MAKING DIGITAL DATA TRUSTWORTHY IN THE CLOUD

Cian Kinsella – CEO, Digiprove
cian.kinsella@digiprove.com

Malaga 7th June 2013
My Background

- Been developing software since 1972
- Commercial and Freelance
- Co-founder of 3 Software Product Companies
- Have had many different responsibilities
- Still developing software, though…

- Digiprove’s technology adds Trust to Digital data
What is the core issue?

- To be able to provide absolute proof of the integrity, authenticity and provenance of digital content when required. (detect any tamper).

- The Cloud adds further uncertainty:
  - Who has access to your data & communications?
  - How many times is it replicated?
  - Where is your data (and copies thereof)?
  - Who has access to your data

- You are placing Trust in an organisation, i.e. people, not technology
When is Trust Critical?

- Digital evidence
- In Court
- For Regulator/Compliance
- For Business decisions
- eDiscovery
- IPR/Copyright
- Retention compliance
- Digital Signing
- Cloud Data and Apps
- Social Media
- Financial Records
Vast majority of content originates digitally:
- Application software
- Email
- Microsoft Word
- Digital cameras
- Pro-tools Sound Production software
- Final Cut Video Production software
- Phones/Tablets: VOIP / Messaging (BYOD)

Most of the rest ends up in digital domain
- Scanned papers and images
- OCR

Usually no way to trace provenance
Characteristics of Digital Data

- It is intended by design to be manipulated
- It can be changed:
  - Text files
  - Microsoft Word and other documents
  - Email archives
  - PDFs
  - Sound recordings
  - Images
  - Database records & Logs
- It is communicated
- It can move location or be replicated
Questions to Consider

- Criticality of Data Integrity
  - What is consequence of integrity failure
- Who needs to trust it?
  - Internal – *basis for operational decisions*
  - Non-exec Directors / Boards – *basis for strategic decisions*
  - Auditors or Regulators – *basis for compliance*
  - External Stakeholders – *basis for trust & confidence*
  - Public/Citizen – *basis for trust and commitment*
  - Journalists – *for attributable quotes*
  - Courts – *evidence*
# How trustworthy does it need to be?

- **Depends on:**
  - Criticality of business
  - Criticality of data

<table>
<thead>
<tr>
<th>I trust our access controls:</th>
<th>Completely</th>
<th>Up to a point</th>
<th>Not Sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation to tamper or hack:</td>
<td>Little or none</td>
<td>Certain parties</td>
<td>Strong interest</td>
</tr>
<tr>
<td>Potential Consequences of undetected tamper:</td>
<td>Negligible</td>
<td>Unknown</td>
<td>Potentially Serious</td>
</tr>
<tr>
<td>Can some of the data become evidence?:</td>
<td>Very unlikely</td>
<td>Possibly</td>
<td>Probably</td>
</tr>
</tbody>
</table>
Regulators
- Often insist on “wet” signatures
- Submissions must be in paper or fax
- Retention of Original paper records

Legal documents, contracts etc.

Certificates, Diplomas

E-Discovery Law
- Spoliation allegations
- Preservation Obligations

Cloud generally

Low Trust should create demand for Digital Data Trust Solutions
Misplaced Trust

- Documents
  - PDFs
  - Scanned Images
- HTTPS web-pages
- (Some) Financial Audit Trails
- Cloud data
- Tweets and other social media
Do people trust Digital Content?
Requirements of a Data Trust Solution

- Have certainty that any content has not changed
  - Since creation date
  - Since authorised amendment
    (must include version history)

- Automatically detect if content changed outside permitted channels
  - Threat detection
  - Early alarm for attempted fraud
  - Early alarm for data corruption

Give visible assurance  
Raise Alarm
Requirements of a Data Trust Solution

- Establish Provenance
  - Who created?
  - Who witnessed?
  - Is signed / contains signatures?
  - Version History

- Other Metadata:
  - Data Type
  - Functional / How created (email, MS Word, camera etc.)
  - Where created – Location
  - Identity (like fingerprint)
<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Digital Signatures</strong></td>
<td>Individual signs content - Binds content to individual and any change in content is detectable.</td>
<td>- Very Secure</td>
<td>- Dependent on “web of trust”, independent certification of identities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Independent proof of identity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Can be extended with independent timestamp</td>
<td></td>
</tr>
<tr>
<td><strong>Digital Fingerprinting</strong></td>
<td>Use one-way “hash” algorithm to create unique fingerprint of each file/blob</td>
<td>- Easy to implement</td>
<td>- No intrinsic security</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Computationally light</td>
<td>- No intrinsic timestamp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Reliable at detecting file corruptions</td>
<td>- No key</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- No provenance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Certain algorithms have flaws (e.g. SHA1, MD5)</td>
</tr>
<tr>
<td><strong>Checksum</strong></td>
<td>Simple algorithm to identify content</td>
<td>- Very Quick</td>
<td>- As for digital fingerprinting</td>
</tr>
</tbody>
</table>
Overview

- Database
- Tamper-proof Audit Trail
- Proof Engine
- Database
- Web Services API

Packaged Solutions
Web Service
Integrated Systems
<table>
<thead>
<tr>
<th>STEP:</th>
<th>Function</th>
<th>Notes / Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital Fingerprinting</td>
<td>Creates unique id/recognition key for each content blob</td>
<td>Uses strong algorithm (e.g. SHA256)</td>
</tr>
<tr>
<td>Secure Fingerprinting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Create timestamp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certify metadata</td>
<td>Establishes provenance</td>
<td>- User name, content abstract, version, GPS co-ordinates</td>
</tr>
<tr>
<td>Create transparent tamper-evident audit trail</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent element</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Automatic Verification       | Gives visible assurance Detect / Alert tampers or corruptions           | - On request  
- Periodic  
- Whenever content is accessed |

**Result:** Creates proof of content & time-stamp and provides independent certification of this. It is impossible (even for Cert. Service Provider) to issue a back-dated certificate. Process does not rely on trust in any person or organisation.
Some Uses of this Technology

- Tamper-evident audit trails (regulated industries)
- Establish provenance & timestamp of legal documents
- Verify integrity of web pages as they are served
- Handwritten signatures on digital documents
- Meet e-Commerce legal requirements on retaining digital data
- Tamper-evident email archives
- Ediscovery lock-down
Applying to the Cloud

- Requires cloud & ground deployment of technology

- Multiple Points of certification:
  - For content created locally, evidenciate before upload
  - For content created in Cloud, evidentiate immediately
  - For externally-sourced content, evidentiate immediately

- On Content presentation, verify locally
We are Interested in

- Ideas about Drivers for Recognition of Problem
- Raising our Profile
- Improving Our Core Offering
- New Deployments of our technology/service
- Academic Opinion and Input
- Other Forms of Collaboration
Thanks

cian.kinsella@digiprove.com