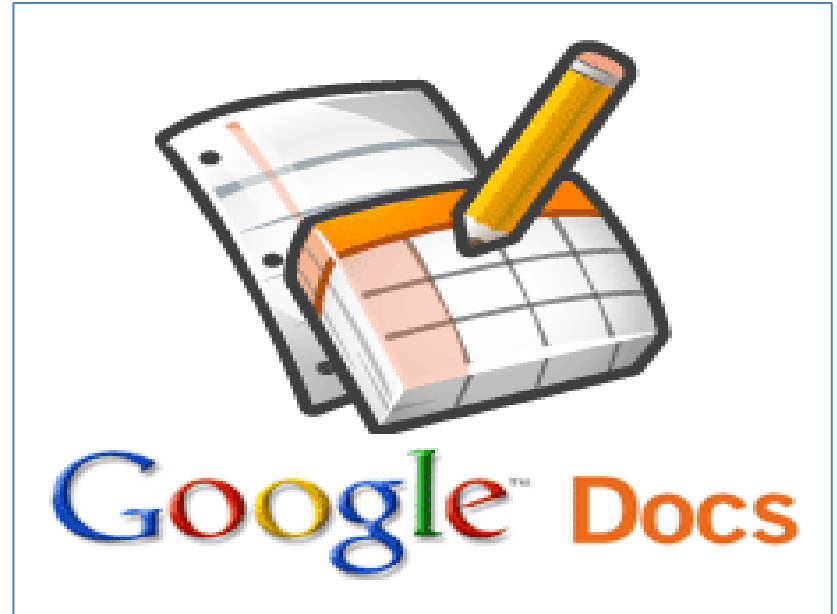
Software as a
Service (SaaS)

Platform as a
Service (PaaS)

Infrastructure as a
Service (IaaS)

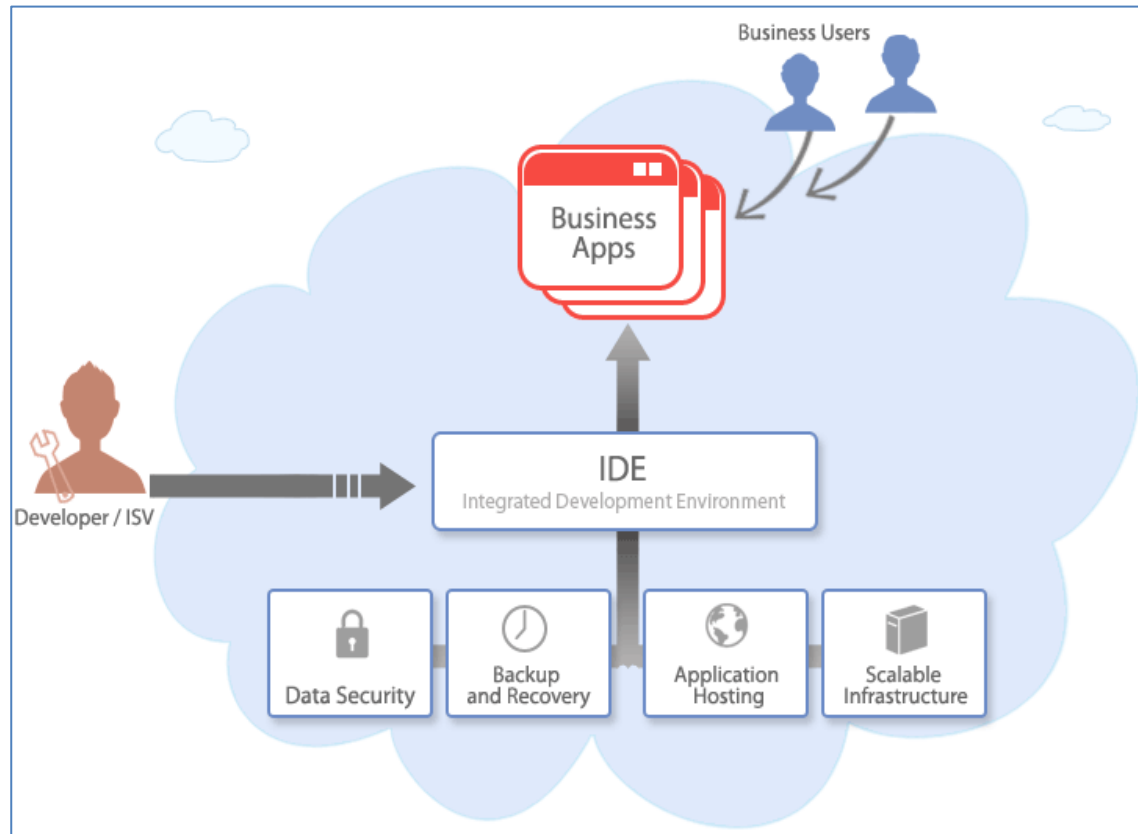
*Service
Models*

Software as a Service (SaaS)



Software as a Service è la forma più comune di Cloud Computing. Consiste nel poter usare le applicazioni del fornitore che risiedono in una infrastruttura nuvola. Le applicazioni sono accessibili da una varietà di tecnologie attraverso un'interfaccia sottile come un web-browser (e.g., web-based email).

Piattaforma as a Service (PaaS)



Platform as a Service è la capacità di utilizzare nell'infrastruttura della nuvola applicazioni create dal cliente o acquisite da altri usando linguaggi di programmazione e strumenti supportati dal fornitore.

Infrastruttura as a Service (IaaS)



Infrastructure as a Service è la fornitura di servizi come descrizione, reperimento, accesso, scarto, immagazzinamento, networks, per cui il cliente può utilizzare e gestire software, che può includere operating systems e applications. Con tecniche di virtualizzazione l'infrastruttura è preparata in piccole unità che sono assegnate al cliente, come acqua o elettricità

Tipi diversi di Nuvole

Ci sono diversi modi di usare la Nuvola. Ciascuno ha il suo gruppo di standard utenti e livelli di servizio.



*Deployment
Models*

Tipi di Nuvole

- **Nuvola Pubblica:** L'infrastruttura è disponibile al pubblico in generale o a un'ampia area di attività e è la proprietà di un'organizzazione che vende servizi di Cloud. Per esempio Gmail.
- **Nuvola Comunitaria:** L'infrastruttura va condivisa da varie organizzazioni e dà supporto a una comunità specifica che condivide gli stessi bisogni (e.g., mandato, requisiti di sicurezza, politica e bisogni legali). Per esempio, una municipalità può usare una nuvola comunitaria.
- **Nuvola Privata:** L'infrastruttura è operata solo per una organizzazione. Può essere gestita dall'organizzazione o da una terza parte e può esistere in loco o al di fuori dell'organizzazione: è la più usata.
- **Nuvola Ibrida:** L'infrastruttura è una composizione di due o più nuvole (privata, comunitaria, o pubblica) che rimangono entità separate ma sono connesse da una tecnologia standard o proprietaria che permette la portabilità di dati e applicazioni). Per esempio, Google e Amazon.

Come viene usata la Nuvola?

- Backup
 - Collaborazione
 - Distribuzione
 - Archiviazione
-
- Storage di Email è l'uso numero 1.

Esempi di Nuvole

- Toyota usa Microsoft per fornire e mantenere traccia delle operazioni di assistenza stradale.
- Los Angeles Police Department usa Google per email. Ci sono problemi per via dei requisiti di sicurezza federali per l'archiviazione di documenti di polizia.
- General Services Administration in US usa IaaS e sta considerando PaaS e SaaS.
- NARA ha creato una direttiva per aiutare gli enti federali a rispettare i regolamenti per i documenti federali nell'uso della Nuvola.

Chi trae vantaggio dall'uso della Nuvola?

- Individui. In particolare liberi professionisti
- Un'organizzazione piccola o di misura media con un bilancio limitato per information technologies (IT) o un dipartimento IT che non è efficiente.
- Grandi enti che selezionano accuratamente i tipi di servizi

Benefici della Nuvola

Riduce i costi

- Non è necessario possedere hardware/software, quindi non ci sono costi iniziali.
- Si risparmiano costi energetici.
- Si riducono i costi per IT, perchè il personale non deve implementare e mantenere un'infrastruttura. Deve solo mantenere la Nuvola, non ciascun computer, come negli anni 60, quando c'erano tanti terminali e solo un computer (mainframe).
- Le risorse sono condivise, quindi si ottiene di più per meno e hardware/software e network sono molto migliori.

Benefici della Nuvola

Scalabilità

- Si può usare solo ciò di cui si ha bisogno e pagare solo per ciò che si usa.
- Si può mantenere il controllo di ciò che si usa.

Benefici della Nuvola

Affidabilità

- Sempre disponibile su richiesta, non importa quanto grande la richiesta sia
- Disponibile ovunque uno si trovi, semplicemente usando un browser.

Benefici della Nuvola

Sicurezza

- Il livello di sicurezza può essere molto alto e certamente più di quanto un'organizzazione possa permettersi, sia dal punto di vista fisico che da quello virtuale.
- La centralizzazione di tutte le risorse documentarie rende più facile proteggerle.

Benefici della Nuvola

Favorisce la collaborazione

- Permette più efficienza nella collaborazione perchè tutto il materiale è in formato compatibile quando si accede ad esso dal web browser.
- Rende possibile interagire avendo l'informazione accessibile da molte aree geografiche allo stesso tempo.
- Pensate a Google Docs.

Rischi della Nuvola

Impossibile pianificare i costi

- Dopo aver calcolato i costi di sottoscrizione, trasferimento, e attivazione, il resto è imprevedibile. Si può anche ricevere un conto per il costo della licenza che non ci si aspetta (Google/LA City).
- Variabilità dei costi a seconda della quantità di uso-non c'è un costo mensile prestabilito.

Rischi della Nuvola

L'affidabilità non è completa

- La Nuvola può andare in bancarotta, sparire o essere venduta. I documenti possono sparire per l'azione di un impiegato (e.g. Zodiac Island vs. CyberLynk)
- La Nuvola può perdere traccia di documenti e può non essere in grado di trovarli perchè a volte il backup fallisce.

Rischi della Nuvola

La sicurezza non è assoluta

- Bisogna affidarsi completamente al provider, ma si può pensare che ci informerà in caso di accessi non autorizzati, dell'esistenza di sotto-appaltatori, di hackers, etc.?
- I documenti possono trovarsi ovunque e possono essere spostati ogni momento senza che si sappia.
- La cifratura potrebbe non essere fatta quando i documenti sono in transito.
- I servers condivisi possono mischiare i documenti.
- Ex.: la FBI ha sequestrato un server del Dallas Data Center per i documenti di 1 persona-50 business usavano il server, e sono stati in grado di riavere accesso ai loro documenti solo dopo molti giorni.

Rischi della Nuvola

Mancanza di controllo

- Non si ha un vero controllo sul materiale nella Nuvola.
- Non si ha controllo su chi condivide lo stesso spazio nella Nuvola
- Le condizioni di servizio e la policy riguardante la privacy possono cambiare.
- Backup può essere fatto o no, senza seguire le direttive specifiche
- Documenti possono essere scartati secondo le direttive date o no
- Non si sa che succede quando l' hardware della Nuvola non è più utile
- Audit di solito non è permesso.

Rischi della Nuvola

Autenticità

- Non è possibile dimostrare una catena ininterrotta di custodia e di controllo sui documenti.
- Non è possibile dimostrare che il sistema di tenuta dei documenti usato dalla Nuvola è affidabile e conseguentemente lo sono anche i documenti
- Nell'assenza di una inferenza di autenticità basata sull'integrità dell'ambiente digitale, l'integrità dei documenti nella Nuvola non si può garantire altro che per verifica diretta su duplicati
- L'analisi dell'integrità del bitstream, dei dati, del computer, del sistema informativo, o del processo di creazione, riproduzione, mantenimento e selezione da parte di terze parti non è permessa o possibile.

Rischi della Nuvola

Privacy

- La European Union Data Protection Directive regola data privacy. Dati personali di residenti nell'Unione Europea non possono essere trattati in paesi che non hanno uguale protezione
- La Federal court del Canada ha affermato che un' organizzazione americana che trasferisce dati in Canada deve rispettare PIPEDA (next slide).

Rischi della Nuvola

Problemi Legali

- La locazione geografica dei documenti è di rilievo per la giurisdizione.
- I segreti commerciali non rimangono tali se si trovano nella Nuvola
- Il privilegio delle comunicazioni tra avvocato e cliente non si considera privilegio se la loro corrispondenza si trova nella Nuvola
- Secondo il Patriot Act, la FBI può ottenere mandati di sequestro per Section 215.
- HIPAA (US health information) e Gramm-Leach Bliley Act (US financial information) devono essere protetti con accordi specifici.

Rischi della Nuvola

E-discovery

- Non è facile isolare i documenti protetti da un ordine di e-discovery e fermare la loro distruzione
- Non è facile conservarli nel loro contesto, o produrne copie autentiche che possono essere verificate.
- Se ci sono copie multiple in più luoghi, non è facile sapere quali copie sono usate e da chi

Un buon contratto evita i rischi più gravi

Esempio di contratto

- Los Angeles fornirà a 34,000 impiegati Google Apps per email e per collaborare nella Nuvola.
- Il contratto richiede danni illimitati per violazione di sicurezza.
- Verifiche (audits) sono permesse.
- Documenti si trovano solo negli United States.
- Se il servizio è inattivo per più di 5 minuti al mese, c'è una penalità.
- I documenti sono cifrati e divisi in componenti digitali tenute in server diversi.
- Da entrambe le parti c'è impegno di non rivelare i dettagli dell'accordo.

Altre cose importanti

- Fare un valutazione dei rischi.
- Paragonare i fornitori e le loro politiche e procedure.
- Identificare tutte le responsabilità per i documenti.
- Prendere questa decisione con il coinvolgimento di tutti.
- Leggere le linee guida per la sicurezza nella Nuvola (Cloud Security Alliance Guidelines).
- Leggere le policies di paesi che hanno fatto quest'esperienza.

Se decidi in favore della Nuvola...

includi tra gli obblighi del service provider responsabilità di records management relative a

- tenuta e scarto
- conservazione
- sicurezza
- confidenzialità
- accesso
- proprietà dei dati
- autenticità e autenticazione
- conservazione per motivi giudiziari
- capacità di interazione col sistema documentario e
- verifica (audit).

Prepara un piano contingente per rimettersi da un disastro.

Se decidi contro la Nuvola...

scrivi una policy che articoli chiaramente le regole per i singoli impiegati, dicendo espressamente se fornitori esterni possono essere usati o no. Se la policy permette tale uso, deve richiedere lo sviluppo di linee guida per la conduzione degli affari dell'ente.

Il contenuto delle linee guida deve essere basato sulla cultura dell'ente, le aspettative dei suoi membri e utenti, il bisogno di avere accesso mobile e/o a distanza ai documenti dell'organizzazione, e il livello di sicurezza necessario.

Ultime riflessioni

- Google adesso ha un online operating system che si chiama Chrome OS-no hard drive è necessario.
- Stiamo tornando agli anni 60...un computer e tanti dummy terminals.
- Questa volta tuttavia non abbiamo a che fare con il mainframe dell'ente, ma con Google!

Acknowledgement & Reference

Thank you to UBC MAS students Jayson Kennedy and Liza Sneider for their research and the cloud-slides design.

Per sviluppi ulteriori tenetevi aggiornati su www.interpares.org & www.digitalrecordsforensics.org