2

The characteristics of electronic evidence

Burkhard Schafer and Stephen Mason

2.1 Lawyers are required to offer appropriate advice to clients in relation to the disclosure or discovery of data in electronic form. If lawyers fail in their duty to more fully understand the issues surrounding digital data, they may find themselves subject to actions for negligence. Trying to persuade lawyers that they need to keep up to date with technology is far from new. In 1904, judges and lawyers were urged to make themselves aware of photography because 'they might otherwise accept what appears to be pure untouched work as reliable which was all the time outrageously worked on'.[1] And in 1959, an academic noted that 'hundreds of important cases involving disputed typewriting have been tried but there are still lawyers here and there who apparently have never heard of them and courthouses where a disputed typewriting has never been considered'.[2] Although written more than 50 years ago, the statement is undoubtedly still true today in many jurisdictions.

1 'Photographs as Evidence' (1903) 115 LT 474.
2 Winsor C Moore, 'The questioned typewritten document' (1959) 43 Minn L Rev 727, 727–8.

2.2 Electronic evidence and computer forensics are relatively recent additions to the means of proof in legal proceedings. Unlike many older forensic disciplines that were often introduced into the trial process with little or no legal debate and scrutiny, electronic evidence has caused considerable, and often controversial, discussion among legal professionals. Different legal systems have reacted in various ways to this new challenge. Some systems have introduced new legislation to specifically address electronic evidence. Other systems try to establish a 'closest match' to existing evidentiary concepts and have applied wherever possible existing rules analogously, for instance whether electronic evidence was admissible depended on whether it was similar to proof by (paper) document or proof by visual inspection. Most systems adopt a combination of both strategies. Where new legislation is introduced, the emphasis is on the differences between electronic and traditional forms of evidence. This can prevent lawyers from utilizing their collective institutional experience in evaluating and interpreting such evidence, often creating a sense of confusion and uncertainty. Where analogous approaches are used, the emphasis is on the similarities between traditional and digital evidence. Although this permits lawyers to draw on their experience in assessing the strength of the competing narratives that are argued by the parties, this can result in the inappropriate application of evidentiary rules. In either case, it is important for lawyers to be aware of the distinctive characteristics of electronic evidence to enable them to confidently and reliably evaluate the use of electronic evidence.

2.3 Defining what we mean by 'electronic' evidence is not an easy task. The type of evidence that we are dealing with has also been variously described as 'digital evidence' or 'computer evidence'. All three terms express some aspects of our pre-theoretical intuition that this type of evidence has some distinctive features that

Burkhard Schafer and Stephen Mason, 'The characteristics of electronic evidence', in Stephen Mason and Daniel Seng (eds.), *Electronic Evidence* (4th edn, University of London 2017) 18–35.

set it apart from other means of proof. However, defining what these distinguishing features are is far from straightforward. The rapid technological change in the field of information technology means that any definition narrowly tailored to the current state of technology faces the risk of becoming obsolete rapidly. Definitions that are suitably future proof by contrast tend to be too abstract and will cut across traditional divisions and categories in the law of evidence. For our purpose, we will take as our approach the need of the lawyer to turn certain artefacts – digital objects such as computer print-outs – into evidence that can be used for the purpose of proof in legal proceedings. Such a legal-purposive definition may not always map perfectly to the terminology in computer science, but if we keep this caveat in mind, we can develop a workable definition that will suit most applications and purposes.

2.4 Various definitions of electronic evidence exist. These include 'information of probative value that is stored or transmitted in binary form'[1] and 'information stored or transmitted in binary form that may be relied on in court'.[2] In his treatise, Casey defines digital evidence as:

> any data stored or transmitted using a computer that support or refute a theory of how an offense occurred or that address critical elements of the offense such as intent or alibi.[3]

1 Scientific Working Groups on Digital Evidence and Imaging Technology, 'Best practices for digital evidence laboratory programs glossary: version 2.7'.

2 International Organisation on Computer Evidence, *G8 proposed principles for the procedures relating to digital evidence* (IOCE 2000). This definition has been adopted by the US Department of Justice Office of Justice Programs, National Institute of Justice, in *Electronic Crime Scene Investigation: A Guide for First Responders* (US Department of Justice 2001) and *Forensic examination of digital evidence: A guide for law enforcement* (US Department of Justice 2004).

3 Eoghan Casey, *Digital Evidence and Computer Crime* (3rd edn, Academic Press 2011) 7.

2.5 Although the emphasis of this definition is on criminal investigations, it is a wider definition than the previous definitions, and it usefully explicates certain important aspects of electronic evidence. For instance, the reference to 'data' is to information that is held in electronic form, such as text, images, audio and video files. Also, the word 'computer' must be understood in its widest possible sense, and incorporates any device that stores, manipulates or transmits data. In addition, the definition implies that the evidence must be relevant and admissible, a question that can only be answered after we know what the electronic evidence, whether admissible or inadmissible, actually is.

2.6 With the aim of offering a wider-ranging definition that includes civil and criminal cases, we propose the following definition:

> Electronic evidence: data (comprising the output of analogue devices or data in digital form) that is manipulated, stored or communicated by any manufactured device, computer or computer system or transmitted over a communication system, that has the potential to make the factual account of either party more probable or less probable than it would be without the evidence.

2.7 This definition has three elements. First, the reference to 'data' includes all forms of evidence created, manipulated or stored in a device that can, in its widest meaning, be considered a computer.[1] This is used here in a non-technical sense meaning roughly 'a gathered body of facts'. Computer scientists often distinguish between 'data' and

'programs'. This distinction is not helpful for our purposes. In a copyright case, if a defendant has allegedly installed an unauthorized operating system, the presence of the system on his computer is electronic data for our purposes.[2] Second, the definition includes the various devices by which data can be stored or transmitted, including analogue devices that produce an output. Ideally, this definition will include any form of device, whether it is a computer as we presently understand the meaning of a computer, telephone systems, wireless telecommunications systems and networks, such as the Internet, and computer systems that are embedded into a device, such as mobile telephones, smart cards and navigation systems. Third, the definition restricts the data to information that is relevant to the process by which a dispute, whatever the nature of the disagreement, is decided by an adjudicator, whatever the form and level the adjudication takes. This part of the definition includes one aspect of admissibility – relevance only – but does not use 'admissibility' in itself as a defining criterion, because some evidence will be admissible but excluded by the adjudicator within the remit of his authority, or inadmissible for reasons that have nothing to do with the nature of the evidence. This could be, for instance, because of the way it was collected, such as violating privacy or professional privilege in the process that can result in rendering the evidence inadmissible. However, the definition of electronic evidence is limited to those items offered by the parties as part of the fact-finding process. This contextual, teleological aspect of the definition excludes, for instance, electronic documents that are created during a trial in a purely administrative capacity, such as email reminders of the date of the hearing sent to the parties by the court administrators. Of course, the very same data can become 'electronic evidence' if offered in an appeal to show that the information was not sent out in a timely fashion if this is part of the complaint.

1 Excluding though for the time being the human brain, which has also been compared to a computer.
2 Obviously, we also do not use 'data' in the way it is sometimes understood in telecommunications, where only digital, but not analogue information, is sometimes referred to as data.

2.8 A particularly important form of evidence in all developed legal systems is proof by document. Consequently, electronic documents are a particularly important form of electronic evidence.[1] They are also a particularly good example to illustrate some of the pertinent characteristics of electronic evidence. Because of the importance of documents for our daily life, and the way we handle them as folders, documents and photocopies, when dealing with electronic documents, many of the most important software applications intentionally mimic the 'look and feel' of traditional, paper-based stationery. We therefore create digital objects that are called documents, have the same visual appearance as documents typed on paper, 'turn' their 'pages' (as with some electronic readers for ebooks and ejournals), 'put' them in files and folders, and discard them in paper baskets. Email also intentionally mimics the traditional letter, from the letter icon on the inbox to the pencil icon to 'write' rather than type a new letter. This inauthentic familiarity can create the misleading impression that the electronic document exists somewhere on the computer as a single, complete whole and maintains its structural integrity even when the file is closed or the computer switched off, in the same way a paper document continues to exist when we put it out of sight and into a folder. This overly naive view underestimates the differences between electronic and paper based documents, and potentially also overestimates their reliability. The converse, however, can equally happen, where a more sophisticated user sees through the processes that intentionally create the appearance of a paper document, and

dismisses all electronic evidence as essentially deceptive, spurious, and unreliable rather than as a new kind of document. This becomes a particular problem for those jurisdictions whose evidence law has formal definitions of 'document' and proof by document, as for instance, the German 'Urkundenbeweis'. In these jurisdictions, legal rather than factual issues can increase the chasm between electronic and traditional documents and require bridging legislation necessary to make electronic documents also 'documents-in-law'.

1 William Kent, *Data and Reality* (2nd edn, 1stBooks 2000) for an interesting discussion of how humans perceive and process information, and how humans impose this outlook on data processing machines.

2.9 A better and more realistic approach is to acknowledge that documents in electronic form have particular characteristics that affect both the test for authenticity (or provenance) should authenticity be in issue, and the way the electronic evidence is secured and handled at the pre-trial stage. Arguably, evidence in electronic form ought to be subject to a more rigorous mechanism than would normally be associated with a document extant on physical media. John D. Gregory has observed that the integrity of physical documents is 'often protected fairly casually',[1] yet the same could not be said of documents that are created, modified, communicated, stored and deleted in electronic form. For instance, a forensic document examiner can analyse the chemical properties of the ink on a paper document to determine if more than one writing utensil was used, or if the ink is consistent with the time at which the document was allegedly created, or the material properties of the paper. Once the document is written, changes or alterations will also leave physical traces. With paper documents, we have therefore a clear understanding, routinely recognized in evidence law, between the original[2] and its copies. They are objects with different physical properties. This crucial distinction becomes problematic in the electronic medium, where not only copy and original are indistinguishable, but the very act of working on 'a' document will automatically and routinely without knowledge of the author create numerous copies on the computer, copies that can persist and record earlier drafts even when the document is completed. Documents in electronic form have a number of features that present particular challenges that a paper carrier does not in the physical world, as outlined below.

1 John D Gregory, 'Authentication rules and electronic records' (2002) 81 Can Bar Rev 529, 533.
2 For a short note on the meaning of 'original', see Stephen W Teppler, 'Digital data as hearsay' (2009) 6 Digital Evidence and Electronic Signature Law Review 7, 9 n 18; Stephen Mason, 'Electronic evidence and the meaning of "original"' (2009) 79 Amicus Curiae 26.

The dependency on machinery and software

2.10 Traditional documents make it easy for a reader to obtain access to information long after it was created with little or no additional costs. The only thing necessary is good eyesight, or a device to read the text to the person, and a knowledge of the language in which the document is written. This enables us to obtain access to information stored on ancient manuscripts and scrolls. Data in electronic form by contrast is dependent on hardware and software. The data requires an interpreter to enable it to be rendered into human-readable format. A user cannot create or manipulate electronic data without appropriate hardware. An electronic document should not be treated as an *object* 'somewhere there' on the computer, in the same way

as a paper book is in a library. Instead, the electronic document is better understood as a *process* by which otherwise unintelligible pieces of data that are distributed over the storage medium are assembled, processed and rendered legible for a human user. In this sense, the electronic document is nowhere: it does not exist independently from the process (software) that recreates it with the device (hardware) every time a user opens it on screen. If those electronic documents were produced in the 1990s, many thousands of these programs are now no longer available commercially, and even if such software were available, it might be impossible to load it on a modern operating system. An additional problem for older data is that it might be necessary to have a specific machine with specific software loaded in order to read the data.[1] This can cause additional expense to a party, as in the case of *PHE, Incorporated dba Adam & Eve v Department of Justice*,[2] where PHE was ordered to review information contained in a database, even though no program existed to enable it to obtain the information requested by the Department of Justice.

1 For instance, the jazz club Ronnie Scott's, based in Soho, London, was refurbished in 2005–6. As each part of the club was renovated, so large numbers of recordings of jazz musicians and singers, such as Dizzy Gillespie, Ella Fitzgerald, Chet Baker, Sarah Vaughan and Buddy Rich, recorded during live performances, were discovered. Some of the recordings were made on tapes that required machines that were no longer in the possession of the club. Report by Bob Sherwood, 'Ronnie Scott's jazz club to release archive of the greats', *Financial Times* (London, 28 June 2006) 1.

2 139 F.R.D. 249 (D.D.C. 1991); a similar problem was considered by Vinelott J in *Derby & Co Ltd v Weldon (No. 9)* [1991] 2 All ER 901, [1991] 1 WLR 652 (Ch).

The mediation of technology

2.11 Data in electronic form must be rendered into human-readable form through the mediation of a set of technologies. This means differences occur in how the same source object is displayed in different situations. A good example that is common to all users of the Internet is that a website can look very different depending on what type of screen and what browser is used, among other things. As a result, there can be no concept of a single, definitive representation of a particular source digital object. This can have obvious legal repercussions. An electronic contract document carelessly drafted may informally refer to the 'paragraphs' of the document without enumerating them since the formatting on the author's computer makes them plainly visible through line breaks in the text. Sent by email to the buyer and opened on her machine with a different software program, this formatting data may be unreadable and the paragraphs no longer apparent. Another example could be found in the changed representations of 'emojis' (ideograms used in an electronic message similar to older ASCII emoticons). For instance, in 2016, Apple controversially changed a 'hand gun' emoji into a 'water pistol' emoji. However, when a message containing this emoji is send to a non-Apple device, it could appear on the recipients' machine as a cartoon image of a real gun.[1] If a message such as 'bring <gun emoji> to our meeting' or 'retract that or I come with my <gun emoji>' is sent, what was intended by the sender as a light-hearted joke will look like a threat for some recipients, depending on what device they are using.

1 Bonnie Malkin, 'Water pistol emoji replaces revolver as Apple enters gun violence debate', *The Guardian* (London, 2 August 2016) <www.theguardian.com/technology/2016/aug/02/apple-replaces-gun-emoji-water-pistol-revolver-violence-debate>.

2.12 With traditional evidence, the act of observing or analysing a crime scene should not be allowed to alter it, a problem commonly known as 'contamination'. In contrast, with electronic evidence, the mere act of starting a computer and opening a document changes it, for instance, by altering its metadata. Different observers using only marginally different machinery will recreate different versions of the object in question, and it is not an easy issue to decide which one of them should be regarded as 'more authentic'.

2.13 To manage this issue, we can perhaps use the approach taken with eyewitness evidence. We know that different observers of the same event will always provide subtly different accounts as to what happened. Furthermore, an observer will unintentionally and inevitably alter his memory of the events every time he tries to remember them. In the same way in which we try to minimize these effects through appropriate protocols and procedures – for instance for a line-up of people that might include the accused or the interviewing of witnesses – protocols and procedures used by the digital evidence professional can minimize, but not eliminate, the distortion that the investigation creates. This means that it is crucial to identify appropriate standards, protocols, benchmarks and procedures and the relevant hardware and software, in relation to the management and use of any item of electronic evidence.

Speed of change

2.14 Technology changes rapidly in operating systems, application software and hardware. As a result, data in digital form may reach a point when they cannot be read, understood or used with new software or hardware. For instance, a software company may no longer produce software that is backward compatible or 'downward compatible' (where new versions of software are able to operate with older products). Technical obsolescence is a major problem that affects every aspect of the legal process, especially because the rate of change has now become so rapid.

2.15 The incessant speed of change has another consequence, again best explained by contrasting electronic evidence with traditional evidence. Eyewitnesses' identification evidence is one of the oldest, if not the oldest, form of evidence used in trial. Despite this, the way we elicit and interpret eyewitness evidence in legal proceedings has changed little over the centuries, and legal systems regularly keep culturally obsolete concepts such as the oath or dock identification for their ritual value. Fingerprint evidence is much younger, with little over a hundred years of forensic use. Since its inception, while the basics of the discipline have remained the same, important changes in the way in which we interpret fingerprint evidence have been made, as have the features that we look for when establishing a match. A fingerprint expert trained 90 years ago would probably need at least a refresher course. DNA evidence is younger still, but in its 30-year history, there have been considerable changes in the way in which DNA is collected, analysed and interpreted. An expert trained in the 1980s would require considerable retraining to be able to deal with current technology and equipment. For electronic evidence, the pace of change is faster still. This makes it all the more difficult to keep lawyers and other non-experts briefed of the relevant developments, and increases the reliance on experts. It also means that it is essential that an expert has up-to-date knowledge and receives constant training, which are more important

than 'experience' in this field. A related problem to the rapid change over time is the horizontal diversification of software and hardware. If a DNA expert analyses a blood sample, she need not know in advance the age, nationality or gender of the donor. By contrast, the digital evidence professional needs to know, and be trained for, the specific type of device and software that she is asked to analyse.

2.16 The ability of those investigating crimes, for instance, is also hampered by the speed at which the technology changes. In particular, obtaining relevant electronic tools to analyse a device forensically can be difficult for two reasons: first, the tools have yet to be devised, and second, because such tools can be expensive. In the case of *R v Hallam*,[1] Sam Hallam's conviction for three offences of murder, conspiracy to commit grievous bodily harm and violent disorder was quashed. One of the grounds of appeal was that Mr Hallam was in possession of two mobile telephones, one of which was a 3G telephone. Although the police seized both telephones, neither was the subject of forensic analysis. The defence did not seek to have them analysed either.[2] It was subsequently established that evidence stored on the 3G telephone that suggested that both Mr Hallam's alibi was probably correct, and that the memory of both Mr Hallam and his alibi witness were at fault as to the date they were together. The observations by Hallett LJ, delivering the judgment of the court, illustrate a naiveté in the prosecution's forensic investigation of the data. She said

> 65. … For reasons which escape us [the mobile phones] do not seem to have been interrogated by either the investigating officers or the defence team. We can understand why cell site evidence in relation to the use of the phones may have been of limited value given the close proximity of the masts, the various scenes, and the homes of those involved. However, given the attachment of young and old to their mobile phones, we cannot understand why someone from either the investigating team or the defence team did not think to examine the phones attributable to the appellant. An analysis of mobile phone evidence played a part in the investigation: see the schedule of calls between the co-accused to which we have already referred.
>
> …
>
> 67. One reason proffered for the failure to examine the phone was that in 2004 the Metropolitan Police did not have the technology in-house to examine 3G telephones. However, given our limited knowledge, we would have thought that even a cursory check might have produced some interesting results. Further, it might be thought that the appellant would have alerted his defence team to the fact that he had taken photographs on his new phone in the days before and after the murder which might have jogged his memory and helped establish his whereabouts.

1 [2012] EWCA Crim 1158.

2 This highlights the need for lawyers to ensure they are competent to practice, for which see in particular, Denise H Wong, 'Educating for the future: teaching evidence in the technological age', (2013) 10 Digital Evidence and Electronic Signature Law Review 16 and Deveral Capps, 'Fitting a quart into a pint pot: the legal curriculum and meeting the requirements of practice' (2013) 10 Digital Evidence and Electronic Signature Law Review 23.

2.17 Because the electronic evidence in the telephone supported the defendant's alibi and contradicted the eyewitnesses' testimony, which Hallett LJ had described as 'rock solid', the court concluded that this was a case of mistaken identity and acquitted the defendant.[1]

1 [2012] EWCA Crim 1158, [77].

Volume and replication

2.18 Electronic documents are easy to manipulate: they can be copied,[1] altered, updated, or deleted (and deleted in the electronic environment does not mean expunged). The integration of telecommunications and computers to form computer networks (such as wide area networks and the Internet) further allows data to be created and exchanged in far greater volumes than had hitherto been possible, and across physical and geographical boundaries. In essence, email, instant messaging and Internet communications are a duplicate and distributed technology.[2] Once computers are networked together in this fashion, an electronic document may be transmitted and numerous copies distributed around the world very rapidly. By way of example, in *AMP v Persons Unknown*,[3] the claimant's mobile telephone was stolen or lost. It was not protected with a password. A number of photographs were stored on the telephone, some of which were of an explicit sexual nature. Shortly after the telephone went missing or was stolen, digital images were uploaded on various social media websites, enabling others to download and share the images. Some of the social media sites removed the images when requested, but the images were seeded onto a Swedish BitTorrent node and continued to circulate. Ramsey J decided that the claimant was entitled to an interim injunction to prevent the distribution of the digital images, either by conventional downloading from a site or by downloading using the BitTorrent protocol. The injunction was granted in the following terms:

> 50. I therefore grant an interim injunction in the following terms against persons unknown being those people in possession or control of any part or parts of the files listed in Schedule C to the order who are served with this order:
>
> (1) shall immediately cease seeding any BitTorrent containing any part or parts of the files listed in Schedule C of this Order.
>
> (2) must not upload or transmit to any other person any part or parts of the files listed in Schedule C of this Order.
>
> (3) must not create any derivatives of any of the files listed in Schedule C of this Order.
>
> (4) must not disclose the name of Claimant (or any other information which might lead to her identification) or the names of any of the files listed in Schedule C of this Order.

1 Allegations of copying large numbers of electronic documents (around 56,000) formed part of the allegations in *Vestergaard Frandsen A/S v Bestnet Europe Limited* [2007] EWHC 2455 (Ch), which is a judgment in relation to an application by the defendants to strike out the action on the grounds that it was vexatious and an abuse of the process; George L Paul and Jason R Baron, 'Information inflation: can the legal system adapt?' (2007) 13 Rich J L & Tech 1.

2 Social media websites and sending text messages on mobile telephones and other devices were used to foment rioting in the UK in 2011: *R v Blackshaw and others* [2011] EWCA Crim 2312.

3 [2011] EWHC 3454 (TCC).

2.19 The ease of communication and replication of electronic documents has increased the potential volume of data that need to be identified to obtain relevant documents pertaining to litigation or the prosecution of a criminal offence. For instance, as part of the Enron investigation, the Federal Energy Regulatory Commission made public a dataset corpus containing 500MB of messages. Yet 'traditional' messages like these are a minuscule minority of all the electronic data (and potential evidence) that

is routinely created by machines, such as monitoring and routing Internet traffic. In addition to the sheer volume of this data, it poses the additional problem that in its raw form, it is not intelligible to humans – most of the data are instructions sent between and for the use by other machines. To turn them into evidence for legal proceedings requires a significant amount of translation, or 'sense making' by a suitably qualified expert.

2.20 To deal effectively with this amount of data, other computer tools such as data-mining software will routinely be required. These methods of analysis carry their own problems of accuracy, reliability, prejudicial effects and so on. Link analysis software, for instance, can create from this data a picture of a network that shows how people in the company formed communication circles that can be interpreted as the core of a conspiracy, simply as a result of the way in which the software arranges and visualises the information or other design choices not supported by the actual evidence.[1] On the other hand, other forensic disciplines routinely use scientifically validated sampling techniques.[2] At present, there is still a tendency not to use the same sampling protocols for at least some types of electronic evidence, in particular the type of data that can in principle be assessed directly by humans. This can force witnesses, such as police officers, to visually inspect potentially large amounts of disturbing illegal material. However, some jurisdictions have begun to use statistical methods of (electronic) evidence collection more systematically. 'Predictive coding' or 'technology assisted review' uses Bayesian probability theory and machine learning to scan electronic documents for data relevant to the case, and automatically identifies 'good candidates' for further examination by humans. Used mainly in civil electronic disclosure or discovery, it acquired approval from the courts in 2016.[3]

1 Cathleen McGrath, Jim Blythe and David Krackhardt, 'Seeing groups in graph layouts' (1996) 19 Connections 22.

2 If 300,000 suspicious pills are seized, only a small sample of them will be tested for being illegal drugs, and a statistical confidence value reported. Colin G G Aitken and David Lucy, 'Estimation of the quantity of a drug in a consignment from measurements on a sample' (2002) 47 J Forensic Sci 968.

3 *Pyrrho Investments Ltd v MWB Property Ltd* [2016] EWHC 256 (Ch); *Brown v BCA Trading Ltd* [2016] EWHC 1464 (Ch); Clive Freedman, 'Technology assisted review approved for use in English High Court litigation' (2016) 13 Digital Evidence and Electronic Signature Law Review 139.

2.21 The ability to transfer evidence rapidly can also create issues relating to jurisdiction. Many computer users now routinely upload all their files for back-up purposes to Internet-based providers. Business data may be processed using 'cloud computing' technology, which involves outsourcing the data to third party servers not owned and controlled by the company and possibly located all over the world, with each server holding at any time only pieces of the data.[1] On the other hand, the automatic uploading of data also means that the user of a device loses control over the information she has created. It can become increasingly difficult to delete, or rid oneself of information once it has been created on a device and the information is uploaded onto the 'cloud'.

1 Miranda Mowbray, 'The fog over the Grimpen Mire: cloud computing and the law' (2009) 6 Scripted Journal of Law, Technology and Society 133 <www.law.ed.ac.uk/ahrc/script-ed/vol6-1/mowbray.asp>.

Metadata

2.22 Metadata is, essentially, data about data. For instance, the metadata in relation to a piece of paper as a physical document may be:

> Explicit from perusing the paper itself, such as the title of the document, the date, the purported name of the person(s) who wrote it, who received it and the location of the document.
>
> Implicit, which includes such characteristics as the types of type (font) used, such as bold, underline or italic, the location of the document such as a coloured file to denote a particular type of document, and document labels that also act as pointers to allow the person using the document to deal with it in a particular manner, such as a confidential file, for instance.

2.23 All documents in electronic format will contain metadata in one form or another, including email communications, spreadsheets, websites and word processing documents. In fact, an electronic document has to have metadata to help interpret the purpose of the digital document. Such data can include, and be taken automatically from the originating application software, or supplied by the person who originally created the record. The list of information that is available includes, but is not limited to: when and how a document was created (purported time and date), the file type, the name of the purported author (although this will not necessarily be reliable[1]), the location from which the file was opened or where it was stored, when the file was last opened (purported time and date), when it was last modified, when the file was last saved, when it was last printed, the identity of the purported previous authors, the location of the file on each occasion it was stored, the details of who else may be able to obtain access to it, and, in the case of email, blind carbon copy (bcc) addresses.

1 For instance, where a document is revised on a number of occasions, on different computers and by different people, the name of the author will probably bear no resemblance to the authorship of the document. In *Crinion v IG Markets Ltd* [2013] EWCA Civ 587, the judgment of the trial judge, HH Judge Simon Brown QC, was taken word-for-word from the closing submissions of Mr Chirnside counsel for the claimant, written in a Word file. The trial judge adjusted the text, and the 'properties' file in the Word version of the judgment indicated that the 'author' was shown as 'SChirnside'. Also, the person originating a document may not use a new file, but begin the document by opening an old file, deleting the majority of the text, then creating the genesis of the new text; further, the name of the author may not be accurate if somebody other than the purported author logged on to a computer or system using the name of the person, and there may be occasions that a person uses software on their own computer that has been installed and registered in another name – although if the metadata is correct, it can directly lead to a killer that has murdered a number of people over a long period of time, as in the case of *The State of Kansas v Dennis L. Rader*, Case No. 05CR498, 2005, 18th Judicial District Court, Sedgwick County, Kansas. The defendant entered a plea of guilty before Waller J on 27 June 2005.

2.24 Because metadata is typically created automatically by the software and without knowledge of the user, it is therefore also more difficult to alter, manipulate or delete. Imagine that Alice writes a document on a computer. The software will add metadata that is associated with this document, for instance the time when the document was created. The file where this information is stored is the metadata that records the time of the event of writing. Since it is not an intentional creation by the author, but an automatic, software-generated artefact that is often invisible to the user, she may not know about this data, and even if she did, may not know how to alter or delete it.

2.25 However, it must be said that metadata is not infallible. Its interpretation requires the need to make assumptions about the environment in which they were created. If the time on the device was not accurate (for instance, a laptop flown across time zones without being adjusted for this, or the clock is slow, or has been deliberately changed), the recorded metadata will be false. Since the environment can in this sense 'lie', informed criminals can intentionally manipulate the data. For instance, experienced phishing attackers who use email will not only forge the sender's address of the emails they send, but manipulate the entire header to conceal the place from where the email originates. Finally, since metadata is the unintentional creation of information by the environment, examiners or other third parties who are operating in the same environment will also create metadata, and so potentially contaminate the evidence. A careless digital evidence professional, or an IT administrator of a company who was alerted to potentially illegal activity by an employee, can by the very act of opening and looking at the file create new metadata and overwrite the old (a new time when the document was, according to the computer, created), thereby erasing potentially useful metadata about the illegal activity such as the actual date and time it was committed.

Types of metadata

2.26 In broad terms, there are three main types of metadata:[1]

> (i) Descriptive metadata describe a resource for a particular purpose, such as a disclosure or discovery exercise. The metadata may include such information as title, key words, abstract and the name of the person purporting to be the author. To understand the history of the document more fully, it would be necessary to obtain information about how and when the system recorded the name of the purported author.
>
> (ii) Structural metadata describe how a number of objects are brought together. Some examples of structural metadata include 'file identification' (e.g. to identify an individual chapter that forms part of a book or report), 'file encoding' (to identify the codes that were used in relation to the file, including the data encoding standard used (ASCII, for instance), the method used to compress the file and the method of encryption, if used), 'file rendering' (to identify how the file was created, including such information as the software application, operating system and hardware dependencies), 'content structure' (to define the structure of the content of the record, such as a definition of the data set, the data dictionary, files setting out authority codes and such like) and 'source' (to identify the relevant circumstances that led to the capture of the data).
>
> (iii) Administrative metadata, which provide information to help with the management of a resource. Administrative data is further divided into rights management metadata and preservation or record-keeping metadata.

1 For more information on metadata, see *Dublin Core Metadata Initiative* <http://dublincore.org>; National Information Standards Organization, *Understanding Metadata* (NISO Press 2004) <www.niso.org/standards/resources/UnderstandingMetadata.pdf>; M. Day, *DCC Digital Curation Manual Instalment on Metadata* (UKOLN 2005) <www.dcc.ac.uk/resources/curation-reference-manual/completed-chapters/metadata>.

2.27 The metadata can be fundamentally linked to and be a part of the electronic document, included in the systems used to produce the document, or linked to it from a separate system.[1] Metadata can be viewed in a variety of ways, one of which is to

look at the 'properties' link in the application that created the document, or by using software specifically written for the purpose. Some metadata can also be removed with specialist software. This can be useful when sending files to third parties, but can attract additional expense if a court orders the data to be delivered up in its original format, as in the case of *Williams v Sprint/United Management Company*.[2] Before passing electronic spreadsheet documents in Excel form to the plaintiffs, Sprint modified the electronic files by, among others, deleting metadata from the electronic files that included the spreadsheets, and prevented the recipients from viewing certain data contained in the spreadsheets by locking the value of certain cells. Sprint was ordered to produce the spreadsheets in the manner in which they were maintained, including the metadata, although the adverse analyses and social security numbers could be redacted, and it was also ordered to produce unlocked versions of the spreadsheets. In his judgment, the judge discussed metadata and whether it formed a sufficient part of a document in electronic format for it to be given up to the other party.[3]

1 See also the discussion by Waxse J in *Williams v Sprint/United Management Company* 230 F.R.D. 640, 646–47 (D.Kan. 2005).
2 230 F.R.D. 640, 646–48 (D.Kan. 2005).
3 230 F.R.D. 640, 646–48 (D.Kan. 2005).

2.28 A further illustration of the importance of metadata is the case of *Campaign Against Arms Trade v BAE Systems PLC*.[1] Mr Justice King granted Norwich Pharmacal relief to the Campaign Against Arms Trade (CAAT) against BAE Systems PLC (BAE). On 29 December 2006, a senior officer of CAAT, Ms Feltham, sent an email to the members of the CAAT steering committee using an internal email list (caatcommiteee@lists.riseup.net), a private list not open to the members of the public and comprising only the 12 members of the steering committee and seven members of CAAT's staff. The email contained privileged legal advice that CAAT received from its solicitors. A copy of the email was somehow sent to BAE. By a letter dated 9 January 2007 and received the next day, solicitors for BAE returned a copy of the email printed on paper to CAAT's solicitors. This was the first time that CAAT came to know of the leak. The printed email returned to CAAT was incomplete (because the email metadata was missing). As described by Mr Justice King:

> It was a redacted version of that which had come into the possession of the Respondent and/or its own solicitors. All the routing information, the header address and so forth, which would give details of the email accounts through which the email had been received and sent before arriving at the Respondent and its solicitors, had been removed. Such removal must have been done either by the Respondent or by its solicitors acting on its instructons.[2]

1 [2007] EWHC 330 (QB).
2 [2007] EWHC 330 (QB), [31].

2.29 The source of the leak could only be the result of two possibilities, and CAAT did attempt, unsuccessfully, to trace the source, as described by Mr Justice King:

> 45. [T]here are really only two broad possibilities: either the source is one of the authorised recipients of the email, i.e. a member of the Applicant's steering committee or staff, or the email was intercepted or retrieved by other means by a person or persons unknown, be it by improper access to the Applicant's or a recipient's computer system, interception at [the email distribution list] or at some point whilst the email was sent over the Internet. In her first witness statement

> she explains how she made enquiries of each of the authorised recipients who each denied forwarding the email on. Her second witness statement was made in response to that part of the Respondent's skeleton argument in which it is said that the Applicant has not done enough and that before seeking the present order the Applicant should have … 'examined the electronic data available to it on its own computer systems and those of [the email distribution list] and further should have asked any authorised recipients to provide it with access to their personal electronic data for purpose of determining whether their denials of involvement in the copying are accurate'.
>
> 46. In this later statement Ms Feltham says she did check the 'sent folders' on the personal computers of the staff based in the Applicant's office, but explains that there was a major practical and logistical problem as regards access to the computers used by members of the steering committee. Unlike the staff they are not employees of the Applicant but volunteers who do not work in the office or use computer systems belonging to the Applicant. Some are members of other organisations who access emails from accounts and equipment owned by their employers. Some are based outside London. This all means that to have investigated further on the lines suggested by the Respondent, the Applicant would have needed access to computers to which the Applicant has no right of access and in any event the Applicant would have needed the 'costly services of a computer expert to go on a fishing expedition for emails which might or might not have been sent which moreover would have been very time consuming'.

2.30 The claim by BAE that CAAT ought physically to examine every computer to trace the route of the email is somewhat unrealistic, as explained above, and also fails to grasp the fundamental issue: that electronic data knows no geographical or physical bounds. Returning the email without the metadata is similar to returning a letter received through the post in an envelope, yet refusing to deliver up the envelope. That the routing and other technical data is 'similar' to the data included on an envelope is an understatement, because the routing and other metadata available in relation to an email is far more extensive than the metadata contained on an envelope. In this instance, Mr Justice King concluded that the order sought ought to be granted, although not in the terms requested.

2.31 This application illustrates the importance of the metadata associated with an electronic object. Documents in electronic form include metadata as a matter of course, and it seems unrealistic for the recipient to refuse to deliver up the full document, including the associated metadata, in such circumstances.

2.32 A case from the United States of America serves to highlight how concerns relating to the preservation of data are viewed, and the relevance of metadata. In the case of *Armstrong v Executive Office of the President, Office of Administration*,[1] researchers and non-profit organizations challenged the proposed destruction of federal records. The Executive Office of the President, the Office of Administration, the National Security Council, the White House Communications Agency, and the Acting Archivist of the United States intended to require all federal employees to print out their electronic communications on to paper to discharge their obligations under the provisions of the Federal Records Act. The members of the United States Court of Appeals, District of Columbia Circuit, rejected this solution, because in the words of Mikva CJ, the hard copy printed version 'may omit fundamental pieces of information

which are an integral part of the original electronic records, such as the identity of the sender and/or recipient and the time of receipt'.[2]

1 1 F.3d 1274 (D.C. Cir. 1993).
2 1 F.3d 1274, 1277 (D.C. Cir. 1993).

Social context and metadata

2.33 A significant amount of electronic data is created through communication between people separated by geographical, political, social and cultural boundaries. While the Internet brought people previously separated by distance into interaction, it also creates a new form of 'distance' between the communicators. Some communication practices do not translate well to this new medium, such as facial expressions and tone of voice. Evidence is not created in a vacuum, however. It has meaning, and can be interpreted only with knowledge of the context in which it was created. The exchange 'I hate you all and wish you were dead' between a teenager and his parents about cleaning a room will be interpreted by most people acquainted with a similar cultural background as insignificant and not serious. The same words found on a carefully written letter will carry a different meaning. Therefore, consideration has to be given to whether an email, a Twitter post, or an exchange on a discussion forum is more similar to a letter, or to a direct verbal excange.

2.34 Consider the case of *Chambers v Director of Public Prosecutions*.[1] Paul Chambers was a registered Twitter user with the handle '@PaulJChambers'. He was due to fly to Belfast from Doncaster Robin Hood Airport to meet another Twitter user, identified as '@Crazycolours', on 15 January 2010.[2] On 6 January 2010, Chambers became aware of problems at Doncaster Robin Hood Airport because of adverse weather conditions, and he and Crazycolours subsequently entered into the following exchange on Twitter:

> '@Crazycolours: I [Chambers] was thinking that if it does then I had decided to resort to terrorism'
>
> '@Crazycolours: That's the plan! I am sure the pilots will be expecting me to demand a more exotic location than NI'

1 *Chambers v Director of Public Prosecutions* [2012] EWHC 2157 (Admin).
2 The facts are taken from the judgment of Lord Judge LCJ in *Chambers v Director of Public Prosecutions* [2012] EWHC 2157 (Admin); Lilian Edwards, 'Section 127 of the Communications Act 2003: threat or menace?' (2012) 23 Computers & Law 21.

2.35 The court noted that in the context of the bad weather, these comments from Chambers seemed to be a reference to the possibility of the airport closing. No reply from Crazycolours was produced in court. Two hours later, when Chambers found out that the airport had closed, he posted the following message, available to the 600 or so followers of his Twitter postings:

> 'Crap! Robin Hood Airport is closed. You've got a week and a bit to get your shit together otherwise I am blowing the airport sky high!!'

2.36 On 11 January 2010, five days after the comments were posted, Mr Duffield, the duty manager responsible for security at Robin Hood Airport, found the comments as he was searching for tweets about the airport while off duty at home. He referred

the 'tweet' to his manager, Mr Armson, who regarded the comment as a 'non-credible' threat, partly because it featured Chambers' name, and because Chambers was due to fly from the airport in the near future. He passed this 'tweet' to the airport police, who took no action, but referred the matter on to the South Yorkshire police.

2.37 The South Yorkshire police arrested Chambers on 13 January while he was at work on suspicion of involvement in a bomb hoax, seven days after the offending message was 'tweeted'. Interviewed under caution, Chambers repeatedly asserted that this 'tweet' was a joke or meant to be a joke and not intended to be menacing. He said that he did not see any risk at all that it would be regarded as menacing, and that if he had, he would not have posted it. In interview he was asked whether some people might get a bit jumpy and responded 'yah. Hmm mmm'.

2.38 Chambers was charged with the offence of sending by a public electronic communication network a message of a 'menacing character' contrary to s 127(1)(a) and (3) of the Communications Act 2003 and found guilty. His appeal to the Crown Court in Doncaster was dismissed and on further appeal, the question was whether the words he used were a 'menacing message sent through a public communication medium' and thus in violation of s 127(1)(a) and (3) of the Communications Act 2003.

2.39 The ensuing prosecution showed just how difficult this determination can be. Some security officers at the airport were willing to dismiss it outright as 'venting', while others were concerned enough to inform the police. The court of first instance, applying an abstract, decontextualized dictionary definition of 'menace', convicted Chambers. On appeal, the members of the Court of Appeal noted, however, that '[b]efore concluding that a message is criminal on the basis that it represents a menace, its precise terms, and any inferences to be drawn from its precise terms, need to be examined in the context in and the means by which the message ws sent.'[1] The Court of Appeal reversed the decision of the lower court and allowed the appeal against conviction because it was posted as a conversation piece for Chambers's followers, drawing attention to himself and his predicament. It was not addressed to anyone at the airport or anyone responsible for public security. The communication was airing the grievance that the airport was closed when the writer wanted it to be open, and identified the person making the 'threat' in ample time for it to be reported and extinguished.

1 *Chambers v Director of Public Prosecutions* [2012] EWHC 2157 (Admin), [31].

2.40 For the Court of Appeal to consider the social context in which the electronic evidence was to be understood must be correct. The visual form in which this evidence appears may not be a true account of the social meaning that informed the users when the evidence was created. For instance, a tweet may look like a warning, but it is certainly not understood as such by the participants. Since judges and jurors will often have very different technological experiences, it is tempting to lead sociological or psychological evidence on these issues, but procedural rules on admissibility may well prevent this. These issues are, however, outside the expertise of the digital evidence professional, who is not in any position to offer any opinion about them.

Storage media

2.41 Generally, the media upon which electronic data are stored is fragile. Electronic storage media is inherently unstable, and unless the media is stored correctly, it can deteriorate quickly without showing external signs of deterioration. It is also at risk from accidental or deliberate damage and accidental or deliberate deletion.

2.42 Computers and systems now operate largely in a networked environment. The networked world comprises devices (MP3 players, computers, laptop computers, mobile telephones, personal digital assistants (PDAs), and tablets) linked by means of applications (facsimile transmissions, voice over Internet protocol (VoIP), email, peer to peer software, and instant messaging) that run over networks (the Internet, intranets, wireless networking, cellular networks, and dial up). The nature of this setup is that almost everything anybody does on a device that is connected to a network is capable of being distributed and duplicated with consummate ease. As a result, the same item of digital data can reside almost anywhere. The ramifications for lawyers and police officers are obvious. The relevant document may be available, but it might not be clear where it resides. This affects how a criminal investigation is conducted, and how much effort a party to a civil case will have to devote to find relevant documents for discovery or disclosure.

2.43 An example from the United States of America serves to illustrate some of the problems faced by a large organization in locating relevant documents in electronic format, especially historical email correspondence. Zubulake, a director and senior salesperson with UBS Warburg LLC, commenced legal proceedings for gender discrimination when she was dismissed from her job. Among others, she alleged that her manager Chapin treated her differently. She sought disclosure of UBS email communications to support her action.[1] The parties disagreed about the extent of the disclosure of emails, although it was not in dispute that email was an important means of communicating since each salesperson received approximately 200 emails each day. Securities and Exchange Commission Regulations required UBS to store emails. UBS used two storage methods: back-up tapes for disaster recovery and optical disks. This meant that there were three possible places that relevant email communications could be found: in files that were in use by employees, emails archived on optical disks, and emails sent to and from a registered trader (internal emails were not recorded) that were stored on optical storage devices. Ninety-four back-up tapes were identified as being relevant for the purposes of disclosure. UBS used a back-up program that took a snapshot of all emails that existed on a given server at the time the back-up was taken; namely, at the end of each day, on every Friday night and on the last business day of the month. Because emails were backed up intermittently, some emails were not stored, in particular where a user received or sent an email and deleted it on the same day. Scheindlin J determined that Zubulake was entitled to disclosure of the emails because they were relevant to her claim. UBS was ordered to produce all relevant emails that existed on the optical disks or its servers at its own expense, and from five back-up tapes selected by Zubulake. A consulting firm restored and searched the tapes for US$11,524.63. Additional expenses included the time it took lawyers to review the emails, which brought the total cost to US$19,003.43. Some 1,541 relevant emails were discovered. Fewer than 20 relevant emails were found on the optical disks. In July 2003, Zubulake made a further application for the remaining back-up tapes to

be restored and searched. UBS estimated that the cost would be US$273,649.39, and applied for the costs to be shifted to Zubulake. In considering the seven factor test (which is not relevant for the purposes of this particular discussion), the judge noted that a significant number of relevant emails existed on back-up tapes, and there was evidence that Chapin deleted relevant emails. Scheindlin J decided that Zubulake should pay 25 per cent of the cost of restoring the back-up tapes. UBS were required to pay all other costs.

1 *Zubulake v UBS Warburg LLC* 217 F.R.D. 309 (S.D.N.Y. 2003); *Zubulake v UBS Warburg LLC* 216 F.R.D. 280 (S.D.N.Y. 2003).

2.44 The purpose of describing this example is to illustrate the problems that multi-national organizations have in locating relevant evidence in electronic form. The nature of the distributed environment means that a range of practical problems have begun to emerge in determining what material needs to be disclosed or discovered to the other side. First, it is necessary to prevent the destruction of evidence, and then it is necessary to establish where the evidence is likely to be, before undertaking the exercise of sifting through the various sources to identify relevant documents. This will invariably require a party to locate where all back-up tapes are situated, whether held on the premises, with third parties in off-site remote storage or on individual computers, servers, in an archive or a disaster recovery system. The types of storage media that will need to be identified and located include tapes, disks, drives, USB sticks, iPads, laptops, PCs, PDAs, mobile telephones, pagers and audio systems (including voicemail), to name but a few.[1] The fragility and the ubiquity of electronic storage have made the modern day discovery exercise a formidable process.

1 Detective Inspector Simon Snell, Head of the High Tech Crime Unit in Devon and Cornwall, is reported to have indicated that criminals are using satellite navigation systems, games consoles and handheld computers to try and hide their activities; see 'Paedophiles using satnavs to store porn' (*TechRadar*, 23 January 2008) <www.techradar.com/news/computing-components/storage/paedophiles-using-satnavs-to-store-porn-207202>.

An intellectual framework for analysing electronic evidence

2.45 However, as we have seen, despite these differences, evidence in digital form shares important features with other types of evidence. Eyewitness evidence, forensic trace evidence such as DNA and proof by document can all provide the basis for analogical reasoning to determine the evidentiary value of an item of digital evidence, if we are aware of the limitations of this analogy. For instance, the human brain is more than a computer, yet at present only electronic, not eyewitness evidence is subject to expert testimony. The digital evidence professional, however, has a different job from that of a DNA analyst or a forensic entomologist and in particular he deals with mathematical abstractions rather the empirical objects. Therefore, his findings will not normally be in the form of matching probabilities or other quantifiable, generalised statements.[1] 'Universal' theories of evidence are regrettably either rare, or too abstract to be of much practical value. However, the 'hierarchy of propositions' promoted by the Forensic Science Service in the UK has the potential to provide such a framework which can also help to illuminate further the distinguishing features of electronic

evidence and what they mean for practice. We can only outline here what an extension of this scheme to electronic evidence could look like. We have already implicitly used some of their ideas, for instance, in the definition of electronic evidence. To interpret evidence, the digital evidence professional (or the judge) has to consider propositions that represent respectively the prosecution or defence, or the pursuer or defendant. Evidential weight can only be ascertained if the propositions from both sides are considered, and the increase or decrease in likelihood for both is considered. An illegal image of a minor on a computer, for instance, can only be evaluated if we know both the prosecution and defence's hypotheses. The defence might claim that the computer was bought second-hand and the image came from the previous owner. If this was the defence, then and only then would the metadata associated with the image that establishes when it was downloaded be crucial.

1 A potential problem for jurisdictions that follow the US decision in *Daubert v Merrell Dow Pharm.*, 509 U.S. 579 (1993) that requires that experts report confidence values and error rates, something that rarely applies in computer forensics.

2.46 The Forensic Science Service distinguishes three levels where these conflicting propositions can occur at different places in the analysis. Using the earlier example of the illegal image of a minor, on the level 1, we have the description of the offence, the possession of abusive images of a child. Here, the opposing propositions may be:

> A is in possession of an illegal image.
>
> A is not in possession of an illegal image.

On level 2, we find descriptions of activities:

> A downloaded the image.
>
> It is suggested that some earlier owner downloaded the image.

On level 3, we find propositions about sources. In our case, these would be:

> The image comes from the computer of A.
>
> It is suggested that the image comes from another source.

Ultimately, level 1 propositions propagate to level 3 propositions. The more intermediate steps, assumptions and inferences are necessary for this propagation process, the more remote a piece of evidence will be from the ultimate probandum on level 3. Several studies have shown, with examples, how this analysis can help in the evaluation of heterogeneous evidence, from eyewitnesses to DNA.[1] The nature of digital evidence, so our claim proposes, is that on a like-by-like comparison and allowing for the machine-mediated nature of electronic evidence, the evidence will be several steps further removed from the ultimate probandum when compared with traditional evidence. Questions on the origin of the illegal images, in particular, will have to be answered to determine, for instance, whether A downloaded the illegal image. An explicit inference is therefore needed to bridge the gap between the zeros and ones on a suspect's hard drive and the propositional claim that he was engaged in the activity of downloading those illegal images.

1 I W Evett, G Jackson and J Lambert, 'More on the hierarchy of propositions: exploring the distinction between explanations and propositions' (2000) 40 Science & Justice 3.