A TALE OF TWO TRACES – DIPLOMATICS AND FORENSICS

Fred Cohen

▶ **To cite this version:**

Fred Cohen. A TALE OF TWO TRACES – DIPLOMATICS AND FORENSICS. 11th IFIP International Conference on Digital Forensics (DF), Jan 2015, Orlando, FL, United States. pp.3-27, 10.1007/978-3-319-24123-4_1 . hal-01449067

HAL Id: hal-01449067

https://inria.hal.science/hal-01449067

Submitted on 30 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Chapter 1

A TALE OF TWO TRACES – DIPLOMATICS AND FORENSICS

Fred Cohen

Abstract This chapter focuses on two examples of legal matters involving archived data, one is a digital archive of born-analog data and the other is a digital archive of born-digital data. Their resolution is explained, and along the way, several challenges and issues related to digital archives, the transition from classical diplomatics to modern diplomatics, digital forensics in the light of current record-keeping systems, and related facts and supporting data points are explored.

Keywords: Questioned documents, archival science, diplomatics, digital forensics

1. Introduction

In October 2013, a house was up for sale and, during the closing period, a dispute arose about square footage that could have stopped the sale or substantially changed the terms of the sale. The issue was ultimately resolved by examining documents from city and county digital archives of born-analog data. This is referred to as Case 1.

In an unrelated legal dispute, two ex-partners were in stark disagreement about whether one of the partners took on a competitive business while the partnership was still operating, and thus acted in bad faith. The issue was brought to light because of the content of a web page depicted on the Wayback Machine of `archive.org`, which is asserted to be an archive of Internet websites over historical time frames (mid 1990s). This is referred to as Case 2.

The issues in these cases are not new. In fact, they are very old and, by looking back to the history of archives, diplomatics and related aspects of records management, insight may be gained about the future of these fields.

1.1 Causality as a Foundation of Science

Foundational to science is the notion of causality. Cause (C) acts through ($\rightarrow$) mechanisms (m) to produce effects (E): $C \rightarrow^{m} E$. This is the basic assumption of science as a whole as well as scientific evidence in the narrow sense of legally admissible evidence. Noteworthy are the notions that correlation is not causality and that effect does not imply cause [2]. In order to create a scientific hypothesis about a legal matter, a hypothesis of a mechanism by which cause produced effect must be formed, with the effect being the traces found and the cause being a hypothesized act of interest to the case.

1.2 Diplomatics

The field of diplomatics is often reported as being founded in 1681, when the renowned French philologist Mabillon [6] published the results of an analysis of approximately 200 documents divided into categories and examined with regard to material, ink, language, script, punctuation, abbreviations, formulas, subscripts, seals, special signs, chancery notes and so on. He created descriptions to allow the detection of forgeries and identified ground truth based on the recurrence of intrinsic and extrinsic elements in documents from the same time and place [3]. In modern terminology, and taking some liberties in usage, he used redundancy to test for consistency. Note that this approach is based on correlation, but causality was also present in the form of known chanceries or scriptoria traditions (cause) and capabilities of scribes over the ages (mechanism). In addition, and perhaps more vitally, no ground truth was available for much of this effort because the documents were too old for eyewitnesses and the documentary evidence supporting the claims was in question along with the claims themselves.

Diplomatics provides much of the basis for the admissibility of evidence and the establishment of criteria for evaluating evidence. The field and its principles of application are still used for questioned document analysis. By extension, its principles and many of its methods are in use or have analogous use in digital forensics. The modern and historic reconstruction of causes acting through mechanisms to produce effects forms an experimental basis for diplomatics, except, of course, that accelerated aging and similar methods are approximations or models.

1.3 Archival Science and Public Records

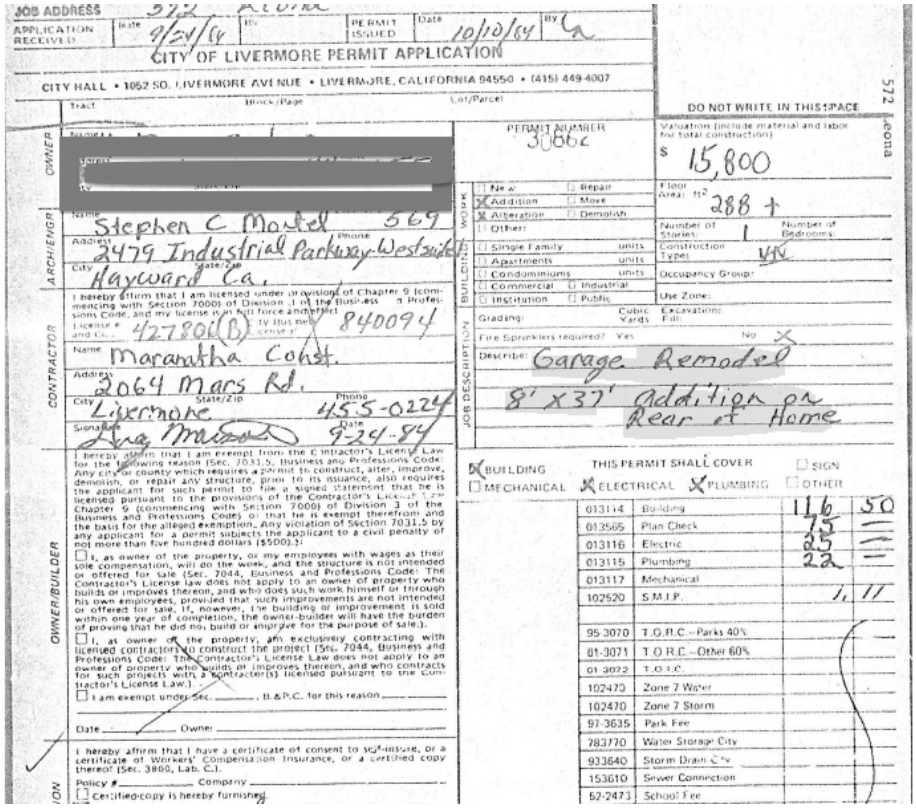
The field of archival science emerged over time due to the need to keep reliable public records (e.g., land ownership documents). Ancient record-

keeping systems date back as far as history and indeed much of history is based on the records collected and retained by archivists at different administrative bodies. In the legal system, public records are generally admissible for the truth of what they self-indicate and are presumed to be trustworthy (i.e., reliable – a true statement of fact), authentic (not corrupted or tampered with) and accurate (truthful, exact, precise or complete) in the legal system, when they are properly introduced and marked with the appropriate seals, signatures and/or special signs by the legal entities that produced them. While diplomatic analysis may be used to attempt to refute these records and/or to rehabilitate them after attempted refutation, they are generally trusted, and built and maintained in a manner that reasonably justifies the trust. This is, at least, their nature in the analog records space.

The analytical approach is based on a set of redundant acts by independent trusted actors forming a set of archival fonds associated with different archival units, programs or institutions. Fonds are aggregations of documents that originate from the same source. The records and fonds include explicit formal elements designed to provide assurance that the records are authentic, accurate and reliable over their life. This is undertaken by providing a chain of custody in a transparent system of record keeping through redundant information that associate actions related to a record with the record in the context of the fonds, and the fonds in the context of the archives. This is sometimes called the archival bond, which is the relation between a document and the previous and subsequent documents produced in the course of a business matter. Paper records are often annotated over time, marked with stamps, altered by updates and changes, and so on.

Figure 1 shows a (redacted) document provided in response to a request for a legal record related to ownership. The document has a series of writings added over time. The record, in this case, is a dynamic document that has been updated to reflect the officially-authorized changes, as shown by the seals of the officials who carried out the actions. As the record evolves over time, it is retained in a chain of custody and supported by the fonds in which it resides, which reflects the dates of access and other related information. As a legal document, this is considered proof of the facts contained in the document and is inherently regarded as reliable and accurate based on the source, and authentic because it could be proffered to a court as a proper form with the seals intact.

This document might have been moved into an archival repository for a period of time and returned to active use later. Such movements would be annotated on the document and/or in the fonds in which it resided over time. The redundant information from the various sources


CITY OF LIVERMORE PERMIT APPLICATION

CITY HALL • 1052 SO. LIVERMORE AVENUE • LIVERMORE, CALIFORNIA 94550 • (415) 449-4007

JOB ADDRESS 572 Peony

APPLICATION RECEIVED 9/24/84

PERMIT ISSUED 10/10/84

PERMIT NUMBER 30862

DO NOT WRITE IN THIS SPACE

Valuation (include material and labor for total construction) $ 15,800

OWNER: Name [redacted]

CONTRACTOR: Name Stephen C Martel 569 Phone; Address 2479 Industrial Parkway West; City Hayward Ca.

I hereby affirm that I am licensed under provisions of Chapter 9 (commencing with Section 7000) of Division 3 of the Business and Professions Code, and my license is in full force and effect. License # 427806 (B) Lic. Class X 840094

Name Maranatha Const.

Address 2064 Mars Rd. Phone 455-0224

City Livermore State/Zip

Signature Ana Marsh Date 9-24-84

I hereby affirm that I am exempt from the Contractor's License Law for the following reason (Sec. 7031.5, Business and Professions Code: Any city or county which requires a permit to construct, alter, improve, demolish, or repair any structure, prior to its issuance, also requires the applicant for such permit to file a signed statement that he is licensed pursuant to the provisions of the Contractor's License Law (Chapter 9 (commencing with Section 7000) of Division 3 of the Business and Professions Code) or that he is exempt therefrom and the basis for the alleged exemption. Any violation of Section 7031.5 by any applicant for a permit subjects the applicant to a civil penalty of not more than five hundred dollars ($500).):

☐ I, as owner of the property, or my employees with wages as their sole compensation, will do the work, and the structure is not intended or offered for sale (Sec. 7044, Business and Professions Code: The Contractor's License law does not apply to an owner of property who builds or improves thereon, and who does such work himself or through his own employees, provided that such improvements are not intended or offered for sale. If, however, the building or improvement is sold within one year of completion, the owner-builder will have the burden of proving that he did not build or improve for the purpose of sale.).

☐ I, as owner of the property, am exclusively contracting with licensed contractors to construct the project (Sec. 7044, Business and Professions Code: The Contractor's License Law does not apply to an owner of property who builds or improves thereon, and who contracts for such projects with a contractor(s) licensed pursuant to the Contractor's License Law.).

☐ I am exempt under Sec. _____, B.&P.C. for this reason _____

Date _____ Owner _____

I hereby affirm that I have a certificate of consent to self-insure, or a certificate of Workers' Compensation Insurance, or a certified copy thereof (Sec. 3800, Lab. C.).

Policy # _____ Company _____

☐ Certified copy is hereby furnished.

WORK: ☐ New ☐ Repair ☒ Addition ☐ Move ☒ Alteration ☐ Demolish ☐ Other:

BUILDING: ☐ Single Family units ☐ Apartments units ☐ Condominiums units ☐ Commercial ☐ Industrial ☐ Institution ☐ Public

Floor Area: ft² 288 +

Number of Stories: 1 Number of Bedrooms:

Construction Type: VN

Occupancy Group:

Use Zone:

Grading: Cubic Yards Excavation: Fill:

Fire Sprinklers required? Yes No X

JOB DESCRIPTION: Describe: Garage Remodel 8' x 37' Addition on Rear of Home

THIS PERMIT SHALL COVER: ☒ BUILDING ☐ MECHANICAL ☒ ELECTRICAL ☒ PLUMBING ☐ SIGN ☐ OTHER

013114	Building	116	50
013565	Plan Check	75	
013116	Electric	25	
013115	Plumbing	22	
013117	Mechanical		
102520	S.M.I.P.	1.	11
95-3070	T.O.R.C.–Parks 40%		
01-3071	T.O.R.C.–Other 60%		
01-3022	T.O.R.C.		
102470	Zone 7 Water		
102470	Zone 7 Storm		
97-3635	Park Fee		
283770	Water Storage City		
933640	Storm Drain City		
153610	Sewer Connection		
52-2473	School Fee		

Figure 1. Case 1 example.

(e.g., transmission from use to the archives and back as signed by parties on each side, placement in the fonds in sequence over time, markings on cover sheets and/or envelopes and/or the documents themselves) can be examined by a diplomatics expert in the context of the methods used by the record-keeping system to make a determination of authenticity or to refute or challenge the presumption of authenticity based on a lack of adequate evidence of (or evidence of inadequate) custody and control.

1.4 Digital Records

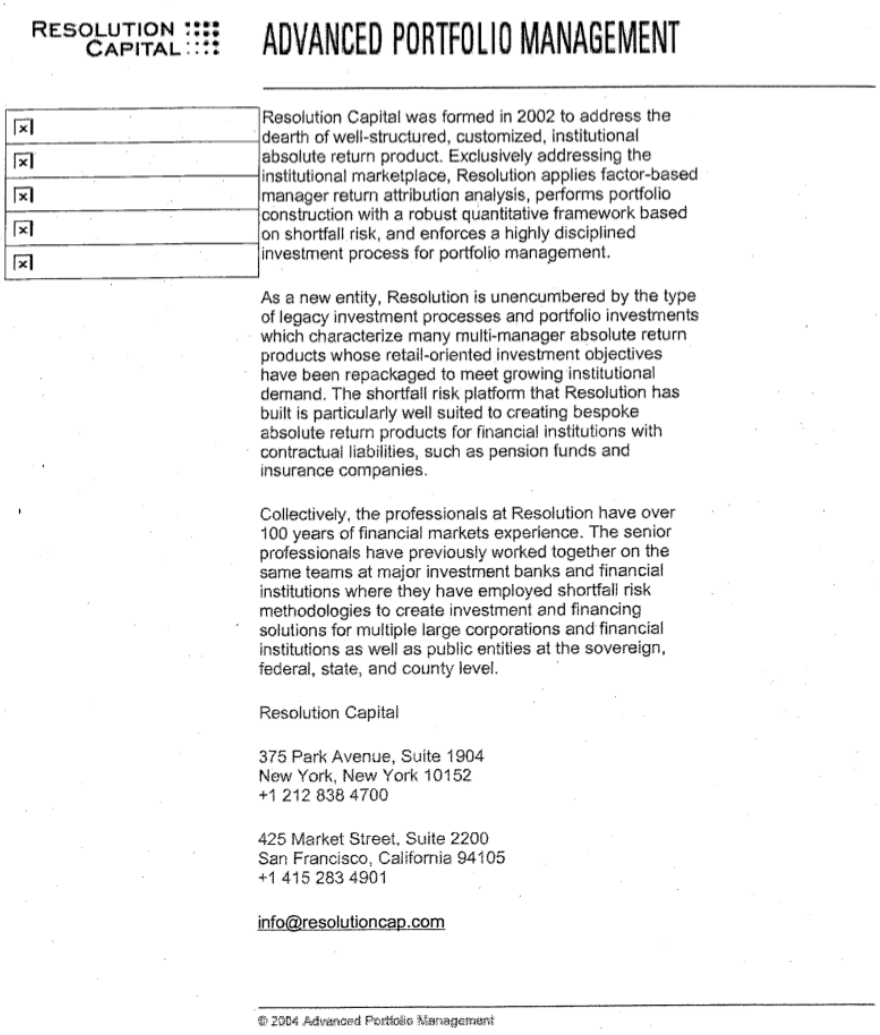
Many of the methods that made analog records reliable over time have not been translated into the new forms of record keeping in the digital records space. For example, the processes involved in property purchase often required a government-authorized actuary to certify that an individual identified by a government document signed a document in the presence of the notary. But increasingly, a digital system allows a

digital signature that is not even created by the individual's own hand. Instead, the self-identified individual agrees electronically over the Internet to adopt a signature form for use in signing documents. The documents are sometimes incorrectly presented (i.e., with incorrect data fields), with the results produced as digital documents reflecting different (and in some cases corrected) content than what was actually presented for signature.

Such record-making and record-keeping systems are potentially very problematic in legal terms, but they are not often challenged. They do not guarantee that what is agreed to is what is presented. They often include and present false information and change it after agreement. They do not provide a copy in the form of the original (identical to the original in all respects but issued after the original), an imitative copy (reproduction of both the form and content of a record) or even a simple copy (transcription of the record content) to the signatories. Moreover, they often do not incorporate an actual signature that is traceable to an individual and that is demonstrably different from other adopted signatures. Documents may be presented differently from how they appeared before (i.e., as a pseudo-original with the pretense of originality), even when and if they are ultimately presented and/or submitted in court as authentic, reliable and accurate.

Nevertheless, these pseudo-original documents are then declared as public records, and from that point forward, they are recognized, treated and presumed as authentic renditions of contracts. They become part of corrupt and inauthentic digital records and eventually make their way into the archived and permanent records of societies. The metadata associated with these records often lacks the fields required by record management systems and archives (if present, they may be incorrect), the mechanisms are not transparent and they not available to the parties to the contracts.

Figure 2 shows an example of a presentation made as part of the collection of potential evidence in Case 2. This depiction of a digital record reflects what, in some jurisdictions, is legally admissible as an archival document and may be given the presumption of authenticity, reliability and accuracy. In particular, note that "Resolution Capital" appears with "Advanced Portfolio Management" on the page. This is a depiction that was saved from a screen image of what was seen when one of the parties gathered what he believed to be evidence in support of his case.



RESOLUTION CAPITAL

ADVANCED PORTFOLIO MANAGEMENT

- ☒
- ☒
- ☒
- ☒
- ☒

Resolution Capital was formed in 2002 to address the dearth of well-structured, customized, institutional absolute return product. Exclusively addressing the institutional marketplace, Resolution applies factor-based manager return attribution analysis, performs portfolio construction with a robust quantitative framework based on shortfall risk, and enforces a highly disciplined investment process for portfolio management.

As a new entity, Resolution is unencumbered by the type of legacy investment processes and portfolio investments which characterize many multi-manager absolute return products whose retail-oriented investment objectives have been repackaged to meet growing institutional demand. The shortfall risk platform that Resolution has built is particularly well suited to creating bespoke absolute return products for financial institutions with contractual liabilities, such as pension funds and insurance companies.

Collectively, the professionals at Resolution have over 100 years of financial markets experience. The senior professionals have previously worked together on the same teams at major investment banks and financial institutions where they have employed shortfall risk methodologies to create investment and financing solutions for multiple large corporations and financial institutions as well as public entities at the sovereign, federal, state, and county level.

Resolution Capital

375 Park Avenue, Suite 1904
New York, New York 10152
+1 212 838 4700

425 Market Street, Suite 2200
San Francisco, California 94105
+1 415 283 4901

info@resolutioncap.com

© 2004 Advanced Portfolio Management

Figure 2. Case 2 example.

1.5 Digital Diplomatics

As a field, digital diplomatics currently takes on two meanings. One meaning is the use of digital computing methods to support classical diplomatics. For example, digital methods may be used for word and phrase analysis to detect differences in scribes and to track scribe usage of terms over time in order to date documents more accurately than was feasible using manual techniques. The other meaning is the use of diplomatic methods to authenticate digital documents, which is the

meaning of interest in this work. The following quote amply conveys this meaning:

> Georges Tessier: *"On peut donc avancer que la critique diplomatique est née dans le prétoire ou sur le forum à l'occasion de débats judiciaires ou de controverses politiques ou religieuses, quand le nœud du litige ou de la polémique était constitué par un document ou une série de documents contestés." Cette citation est tirée de L'Histoire et ses méthodes (La Pléiade, 1961) dont Georges Tessier a signé le chapitre Diplomatique* [1].
>
> Google Translation: *"It can be argued that the diplomatic criticism is born in court or forum on the occasion of judicial proceedings, or political or religious controversy when the crux of the dispute or controversy consisted of a document or series of documents in dispute." This quote is from The History and Methods (The Pleiades, 1961) where Georges Tessier signed the chapter Diplomatics.*

In this context, the relevance of digital diplomatics may be reasonably explored relative to digital forensics.

1.6 Forensic Science

Forensic science is often cited as coming to clarity through the work of Locard [5]. Locard identified that, when objects come into contact, they each leave parts (traces) of themselves on each other. The mechanism of objects coming into contact and leaving traces is called "transfer" and, thus, we have the scientific notion of causality fulfilled by contact (cause) acting through transfer (mechanism) to produce effect (traces). Traces may be humanly observable (e.g., chunks of rock or mud) or latent (observable only through the use of tools as in dust and microscopic particles). Locard undertook studies showing layers (e.g., of mud or dust) indicating sequences of places visited (e.g., layers of mud on shoes) when a person transited a city, and associating the transferred traces to locations based on unique properties (e.g., strands of a particular wool from the only factory that produced it in the city).

As forensic science has moved forward, many methods have been developed based on the concept of transfer. Methods, such as tool mark analysis, have also been developed from the earlier diplomatics area.

1.7 Digital Forensics

As digital systems came into widespread use, the legal system had to deal with evidence in the form of traces of activities within and between digital systems. The study of digital traces relative to the legal system was identified as digital forensics. However, as in the case of digital diplomatics, another meaning is used. Specifically, digital forensics is also used to describe activities associated with the investigations of eve-

nts in the digital arena, a much broader field that is closely related to detection and response regimens in computer security.

As a fundamental notion, it has recently been recognized that digital evidence is still trace evidence, is almost always latent in nature and is not transfer evidence. Rather than being transferred in the sense of Locard, digital evidence is formed from traces produced by the mechanisms of operating digital systems, typically as stored states from finite state machines that transform state and input into the next state and output.

Using the term of art from diplomatics, "digital traces" are produced by transmission rather than transfer. In the digital arena, transmission producing traces is typically also transmission in the sense of electromagnetic, optical, sonic or other emission and reception of signals. Specifically, events in one context produce signals that are sensed in another context and memorialized in the form of optical patterns, configurations of particles, magnetic orientation or whatever traces are supported by the transmission or fixation media.

Additionally, in the digital arena, the latent nature of evidence is such that a copy in the form of an original almost never occurs. Instead, a sequence of bits represented in a fixed form in/on a medium may be reproduced (at the bit level), while presentation in a human-readable form is normally an imperfect reflection of the original documentary form. For example, when the information associated with a digital record (e.g., financial transaction) is originated, the form of entry (e.g., web-based entry of a purchase form) is typically very different from the form of transmission (e.g., series of datagrams sent over the Internet as waveforms in transmission media), storage (e.g., sequence of bits in a database storage area of a disk drive or in a positive feedback loop in active memory) and presentation (e.g., line item in a bank statement or entry in a spreadsheet downloaded by an accountant from the financial institution and used for tax purposes). These notions are not widely recognized or stated in digital forensics today, even though they are always present.

The notions of authenticity, accuracy and reliability are always at issue in the digital forensics arena relative to the classical notions of documentary form. The notion that using methods such as cryptographic checksums to verify the lack of alteration of a bit sequence does not even begin to address the issues of authenticity of a record in presentation and reliability in the sense of relationship to original writing or any sort of ground truth. Causality works differently.

In the following sections, distinctions, if they exist, are identified using classical diplomatics, classical forensics, digital forensics and digital diplomatics in the realm of questioned document examination. In par-

allel with this exposition, the case studies are examined in detail. These cases are not large or important on their own, but instead reflect the many everyday legal issues that naturally occur in human interactions and sit at the heart of how people interact with the legal issues faced in these areas. Finally, the cases are resolved and the conclusions are presented.

This work views questioned digital document examination as the fusion of diplomatics and forensics. It may reasonably be called digital diplomatics and/or questioned digital documents without reasonable differentiation. Indeed, the reconciliation of diplomatics and forensics in this arena is a historic merger and unification of the individual concepts and fields of study.

2. Digital Diplomatics and Forensics

This section presents the background of the two cases and discusses document admission and related information regarding records.

2.1 Case 1 Background

Case 1 involved a dispute over the square footage of a house. The seller claimed the square footage specified when the house was purchased and as reflected in the taxes paid over the duration of ownership and for some unknown period prior to the purchase. The buyer, a civil engineer, upon assessment, noted a different square footage in the inspector's report and proceeded to do an independent measurement, which produced a third square footage result.

If left unsettled or settled in various ways, this situation could lead to charges of fraud, damage to reputation, tens of thousands of dollars difference in the sales price, retroactive tax readjustment and delayed-closure or non-closure of the sale. None of these situations were in the interest of the parties and the settlement of the dispute rested on documentary evidence in the form of records from various sources, including the prior sale documentation, tax documentation and city and county records related to remodeling, permits and inspections. The measurements themselves were also at issue because different measurements (e.g., inside dimensions, outside dimensions, livable space and permitted use areas) are based on different definitions in different overlaid jurisdictions (e.g., taxation, county building and city building).

2.2 Case 2 Background

Case 2 involved a dispute between ex-partners in a financial venture. The business failed and each partner went his own way seeking to start

a new financial venture, with the ownership of the domain name remaining with one of the partners. Several years later, upon viewing what was believed to be an image of the prior website using the Wayback Machine at `archive.org`, the partner who did not retain control of the website was unhappy to discover that, according to the displayed content, resources belonging to the previous business were used to advertise the new business prior to the termination of the partnership. This led to charges of misappropriation of resources, customers and business from the partnership and failure to faithfully fulfill fiduciary and other duties as a partner.

In this case, the dispute was based on the form and appearance of the document (i.e., the depicted website) as seen in the archival site. The depiction was clear as it could be. A date selected by the user and indicated in the URL at the top of the web browser page showed content from the prior business simultaneously displayed in a single web page with material from the new business. If the depiction reflected reality, there would be little question that a case could be made by the complainant. The only case that could reasonably be made by the accused party was to question the document presented by the archive.

2.3 Admitting the Documents

While the subtleties of an Internet archive versus other types of archives and the question of how to resolve seemingly inconsistent information from different official records may be vitally important to the issues at hand in the two cases, there seems to be no question that the documents would normally be admitted in legal proceedings.

The Wayback Machine is an automated storage system while an archive preserves documents. Preservation is a process in which an archivist identifies, authenticates, protects, describes, builds retrieval systems, provides access to and acts to protect the material being archived. The term Internet Archive in the context of the Wayback Machine is a misuse of the term of art, "archive." Of course, humans have trusted archives for centuries and the individuals who operate `archive.org` have demonstrated excellent marketing skills in using the term.

The legal status of government documents is normally that they are admitted and presumed reliable, authentic and accurate. Thus, the documents supplied in Case 1 operate under this legal presumption.

The Internet Archive is a bit more nebulous in that it is a website operated by a non-profit (i.e., public interest) corporation, much like a museum or other archive. However, this is not what the Wayback Machine is. The Wayback Machine is not like a museum or an archive

because there is no curation or assurance of protection and permanent authenticity from the moment of acquisition.

Ancient documents ("ancient" is a term of art) are usually admitted under the presumption that they were not forged in advance in anticipation of some future litigation that could not have been anticipated by an archivist. Disregarding the question of how old is old enough, a strong case can be made that the Wayback Machine was not operating intentionally to create a forgery and no claim was asserted that the information it stored was altered in any nefarious manner. The presumption for such documents is, *de facto*, also that of being reliable, authentic and accurate, even if this is not based on the same legal or technical footing as public records. Here lies the rub.

Archives that hold public records are normally designed by archivists and record-keeping specialists to reasonably assure trustworthiness. In the paper world, a chain of custody is established by independent and redundant trusted parties. These parties attest to signatures (i.e., seals) that become part of a document as it moves from party to party for signatures; take custody of the document and retain it in a secure location; track it in the fonds through numbering, ordering, cross referencing and other related processes; indicate how, when, from whom and other characteristics as the document is ingested, stored, moved, retrieved, transmitted, examined, copied, migrated and so on; and generally keep records of their activities in a transparent manner and make them available for examination.

This all depends on trust in the custodian as an entity who has not altered the records and has not allowed anyone else to do so. This latter requirement – of not allowing others to alter records – is problematic in the Internet because it was not designed for this purpose.

Examination can detect inconsistencies in and between records and fonds and this supports trusting (or challenging) the trustworthiness of the records. But this is not the case for depictions presented by the Wayback Machine. Collections are made at seemingly arbitrary times from subsets of automatically-selected websites. The components that manifest a visualized web page are collected at different times, stored with only a single reference to a collection date and are not attributed or tracked in the many ways that an archive is managed. The Wayback Machine is not a system of records as much as an amateur collection, but it is sometimes treated as if it is a traditional archive.

In the digital world, alteration can happen unintentionally or intentionally, the protection mechanisms of the Wayback Machine are not transparent and the adequacy of the machine and its mechanisms has not been established by a rigorous scientific process. The Wayback Ma-

chine does not apparently follow the rigors of archival science or records management and, thus, it should be inherently obvious to an expert in the field that it does not have the same status as public records or archives with regard to maintaining and operating within the standards of care. This is also the case for many other Internet-based sites that assert archival or records status, and this is one of the important reasons why a science needs to be developed in this regard and why diplomatics must be developed as a field to question the stored documents.

The situation is further complicated by the fact that the mechanisms of the Wayback Machine change over time, are not externally documented or transparent, and do not follow widely accepted archival principles. In fact, after the findings discussed here were made public, the Wayback Machine was changed with minimal notice and little apparent transparency. Thus, no external repeatability – a basic requirement of a scientific field – exists for the changes and performing an accurate reconstruction is problematic.

In the legal realm, depending on the bent of the judge and case precedence that may reflect previous mechanisms or poorly tested assertions, depictions that are not accurate, reliable or authentic may ultimately be admitted, presumed trustworthy and treated with a weight similar to that of records maintained by government bodies or real archives.

2.4 Related Information on Records

While this situation may appear problematic, the reality of digital records is, in general, quite tenuous compared with other forms of records. Some of the following examples from personal correspondence are informative and some may be recognized as having been reported in the popular media:

- A global non-government agency indicated that, in some cases, it holds records where 80% are of unknown type. When asked whether assistance was desired in trace-typing them, the response was that, while it must maintain the records, it had neither the resources nor the desire to type them. The obligation of the non-government agency thus stops at proper retention.

- Migration of records from system to system over time is necessary to retain the utility of the records because digital systems fail and older systems are no longer available. Moreover, newer systems rarely support all the mechanisms of the older systems.

  Conversion is thus part of record migration. The result is that a migrated record is often, at best, an imitative copy, sometimes a

simple copy and sometimes a pseudo-original copy. Records may never be viewed as they were initially formed and the loss of utility during conversions is not uncommon. Key parts of the migration problem faced in digital archiving are creating the mechanisms needed to produce copies of the records in one form or another and identifying and recording the nature of the changes associated with conversions and non-original mechanisms in terms of what is depicted and what is no longer depicted.

- Forensic archives of legal matters often contain large volumes of data in unusable form because they have not been migrated or converted and the original mechanisms and/or contexts may no longer exist to meaningfully reconstruct or use them. Given that legal appeals processes may come many years later, this evidence may no longer be viable if re-examination or retrial are required.

- Many modern devices and systems are complex and lack transparency to the point that their operational mechanisms are not discernible. Furthermore, automated patching often results in situations where the exact versions are not available and may be difficult or impossible to accurately reconstruct. Thus, establishing causality in reconstruction may not be accurate, and the task of reconstruction becomes very complex in this light.

- Some governments have admitted the use of covert methods to alter the apparent operation of mechanisms, making them act in ways unknown even to their manufacturers. This potentially shakes to the core the idea that archives accurately reflect the reality of what took place. The competition to rewrite history and current affairs in the digital realm would seem to present problems for the trustworthiness of digital records for legal purposes.

- Some nations and other similar entities use digital records to reflect their operations, including without limit, the original writing and official codifications of their laws and legislative history. This includes scanning documents that become part and parcel of the legal constructs of their societies as well as born-digital records.

  Recent revelations indicate that scanning devices no longer simply make representations of pixelated color values in digital form with known accuracy and precision limits. Instead, some of the devices read the content of documents and rewrite them, sometimes replacing digits, words, spelling and other elements of content with corrected versions. The very laws codified in statute and used to

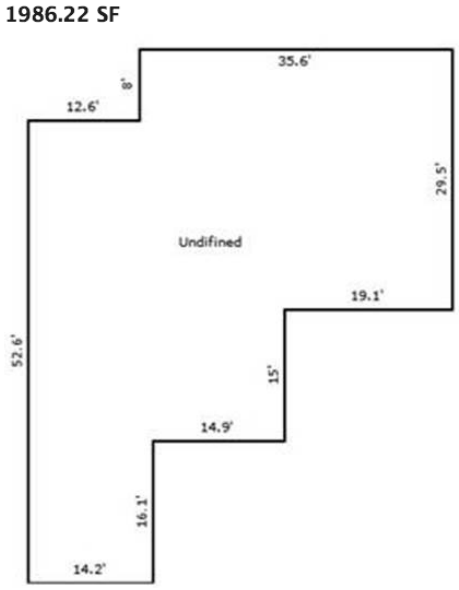


Figure 3. Inspector's version of the house.

make decisions about peoples' lives cannot be relied on to accurately reflect the laws as passed, and documents such as financial records may not be accurately recorded for future use such as taxation and levying fines.

It appears that there is a desperate need for a science of questioned digital documents. This field might be reasonably called digital diplomatics, named after the existing diplomatics field that was created for the same purpose in the non-digital realm. Part and parcel of diplomatics was the development of archival science and records management. The same path would seem to be a reasonable trajectory for digital diplomatics leading to and helping guide digital records management, digital archival science, digital records forensics [4] and the broader field of digital forensics.

3. Case Resolution

This section discusses how the cases were resolved.

3.1 Case 1

Figure 3 shows the drawing of the property produced after an analysis by a building inspector. The corresponding square footage was calcu-

Figure 4. Overhead image of the house.

lated as 1,986.22 sq. ft. This is obviously at odds with the overhead image of the house shown in Figure 4. The inspector was looking at the records of livable interior space in city permits, the full details of which were no longer available from the relevant time frames.

The overhead image in Figure 4 is a Google Maps aerial of the house at about the time of sale. Note the substantial difference between the shape in the inspection report and the actual shape in the overhead image. The dispute at this point was whether and to what extent a remodeling of the former garage was properly accounted for in the calculation.

The issue was ultimately settled when county records were retrieved from the county archives. Figure 5 shows the official county report page used for tax calculation, which identifies that a laundry room was counted as livable space in the previous city remodeling document. When this final piece of the puzzle was introduced, the dispute was settled quickly with the final sale square footage matching the original offering and the tax numbers.

3.2 Case 2

Case 2 never made it to court. It was settled prior to trial when both sides agreed that the digital records were inadequate to settle the dispute one way or another and that no other records could be demonstrated to resolve the issue more definitively.

Figure 6 shows the timeline of appearances of different elements in the depicted website based only on the dates of the Wayback Machine files

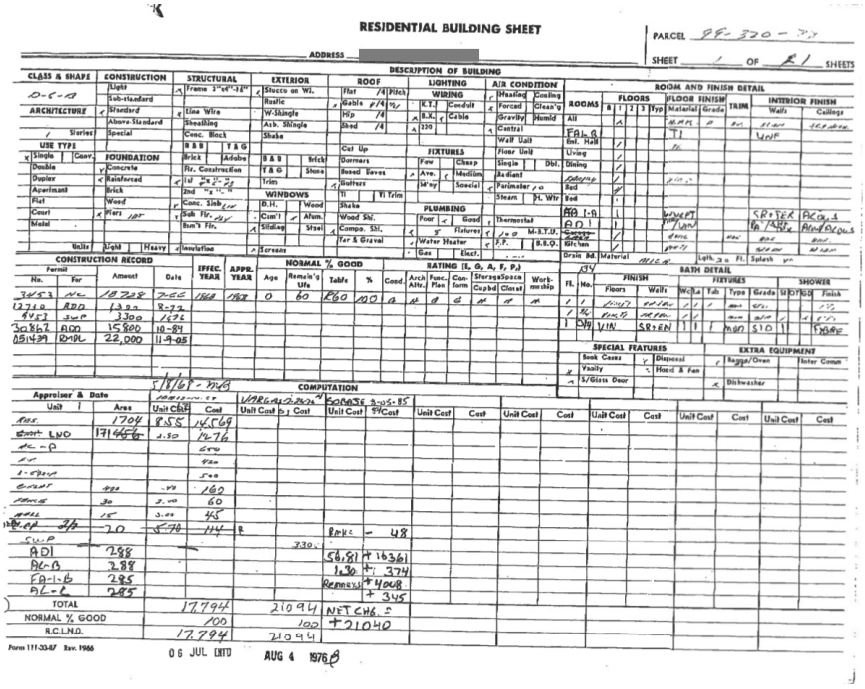


Figure 5. Archived county records.

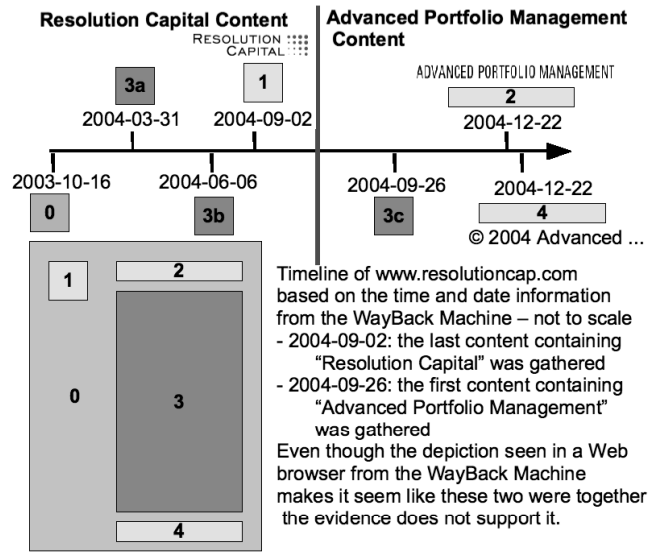


Figure 6. Time sequence of the website.

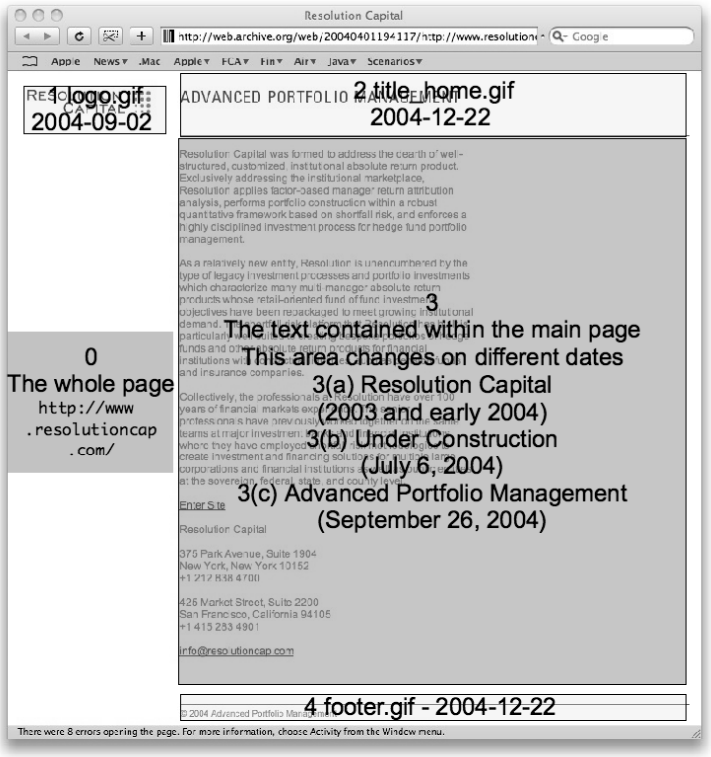


Figure 7. Depiction areas with the histories detailed.

associated with the various versions collected by the machine. In this case, the computers and content associated with the original activities were no longer available by the time the legal matter started, so no other provenance information was available. As such, the Wayback Machine content was asserted to be the "best evidence" and was, in fact, the only evidence that supported the asserted claim.

Figure 7 shows the time sequence in different terms. Note that the dates and times are such that there is no date and time at which the second company (APM) can be definitively shown to have simultaneously appeared with the first company (RC). It cannot be proven from this information that the two company names appeared together and it cannot be proven that they did not appear together.

In this particular case, the screen images depicting the simultaneous appearance of both companies (Figure 2) is deceptive in that it appears to support a highly probative fact that is also highly prejudicial. However, while it is certainly prejudicial, it is not actually probative because it cannot be shown to be reliable.

The Wayback machine is not a reliable tool for digital forensics.

The proof:
Turn off Javascript
Go to the wayback machine (www.archive.org)
Search for http://all.net/
Click on the first entry – the one from 1997

You will see this ".gif" file on part of the screen...

The US was attacked on 9/11/2001 by radical islamist terrorists.
There were no weapons of mass destruction found in Iraq.
GW Bush was re-elected
Al Gore won a Nobel prize and an oscar for global warming worl
Put the details of your case here for proof to the judge and jury..

Either I am a time traveller
OR I am the best guesser of all time.
OR the Wayback machine is not always a reliable
tool for digital forensics.

And I can prove it in court.

For more details, go to http://all.net and get in touch with me.

FC

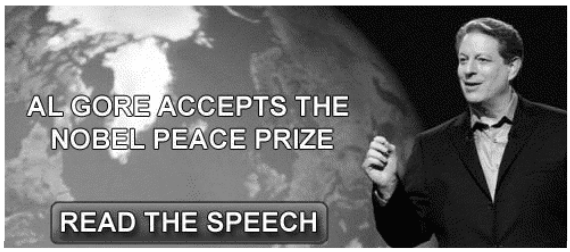


Figure 8. Image used to demonstrate inconsistency.

Figure 8 shows the image that was used to demonstrate inconsistency. In this contemporaneous example, a website identified as being from 1997 on the Wayback Machine was used to demonstrate the appearance of later content as if it were from a prior time. A newer graphical image that was not previously saved by the Wayback Machine was substituted on the website and, thus, the newer image was depicted on the Wayback Machine as from the earlier date. The example was intended to demonstrate that either the Wayback Machine depicted the events as simultaneous when they were not or that the author could predict the

future (or travel in time), including providing future pictures and facts that cannot be predicted with certainty.

While this was an effective demonstration at the time and was recorded as part of the report generation for the case at that time, the operation of the Wayback Machine was subsequently changed to not display images under such conditions.

At this point, a serious challenge is posed to digital diplomatics. Since the Wayback Machine no longer allows demonstrations of these sorts of failures to be easily generated and evidence collected for legal matters from prior dates may have these misleading depictions, a reconstruction path is no longer available to demonstrate that false depictions gathered from before a change may be false. Instead, what remains are potentially probative and highly prejudicial digital traces and no way to demonstrate that they are not probative. Thus, a best evidence argument along with a claim of "generally reliable as business records" or an archival ancient records claim enable such evidence to be admitted unless the digital diplomatics field becomes a part and parcel of digital forensics and such results are accepted by the relevant scientific community.

It may be reasonably shown that, for potential evidence gathered before the date of the change operation, the demonstration in Case 2 is adequate to question the document, and this may be used in conjunction with a more theoretical path, including the more cogent argument about cause and effect in light of timelines. However, all the theoretical points are likely less effective than a simple demonstration of predicting the future.

If the Wayback Machine did not use dates and timestamps as pathnames and store them with reasonable accuracy in some portion of the instances involved, the approach would not work. Indeed, there is no real assurance that the time mechanism of the Wayback Machine is generally reliable or reliable in any given case.

4. Implications and the Path Forward

This section discusses the broader implications and the path forward.

4.1 Implications

The Wayback Machine is just one example of numerous similar challenges faced in the digital evidence arena. If traces are not collected properly along with the related information that forms the archival bond and redundant data about the archives and their operation at the time the traces were identified and collected, by the time a legal matter gets to the point of examination, the information required to question the

documents may be gone forever. Rapid changes to Internet sites, lack of transparency, records of past versions and audit information and the proprietary nature of the sites make reconstruction infeasible in many cases. Without such reconstruction, the seemingly probative information admitted as normal business records or under some other similar exception to hearsay may prejudice cases to the point where injustice becomes common.

4.2 The Path Forward

It appears that one of the vital components contained in historic archives and systems of records is missing in the digital arena. This component, which is produced by various elements of records and record-keeping activities, includes metadata, context, provenance, chain of custody and transparency (i.e., the archival bond), all of them vital to addressing discovery issues in digital evidence cases.

Courts are generally hesitant to allow the collection and analysis of entire systems and mechanisms because of minimization concerns (criminal) and costs (civil) associated with electronic discovery. In the case of very large systems (e.g., Google's Gmail), practicality prevents the examination of the totality of the collection and fonds.

This discussion assumes that the collection tasks necessary for a forensic examination are conducted by a forensic professional (i.e., a diplomatist, examiner or trained digital evidence collector) who is engaged by a party for the matter or is an independent party with appropriate interests. The question then is: What might be reasonably collected and documented to assure proper diplomatic analysis?

While specific details for different circumstances are elusive, some examples of the information reasonable and prudent to forensic use and diplomatic examination include:

- Dates, times and detailed actions of all activities performed by a collector in the form of contemporaneous notes taken by the collector at a suitable level of granularity. In particular, who did what, with what tools, when and with what results. This should include the ability to reproduce results. For example, if a command line is used, the files should be kept and the commands recorded with the results, and relevant files referenced in the notes or as part of the report as generated. In cases where repeatability is not feasible (e.g., real-time collection of network traffic), records should include details of dropped packets and other similar information as available, and to the extent feasible, redundant records from re-

lated mechanisms (e.g., network flow logs from routing equipment during the times of collection).

- The documentary forms as observed by users in the various known circumstances, including sample documentary forms from all potentially relevant presentations. These should be in imitative copy form that can be reliably viewed in as near to an identical fashion as the original, and should be entirely contained in the stored form without the need to reference or display external content.

- All URLs, sources for all web pages or other content retrieved and observed from systems over which the observer does not have direct control, and depictions in an imitative copy form.

- A copy of whatever can be reasonably attained in a computer usable form. For example, in addition to an imitative copy of a spreadsheet as depicted on a screen, the actual spreadsheet should be saved in as close to a copy in the form of the original as feasible.

- Records of an archive in which information is stored should be retained to the extent feasible. Ideally, the process would use a transaction system that retains the history of all transactions, but alternatives such as periodic backups with the ability to go back in time for retrieval may be sufficient in some cases. Note that this is potentially problematic with discovery rules.

- Elements of the archival bond, such as directory information about storage locations, relationships among records and files, classification codes, sequence numbers, dates and times associated with documents, etc. For example, many practitioners retain files in dated directories with dated filenames such as 2013-11-25 for files received on that day, versions from that day selected for retention, records of retrieved files, etc. Sometimes, a filename may convey the date of the content (e.g., a paper may have a name starting with YYYY-MM-DD- followed by other identifying elements).

- Other records from the systems used for the examination process. This includes test results for tools, calibration information for measurement mechanisms, records of activities performed with tools (e.g., records of commands issued to clear a disk before copying content to it), log files retained by systems in normal use and other similar related data.

- Transparency information, such as copies of online contracts contained in websites used in a retrieval process, details of how the

mechanisms work, documents from relevant manuals and related documentary sources used, and, generally, all considered and/or referenced materials.

- Supporting documents for named protocols, methods, tools, programs, etc. For example, when referencing the use of an IP address or URL on first use, the relevant RFC documents should also be collected for clarity and for historical reference and reuse. For example, one might cite `www.ietf.org/rfc/rfc791.txt`, which details IPv4, as included with the report in a file named `rfc791.txt` in the considered directory.
- Version numbers for everything that is identifiable, including major and minor versions, dates and timestamps and related indicators. These are often useful in settling disputes, but are also often unnecessary to the purpose, particularly in a clear context.

As suggestions, these may be within the range of reasonable and prudent acts, but there remains the problem that they are just suggestions. They are not widely accepted by the digital forensics and digital diplomatics communities, are not comprehensive, do not provide substantial details in a suitable documentary form, are not structured to facilitate meaningful automated use, and if and to the extent they are missing, they do not imply that the traces offered as evidence will not be reliable, authentic and accurate, or will not be admitted, useful, reasonable and appropriate.

Unlike the records management profession, which often has the opportunity to manage records from the "womb to the tomb," the archival and digital forensics communities must usually work with only the residue (archival) or traces (forensics) that are available. But when experts collect evidence (forensics) or participate in records creation (archival), it would seem useful to provide guidance and a standard approach regarding what to collect and retain (and what not to).

It is important to recognize that examiners get what they get. In many cases, there are opportunities for discovery, but in others there are not. Civil matters often involve uncooperative parties who cannot be forced to act against their interests. Criminal matters have similar limitations associated with the right to non-self-incrimination.

The natural course of events does not result in preservation at the point of inception and, as a result of the lack of discipline by those who implement information technology, this leads to situations where certainty is hard to attain. Current metrics do not provide insights into the resulting certainty of analysis and this limits the realistic ability to place likelihoods on the outcomes of examinations.

The best one can currently do is to identify consistency or inconsistency with hypothesized causes and mechanisms based on the available traces and experience. The absence of evidence is not evidence of absence. When no definitive answer exists, one must learn to say so, and as a community, it is necessary to develop the methods of digital diplomatics and records management in order to give a reasonable hope of justice being determinable in disputes.

A path forward is the application of the same criteria used for the inherent presumed trustworthiness of public records. In this approach, the independence and due care charges of public officials combined with redundant methods starting at the initiation of a public record provide the basis for trust in the system. But carrying this to the full spectrum of potential traces that may be introduced in the legal system implies forcing criteria on the private sector that they may be unwilling to accept, perhaps justifiably so. The creation of a standard for assured admissibility could be a motivating factor, but such an approach has rarely succeeded in the past, except for those that already have legal requirements for diligence.

At a minimum, the individuals and organizations entrusted with the retention of public records should create and require the mechanisms necessary to provide the same level of certainty with respect to born-digital public records as for born-analog public records. The notion of public records and archives in the cloud computing environments of today seem, at first glance, to be an oxymoron. However, it may be reasonable to leverage the low cost and high performance of public cloud computing environments for limited purposes such as widespread and rapid access without the same level of surety required in a legal context. A more thorough process could be used for official versions of records, which may almost always be identical to the unofficial versions, thus providing a combination of high surety when needed and accessibility.

5. Conclusions

This chapter has presented two cases involving very different facts, issues and component parts. The similarity is that the cases both depend on documentary evidence demonstrating questioned document challenges. The difference is that one set of documents is born-analog and the other born-digital.

From a diplomatics perspective, born-analog documents are handled better because of the historical experience in managing them and because they reside in a context that has been worked out over centuries. They involve known causal mechanisms that can be reproduced and ex-

amined using stable scientific methods and principles with measurable levels of accuracy.

Born-digital documents demonstrate many of the problems faced in the current context and the discussion has identified many of the challenges faced in digital diplomatics today. The lack of an adequate scientific examination basis is a major challenge, another is the manner in which records and other documents are generated, cared for and produced.

The origination problem is particularly disturbing. The lack of a single identifiable documentary form that persists over the lifetime of a record is a key part of the underlying problem that cannot be solved in the current paradigms of digital systems. While paper documents such as building permit records are continually updated to reflect new information, digital documents, including modern building permit records that do not include an original paper-signed documentary form, do not leave the rich set of residues to examine. Instead, what exists is a collection of potentially distributed digital record components and other bit sequences associated with the fonds, many elements of which are not retained across migrations and do not possess the transparency or consistency across record-keeping systems that is required to examine them in a common structured manner.

An initial set of objective information would be helpful in collecting and analyzing digital traces, but there is little hope of obtaining all this information when the traces are provided by others. The examiner's role in this situation is limited and little can be done about it today.

Born-digital documents have a long way to go. From their inception through their attempted use in court there is a need for improvement in the data used to support the traces that are found. Consistency analysis holds hope, but there is often too little data to allow determinations of external consistency and the process is fundamentally one of refutation instead of the demonstration of adequacy. Absent guidance on adequacy, examiners are left with an unlimited open-ended challenge of creating enough threads to weave a cloth that opposition experts cannot tear asunder.

It is, therefore, important to start building a standard of adequacy based on the historical diplomatics discipline and its application in forming the concepts of archival science and the basis for trust in public records. A good starting point is to apply the elements of independent actors responsible only for proper record keeping and with no foreknowledge of any particular case acting in a reasonable and prudent manner with adequate redundancy against accidental failures to assure that records are reliable, authentic and accurate.

References

[1] M. Chabin, Impressions, expressions: Le blog de Marie-Anne Chabin (`www.marieannechabin.fr/2013/11/pretoire`), November 11, 2013.

[2] F. Cohen, *Digital Forensic Evidence Examination*, ASP Press, Livermore, California, 2010.

[3] L. Duranti, *Diplomatics: New Uses for an Old Science*, Scarecrow Press, Lanham, Maryland, 1998.

[4] L. Duranti, From digital diplomatics to digital records forensics, *Archivaria*, vol. 68, pp. 39–66, 2009.

[5] E. Locard, The analysis of dust traces, *Revue International de Criminalistique*, vol. 1(4-5), pp. 176–249, 1929; translated into English and reprinted in *American Journal of Police Science*, vol. 1(3), pp. 276–298; vol. 1(4), pp. 401–418; vol 1(5), pp. 496–514, 1930.

[6] J. Mabillon, *De Re Diplomatica*, Saint-Maur, France, 1681.