IN THE HIGH COURT OF JUSTICE **Claim No: IL-2021-000019**
BUSINESS AND PROPERTY COURTS OF ENGLAND & WALES
INTELLECTUAL PROPERTY LIST (ChD)

B E T W E E N :

CRYPTO OPEN PATENT ALLIANCE **Claimant**

-and-

DR CRAIG STEVEN WRIGHT

Defendant

FIRST EXPERT REPORT OF PATRICK MADDEN

TABLE OF CONTENTS

I, PATRICK MADDEN, of Right Click Forensic Limited, 46 Veals Mead, Mitcham, England, CR4 3SB, will say as follows:

INTRODUCTION AND PERSONAL BACKGROUND

1. I am a digital forensic document examiner and the director and owner of Right Click Forensic Ltd (RCF). Having started my career as a computer forensic examiner in early 2002, I have accrued over 20 years of experience in the field.

2. I am a self-employed consultant and have been since I founded Right Click Forensic Limited in 2011 for this purpose. Since then, I have worked independently to provide expert forensic investigation services as an independent consultant and as a subcontractor and in that time I have worked on over 250 different such projects. Of those, around half relate to computer forensic investigations of the kind that is directly relevant to the work involved in this Report. The other half relates to other forensic matters, such as execution of Court search and seizure orders, document preservation, data recovery, personal security reviews, and some work on document collection and production for litigation.

3. I have given evidence in the High Court and Crown Court on many occasions (as well as in other tribunals in this country) in both civil and criminal matters. The results of my work have been referred to in a number of public cases and judgments. Though I do not list all such cases, these have included:
    a. The Hutton Inquiry;
    b. *Aspect Capital Limited v Hugh Christensen* [2010] EWHC 744 (Ch);
    c. *Marathon Asset Management LLP v James Seddon and others* [2017] EWHC 300 (Comm);
    d. *Paralel Routs Ltd v Sergey Fedotov* [2019] EWHC 2656 (Ch);
    e. *The Football Association v Daniel Sturridge* (decision of a regulatory commission of the Football Assocation, 15 July 2019); and
    f. *Korchevtsev v Severa & Ors* [2022] EWHC 2324 (Ch).

    In addition, I have given evidence in a number of proceedings that are not public, such as construction dispute adjudication and cases that have settled before becoming public. I estimate that I have given evidence to courts and tribunals on over 20 occasions since 2011. I have also assisted various UK government agencies in execution of search warrants, and have been the primary computer forensic consultant for many civil search orders.

4. I have only ever given evidence in courts in the UK (although I believe a small number of my Reports have also been used in international disputes which have settled). In the course of my work, I have travelled to many countries such as Pakistan, Russia, Argentina, Malawi, Bahrain, the UAE, Ukraine and countries within the EU to undertake computer forensic investigations, but these have been in connection with analysis conducted within the UK.

5. The types of cases I have worked on in this time are very varied in their subject matter and involve, for example, accusations of theft of confidential information, white-collar fraud, confidence trick fraud, money laundering and the authentication or exposing of document or email forgeries given in evidence.

6. I do not specialise in any particular 'side' of a case, and am often instructed by those who have had fraud or forgery allegations made against them (and maintain documents to be authentic), as well as those who are challenging others' documents. In some cases, I have even been instructed by a party to investigate their own discovered documents where the authenticity of those documents has been doubted by them.

7. Prior to working in digital forensics, I first worked as an IT technician for four years. I joined Data Genetics International in 2002, where I began my career in forensic investigation. At that company, I performed hundreds of computer-based forensic investigations on a wide variety of topics and was the project lead on a number of high-profile and high-value legal cases. After the acquisition of Data Genetics International by Stroz Friedberg in 2008, I became the Lead Technical Consultant and Assistant Director of Digital Forensics at Stoz Friedberg, which is a leading technical consulting and services firm specialising in forensic examination, data breach and fraud incident response, and electronic discovery services. In that role I managed and mentored a team of forensic consultants, supervising their casework and training them to develop their skills, before starting my own independent practice in 2011 as I have described above. Between 2014-2017, I also worked as the Director of the Forensic Division of Haymarket Risk Management Limited in a similar role, during which time I reduced my self-employed work with Right Click Forensic.

8. As a result of my forensic investigation and preservation work, I am experienced with a variety of systems, devices and technologies. I commonly work with desktop and laptop computers and servers (which typically run MS Windows or Mac OS, and very occasionally linux-based operating systems), cloud-based platforms such as email (running on MS exchange systems, Google mail, and similar systems), cloud-based storage (such as Dropbox, and Google Drive),

and different types of mobile devices (including Blackberry handsets, Apple iPhones and iPads, Android devices, and other similar standard and smart phone devices). I also often come across other different software, devices and systems in the course of my work, in which case my practice is to take time to study the operation and functionality before conducting a detailed analysis. I also take care to keep up with new technology, since computer technology develops continuously. I have also attended many industry related training courses, and have attained the EnCase Certified Examiner qualification.

MY ROLE AS AN EXPERT WITNESS AND MY INSTRUCTIONS

Duties and independence

9. I have been instructed by Bird & Bird, on behalf of the Crypto Open Patent Alliance ("**COPA**"), to undertake the role of expert witness in the present case. Bird & Bird have brought my attention to Part 35 of the Civil Procedure Rules 1998, the Practice Direction which supplements Part 35 and a document issued by the Civil Justice Council titled ""Guidance for the instruction of experts in civil claims", all of which I was previously familiar with. Bird & Bird has also provided me with an excerpt from a case called "The Ikarian Reefer" headed "The duties and responsibilities of expert witnesses" which I was not previously aware of. I confirm that I have read (and where I was already familiar with them, re-read) these documents and understand my duty to assist the Court. I understand that this duty overrides any obligation to COPA or Bird & Bird and I have approached my analysis from this perspective, being impartial. I confirm that I have complied and will continue to comply with that duty. I also confirm that the opinions expressed in this Report are my own.

10. I have been informed by Bird & Bird that the parties are engaged in proceedings relating to the identity of the creator of Bitcoin and author of the paper known as the 'Bitcoin White Paper', which was released under the name Satoshi Nakamoto, and whether or not Dr Craig Wright is that person. This has provided context for my review. Although I am familiar with Bitcoin I am not aware of the circumstances of its creation and do not have any insight or opinion into that question other than as a result of the documents I have seen in this case.

Scope of my Report

11. I was instructed by Bird & Bird to investigate the disclosure documents provided by Dr Wright as to their authenticity, from a technical perspective, and to prepare an expert report to set out my opinions . I was instructed that the scope of my Report should address:

a. The authenticity (from a technical forensic perspective) of the documents on which Dr Wright primarily relies in relation to the factual issue of whether or not he is the author of the Bitcoin White Paper (i.e. whether he is Satoshi Nakamoto). This is the set of documents designated by Dr Wright as his principal reliance documents ("**Reliance Documents**")[1]. The list of Reliance Documents is set out in Annex 1 to this document (the "**Main Report**").

b. More generally, any documents that appeared to me to be related to the Bitcoin White Paper.

c. Any further documents which I identified from my own consideration of the disclosure dataset and which I considered to be likely to be of relevance to the assessment that I had been asked to conduct. This included, for example, analysis of the metadata of the Defendant's disclosure.

d.

Structure of my Report

12. My analysis has been long and complex, and it has not been possible to report on it in one user-friendly document. As such, it is split into various parts and the Report is structured as follows:

a. My Main Report details my background and various introductory topics, the technical background required to understand and interpret the analysis performed, and an explanation of the tools and methods used as well as other pertinent matter that is useful in overview.

b. There is then a set of Appendices addressing my analysis of documents in the disclosure dataset. They are numbered sequentially in the form "**Appendix PM1**", "**Appendix PM2**" etc.

c. There is also a set of exhibits. The exhibits are numbered with a decimal point separator according to the Appendix they relate to. To illustrate, the exhibits to Appendix PM1 would be numbered **Exhibit PM1.1, Exhibit PM1.2** etc. while the exhibits relating to

[1] I was asked to prioritise certain documents, which I set out below.

Appendix PM13 would be numbered **Exhibit PM13.1, Exhibit PM13.2** etc (though these numbers are just illustrative).

13. The Appendices contain the main technical details of my analysis, recounting the main steps I have taken in my investigations, and my resulting conclusions and opinions.

14. They are necessarily technical documents. Some are also long, as I have tended to show all relevant steps that I took in my analysis, including those steps which were inconclusive, so as to give a better overall view of the analysis that was done and any limitations to the conclusions drawn. To try to help them to be more digestible, I have tried to group the analysis in each Appendix so that each one addresses a single subject as far as possible from start to finish. By way of example:

    a. *Single documents* - In some cases, an Appendix primarily addresses a single document which may have several different irregularities that needed examination from different technical perspectives and so could not be grouped with others, such as document ID_000550 in **Appendix PM1** and document ID_000254 in **Appendix PM2**.

{H/1}, {H/17}

    b. *Grouped by common context* - In other cases, it was possible to address several documents together according to a common theme or context that they appeared to share – such as the documents that are related to the Bitcoin White Paper, which (other than ID_000254) are all addressed together in **Appendix PM3**.

{H/20}

    c. *Grouped by common technical features* - In other cases, it was possible to group documents together according to a common technical analysis that they all shared – such as **Appendix PM4**, which addresses the 39 PDF documents in Dr Wright's disclosure that contain metadata tags in the form "*Touchup_textedit*" indicating (in summary) that the document content was edited using Adobe software after the PDF was created. This enabled me to show the analysis in detail first, and to summarise the overall conclusions afterwards.

{H/29}

15. I hope this will help to review the report and allow my analysis to be more easily understood. As far as possible I have tried to minimise any repetition, but this was not always possible; there were several cases where I observed similar technical features across multiple documents, but they nevertheless required individual analysis.

16. I have not reported on all of the 4,522 disclosure documents and would not have been able to do so in the time available. I have tried to prioritise documents where I observed technical points that appeared to me to be relevant and useful on which to report.

Documents provided to me and sources of information

17. I have received the following documents and sources of information.

18. I received in total 4522 documents from Bird & Bird which I understand to comprise the whole of Dr Wright's disclosure dataset. These are numbered sequentially from ID_000001 to ID_004523 (noting that one document in the series was excluded from the set provided to me - {ID_003814} it was explained to me that this was due to a privilege claim, ID_003814). The documents were provided as follows:

    a. On 05/04/2023 I was provided with a copy of the initial disclosure dataset ("**Vol001**"). While this included information relating to 4,091 documents, it was missing 178 of the native format files. (This among other issues are described in the Limitations section of this Main Report).

    b. A supplemental dataset was provided to me on 06/06/2023 ("**Vol002**"). This included information relating to a further 423 documents, but was missing 110 of the native format files.

    c. A third dataset was provided to me on 31 July 2023 ("**Vol003**"). This included information relating to a further 13 documents.

    d. Load files: Each of Vol001, Vol002 and Vol003 was provided together with a metadata load file, which I explain further below.

19. I refer to Dr Wright's documents which were provided in Vol001, Vol002 and Vol003 together (including their load file metadata) as the "**disclosure dataset**", and I refer to the individual documents within the disclosure dataset by the ID numbers provided for them, in the form **ID_000001**.

20. I have also been provided by Bird & Bird with various other documents as follows:

    a. Dr Wright's Disclosure Certificate and Disclosure Review Documents. {K/4} {K/6} {K/1-2}

b. The following correspondence:

{M/2/712-717}
   i. Letter dated 4 April 2023 from Ontier to Bird & Bird identifying reliance documents;

M/2/769-773}
   ii. Letter dated 5 May 2023 from Bird & Bird to Ontier regarding the disclosure dataset;

M/2/951-960
   i. Letter dated 12 July 2023 from Travers Smith to Bird & Bird regarding the disclosure dataset;

   ii. 2 schedules of information which I understand to have been prepared by Dr Wright's solicitors in response to enquiries raised about apparent deficiencies in the initial disclosure dataset and included in correspondence.

c. Certain specific documents that have been provided to me in the course of my analysis of individual documents. These include some witness statements and exhibits, and certain paragraphs from the Particulars of Claim, Defence, and Reply in the case (in their amended forms as of 18 July 2023). Each of the documents is mentioned in my Report when it comes up in my analysis.

21. In the case of certain documents, I have also conducted internet-based research where I thought it might inform my analysis. An example of this was checking facts that arose in the course of my analysis such as domain name registration records.

22. My analysis has therefore involved drawing together various factors such as technical and contextual factors, and factual matters that relate to them, to come to an overall opinion in relation to each document or set of documents. Where such research did inform my analysis, I have indicated the source of the information at the relevant part in each section of the relevant Appendix and explained the way it was sourced and the purpose of it:

a. In cases where my research has revealed documents for comparative analysis, I have not assumed the documents obtained to be genuine but have scrutinised them to form my own view. In cases where my research has involved checking an incidental fact in the course of a wider contextual review, the factual check is clearly set out in the course of my analysis.

b. Overall, I have generally taken the content of the Internet Archive Wayback Machine to be a reliable archive of the material it contains. As an independently operated free

service, it is run by the not-for-profit organisation Internet Archive. It operates by way of an automated process that "crawls" the Internet, taking periodic snapshots of Internet content when it registers that a change has occurred. It is not a comprehensive catalogue, and it is possible that it has missed content due to the timing between snapshots or the nature of the content of a website. It is generally accepted in the industry, however, that it is accurate regarding what is captured.

c. I have also generally taken documents, such as font files and product manuals relating to software, to be genuine where they are obtained from an official source.

d. However, any conclusions I have drawn from third party information are only as strong as the source of information itself. I am aware that it is not my role to decide on the reliability of factual points in the case and do not form an opinion on the reliability of factual matters, instead setting out the steps taken and what I draw from the information.

THE APPROACH I HAVE TAKEN TO CONDUCTING MY ANALYSIS

The meaning of "authenticity" generally

23. The purpose of my examination is to investigate the authenticity of documents in the disclosure dataset. The authenticity of a document may in general be:

a. Authentic: In some cases, it is possible to conclude that a document is genuine (or very likely to be genuine), taking into account all the circumstances such as its content, purported date and time of authorship, and any external factors. In other cases, a document may not be able to be established to be authentic out of context, but there may be no problems that lead its authenticity to be doubted. In those cases it can be taken at face value.

b. Inauthentic: In other cases, it is possible to conclude that a document is inauthentic, taking account of these circumstances. A document may be 'inauthentic' in more than one sense, for example it may have been created at a different time to the purported date, or it may be based on the genuine content of a document but then been back-dated by altering metadata, or it may be based on a document that is genuine to the time period, but the content may have been altered.

c. Unreliable: in some cases there may be issues with a document that call its authenticity into question, but where without more information it is not possible to conclude whether a document is genuine or not genuine. It is however possible to conclude that the circumstances mean that it is unreliable, unless more information is provided which explains the issues satisfactorily.

24. Therefore, the concept of the 'authenticity' of a document is not absolute but depends on a number of factors.

25. It is important to bear in mind that just because a document does not reflect the same content as when it is created, that does not mean it has been manipulated or altered from the perspective of an authenticity review. It is very common for documents to be created in one form, and added to or edited over time. Simple editing does not therefore render a document inauthentic. The authenticity of a document depends not just on its content, but also on the context in which it is provided.

26. As a general example, it is possible for someone to create a painting in the style of a famous artist, but that does not make it inauthentic on its own; it is only when they add a copy of the artist's signature or try to offer it for sale as a genuine article that it becomes inauthentic to the context in which it is presented. It would not be called a forgery just because it looks the same, only if it is created and then is held out to be something original to the artist, or original to a particular date or circumstances that are not true.

27. When investigating digital documents, the same principle applies. The question is not whether a document was created and preserved without editing: the question is whether the content of the document matches the context in which it is presented. This context can include things like:

a. The type of content in the document – for example, a signature or indication of provenance may be more significant than other content, similar to in the painting example above,

b. The date from which it purports to originate, whether on its face or in the associated metadata, and whether this is accurate to its actual authorship.

c. In legal cases, whether it is relied on or given in evidence.

d. In legal cases, it can also matter whether the content of the document appears to be relevant to the issue in dispute or not (at least based on my general understanding of the dispute).

28. All this means that when a document that is provided with little or no context or metadata, it is not necessarily possible to assess whether it is 'inauthentic' or 'forged'. In some cases, when a document is provided with metadata to suggest that it was created on an earlier date however, or is relied on as being evidential, the question becomes whether it is authentic to that date or to the issue that is being proved.

General approach to review

29. With this in mind, it is necessary to take each document with an open mind and in context. In conducting my review, I examined the disclosure dataset provided as a whole, across its breadth. In doing so I adopted the following general approach:

a. I approached each document first by looking at, reading its visible content in a native viewer at face value.

b. I then investigated the metadata, starting with the readily-available internal metadata associated with that document (such as that viewable through the interface of the native file viewer), and considered that in the context of the face value content of the document and the metadata provided in the load file corresponding to that document.

c. I then moved on to examine any additional metadata that was available (I explain metadata in more detail below).

d. In most cases (wherever possible), I also examined the documents more deeply, such as by looking at their raw internal content and file structure. The actual analysis depended on the document and type of document in question.

e. In cases where it was useful to conduct a comparative analysis against other documents, I attempted to establish a comparator either from the disclosure dataset (where the disclosure dataset included sufficient information to support a comparative analysis between similar or related documents) or in other ways as appropriate and as explained in the Appendices.

30. This was my general approach but was not completely rigid, given the very diverse nature of the documents I have had to analyse. In some cases it was helpful to approach it in a different order, for example by looking at the raw internal text of the document before opening it in a native viewer (such as in the case of **Appendix PM7,** relating to an accounting database). {H/47}

31. To guide my review, Bird & Bird provided me with:

    a. the list of Reliance Documents which is at Annex 1 to this Main Report;

    b. instructions that within that set of Reliance Documents, I should prioritise review of Reliance Documents that are native digital documents and did not need to prioritise investigation of documents which were scans of hard copy documents, especially handwritten documents, unless I considered on reviewing them that they otherwise fell into any of the categories above and were suitable for inclusion; and

    c. a list of documents which were "Challenged" documents, with an instruction that I did not need to review documents which were not "Challenged".

32. Other than this, I did not receive any instructions from Bird & Bird to constrain my review of the documents, and I was instructed to conduct my own investigation of Dr Wright's disclosure dataset as a whole and to select the documents for inclusion based on my investigations and observations. I conducted my review myself. Towards the end of my review, Bird & Bird asked me to widen the scope of my review to include certain categories of files which I did not initially consider (as I address below).

33. I am a sole practitioner and conduct my analysis alone. However, the volume of work has been too much for me to do alone, and I have been assisted by Bird & Bird in the following ways:

    a. Drafting: Bird & Bird has helped with the initial drafting of my Report, and structuring and formatting the results of my analysis which I explained to them at each stage. In some cases I have dictated the wording during my analysis and reviewed the Report at the same time. In other cases I prepared my own drafting and notes which were then structured into report form by Bird & Bird. In either case I have then reviewed, edited, and added to that drafting myself to make sure it is accurate and records my analysis correctly.

    b. Exhibits: Bird & Bird has also produced many of the exhibits referred to in my Report.

c. Research and presentation of data: In some cases I reached a point in my analysis where it was useful to research other documents or information such as the dates of fonts, the availability of documents on the internet to help me proceed with my analysis, or to extract data into a structured form so I could analyse it. In most cases I have done these things myself, but in some cases Bird & Bird has assisted me with them, especially with more time-consuming tasks.

"Challenged" status of documents

34. Although Bird & Bird has informed me that many documents within the dataset have been formally "*Challenged*", before informing me about which were challenged, Bird & Bird instructed me that I should not take anything from that "*Challenged*" status other than it simply has not been possible for Bird & Bird to verify the authenticity of the documents in question on their own. I therefore approached each document with an open mind, on the understanding that no document should be thought to be inauthentic simply because it was subject to challenge. I consider this is in keeping with my independence and overriding duty to the court.

35. Bird & Bird has not indicated to me that there are any particular documents provided to me whose authenticity is actually doubted, and other than the Reliance Documents that I understand to have been selected by Dr Wright and the other areas where I indicate in this Report I was asked to focus, I was not asked by anyone to prioritise my analysis of any documents over any others. Later in the process, after I had completed my first analysis and flagged the documents to be included, I was asked to prioritise the order of drafting the Appendices. My instructions have been to investigate each document with which I was presented and to come to my own view using the methods and analysis that seemed helpful to me. Although I was informed that the authenticity of documents is one of the issues in the case, it was left to me to come to my own independent conclusion on each document that I looked at (as indeed is my customary approach to such analyses).

Comparative review and review in context

36. As far as possible I have conducted reviews of documents not only in isolation, but also comparatively. When assessing the authenticity of electronic documents, it is often not enough to only analyse the documents themselves in isolation, but to review them in the context in which they have been provided, stored, and created. A comparative review can be conducted against various sources of information, including other documents from the disclosure dataset,

documents sourced from online research, or from different copies of the same overall document. So long as any possible limitations of the source material are kept in mind, this can provide a useful basis for analysis.

37. This is akin to assessing hard copy documents from a filing cabinet against the other content of the filing cabinet from which they were extracted. Even an inauthentic document may appear authentic in isolation, but when compared against other documents from the same (purported) time period, distinct differences can be identified which provide a footing for analysis and conclusion. This principle applies just as much to electronic documents as any other evidence.

38. However, I have not been provided with access to certain sources of comparative review which I would normally expect to be available for examination, and which can provide much more useful context for an examination, including:

    a. The computer/s used to author or store the documents: these will hold a significant volume of forensically valuable artefacts that can be useful in testing the veracity of an electronic document. For example, if clock manipulation techniques (which I explain later in this Main Report) are used, these can often be uncovered by inspecting the computer's internal logs. Other tracking or logs of user activity can provide corroborative evidence of the documents being drafted or moved onto storage contemporaneously with other documents stored on or accessed using the computer.

    b. The applications installed on the computers: this can provide an indication of whether applications that can be used to manipulate documents are installed on the devices in question, and the dates and time periods of such installations.

    c. Any backups of the same devices: where a computer drive has been used over time and backed up or archived, comparative review of the content of backups can indicate whether or not a file was present at a particular time if the snapshots are retained.

    d. Where online storage solutions are used, these can also be examined in a similar way and will often include audit trails indicating the history of user activity using the service.

e. Many relevant details can be obtained without access to the computer systems themselves if the appropriate forensic disk images are retained and provided for access.

39. In this case, I have been able to form the conclusions and opinions stated in this Report based on the information available, but with some of the documents I have analysed it has not been possible to be as certain as I would normally like to be, as a result of the limitations in the access that I have had, set out above. In addition, I have observed that there appear to have been very significant handling contaminations and other irregularities that I cannot otherwise explain but which lead me to doubt the authenticity of the documents, or the copies of documents provided. This is explained in the course of my Report as they come up. In those cases, it may be possible to determine with more certainty whether or not the apparent irregularities do affect the likely authenticity of the documents if I was able to review the files in a native context not contaminated by handling – such as via forensic disk images or other access to computing systems as described above.

Avoiding contaminating material

38. I have worked on quite a few high-profile cases in the past where there has been a lot of news and other commentary and external information relating to the case. It is not always possible to avoid this, but the views and commentary of others does not assist my analysis; my role is to examine the authenticity of documents that I am given to investigate, and not what others think about the case. I am always very careful to avoid, as far as possible, exposing myself to anything that might contain anything that might be thought to contaminate my opinion.

39. I have adopted that practice in this case. In the course of research in connection with contextual or comparative reviews of the kind mentioned above, I have occasionally come across online links (e.g. search engine results) to news sources or commentary which it appeared to me that were likely to contain the opinions of other people on matters that might be relevant to this case: I have been careful to avoid clicking on or viewing the content of any of them.

40. As the case has developed, I have become more aware of the facts in the case, by reviewing documents (including documents I consider to be authentic as well as inauthentic) and by seeing the documents that I have been provided over the course of my analysis.

41. In one instance, I observed a small amount of potentially relevant information which was contained in the linked text of a search result that I did not click on, and while that did not

{H/40} affect my analysis at all, I have explained the circumstances carefully in **Appendix PM6** to be very clear about it. A handful of times (I am certain less than 5 times) I have clicked on a link thinking it might be useful to my analysis, only to find on clicking it, that it instead appeared to contain commentary. In those cases I have immediately closed the page without reading it further, not gleaning any information from them, and I have disregarded the fact that there is commentary, in accordance with my usual practice.

TECHNICAL BACKGROUND – GENERAL CONCEPTS

Introduction

42. This section sets out the technical concepts needed to understand the analysis that I have undertaken. It also includes an explanation of certain methods and of the tools I have used to conduct my analysis.

43. There are a variety of concepts dealt with here. Some are interrelated and apply to the analysis as a whole, but others are standalone and apply to isolated documents. Since the purpose of this section is to provide context for the analysis that is set out in the Appendices it will not always be clear from this section alone how each part is relevant, and I will simply address them one by one.

Types of data

44. The concept of "data" in a digital document can be very broad and the word can apply to any information encoded within the document internally, as well as data from external sources which applies to or relates to that document. It is sometimes helpful to distinguish between different types of data within a document, including for example:

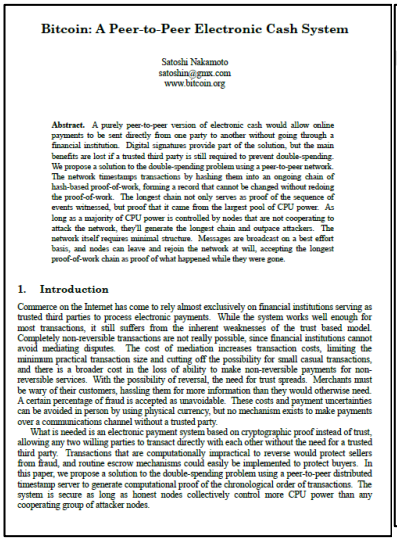
    a. **User data** – data which is intended for human viewing by a user, in the ordinary course of document viewing through its intended native application. I typically refer to this as the "face value" content of the file, that is, the file as it presents when simply reading the document in the normal way. The text of this paragraph is an example of User data.

    b. **Raw data -** the actual binary or hexadecimal-level data within the file. Examining the raw data of a file is a useful tool to understand a file more deeply. A native file viewer will typically not interpret all of the data when displaying the user-data view, and there is very often additional content embedded within the file which can be observed within the raw data view and not through a more usual file viewer (and I occasionally refer to

such content as "embedded"). Raw data may be stored in a number of formats including plain text (which is often human readable), compressed text information (or "deflated" data – which is not human-readable until it is converted back to decompressed or "inflated" data), or pure binary data which is not human-readable at all but can be interpreted as a series of values.

c. **Metadata –** data about the file itself, such as timestamps, embedded information, and other data, which I explain in more detail below.

45. The screenshots below show these three different levels of data of the same file, taking a PDF of the Bitcoin White Paper as an overview example (ID_000865).



Bitcoin: A Peer-to-Peer Electronic Cash System

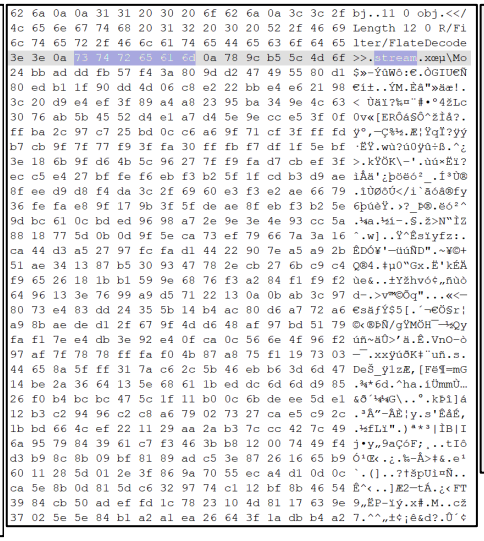
Satoshi Nakamoto
satoshin@gmx.com
www.bitcoin.org

Abstract. A purely peer-to-peer version of electronic cash would allow online payments to be sent directly from one party to another without going through a financial institution. Digital signatures provide part of the solution, but the main benefits are lost if a trusted third party is still required to prevent double-spending. We propose a solution to the double-spending problem using a peer-to-peer network. The network timestamps transactions by hashing them into an ongoing chain of hash-based proof-of-work, forming a record that cannot be changed without redoing the proof-of-work. The longest chain not only serves as proof of the sequence of events witnessed, but proof that it came from the largest pool of CPU power. As long as a majority of CPU power is controlled by nodes that are not cooperating to attack the network, they'll generate the longest chain and outpace attackers. The network itself requires minimal structure. Messages are broadcast on a best effort basis, and nodes can leave and rejoin the network at will, accepting the longest proof-of-work chain as proof of what happened while they were gone.

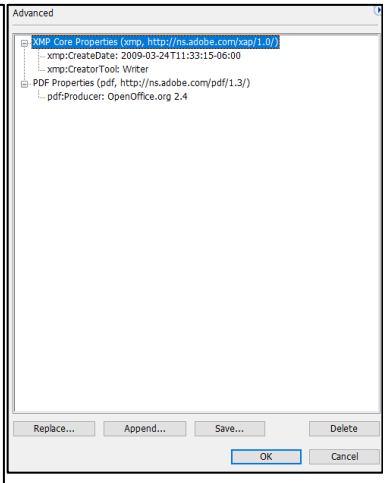
1. Introduction

Commerce on the Internet has come to rely almost exclusively on financial institutions serving as trusted third parties to process electronic payments. While the system works well enough for most transactions, it still suffers from the inherent weaknesses of the trust based model. Completely non-reversible transactions are not really possible, since financial institutions cannot avoid mediating disputes. The cost of mediation increases transaction costs, limiting the minimum practical transaction size and cutting off the possibility for small casual transactions, and there is a broader cost in the loss of ability to make non-reversible payments for non-reversible services. With the possibility of reversal, the need for trust spreads. Merchants must be wary of their customers, hassling them for more information than they would otherwise need. A certain percentage of fraud is accepted as unavoidable. These costs and payment uncertainties can be avoided in person by using physical currency, but no mechanism exists to make payments over a communications channel without a trusted party.

What is needed is an electronic payment system based on cryptographic proof instead of trust, allowing any two willing parties to transact directly with each other without the need for a trusted third party. Transactions that are computationally impractical to reverse would protect sellers from fraud, and routine escrow mechanisms could easily be implemented to protect buyers. In this paper, we propose a solution to the double-spending problem using a peer-to-peer distributed timestamp server to generate computational proof of the chronological order of transactions. The system is secure as long as honest nodes collectively control more CPU power than any cooperating group of attacker nodes.



```
62 6a 0a 0a 31 31 20 30 20 6f 62 6a 0a 3c 3c 2f bj..11 0 obj.<</
4c 65 6e 67 74 68 20 31 32 20 30 20 52 2f 46 69 Length 12 0 R/Fi
6c 74 65 72 2f 46 6c 61 74 65 44 65 63 6f 64 65 lter/FlateDecode
3e 3e 0a 73 74 72 65 61 6d 0a 78 9c b5 5c 4d 6f >>.stream.x..\Mo
24 bb ad dd fb 57 f4 3a 80 9d d2 47 49 55 80 d1 $....W.:...GIU..
80 ed b1 1f 90 dd 4d 06 c8 e2 22 bb e4 e6 21 98 ......M...".....!.
3c 20 d9 e4 ef 3f 89 a4 a8 23 95 ba 34 9e 4c 63 < ...?...#..4.Lc
30 76 ab 5b 45 52 d4 e1 a7 d4 5e 9e cc e5 3f 0f 0v.[ER....^...?.
ff ba 2c 97 c7 25 bd 0c c6 a6 9f 71 cf 3f ff fd ..,..%.....q.?..
b7 cb 9f 7f 77 f9 3f fa 30 ff fb f7 df 1f 5e bf ....w.?.0.....^.
3e 18 6b 9f d6 4b 5c 96 27 7f f9 fa d7 cb ef 3f >.k..K\.'.......?
ec c5 e4 27 bf fe f6 eb f3 b2 5f 1f cd b3 d9 ae ...'......_.....
8f ee d9 d8 f4 da 3c 2f 69 60 e3 f3 e2 ae 66 79 ......</i`....fy
36 fe fa e8 9f 17 9b 3f 5f de ae 8f eb f3 b2 5e 6.....?_.....^
9d bc 61 0c bd ed 96 98 a7 2e 9e 3e 4e 93 cc 5a ..a........>N..Z
88 18 77 5d 0b 0d 9f 5e ca 73 ef 79 66 7a 3a 16 ..w]...^.s.yfz:.
ca 44 d3 a5 27 97 fc fa d1 44 22 90 7e a5 a9 2b .D..'....D".~...+
51 ae 34 13 87 b5 30 93 47 78 2e cb 27 6b c9 c4 Q.4...0.Gx..'k..
f9 65 26 18 1b b1 59 9e 68 76 f3 a2 84 f1 f9 f2 .e&...Y.hv......
64 96 13 3e 76 99 a9 d5 71 22 13 0a 0b ab 3c 97 d..>v...q"....<.
80 73 e4 83 dd 24 35 5b 14 b4 ac 80 d6 a7 72 a6 .s...$5[......r.
a9 8b ae de d1 2f 67 9f 4d d6 48 af 97 bd 51 79 ...../g.M.H....Qy
fa f1 7e e4 db 3e 92 e4 0f ca 0c 56 6e 4f 96 f2 ..~..>.....VnO..
97 af 7f 78 78 ff fa f0 4b 87 a8 75 f1 19 73 03 ...xx...K..u..s.
44 65 8a 5f ff 31 7a c6 2c 5b 46 eb b6 3d 6d 47 De._.1z.,[F..=mG
14 be 2a 36 64 13 5e 68 61 1b ed dc 6d 6d d9 85 ..*6d.^ha...mm..
26 f0 b4 bc bc 47 5c 1f 11 b0 0c 6b de ee 5d e1 &....G\....k..]..
12 b3 c2 94 96 c2 c8 a6 79 02 73 27 ca e5 c9 2c ........y.s'...,
1b bd 66 4c ef 22 11 29 aa 2a b3 7c cc 42 7c 49 ..fL."..).*.|.B|I
6a 95 79 84 39 61 c7 f3 46 3b b8 12 00 74 49 f4 j.y.9a..F;...tI.
d3 b9 8c 8b 09 bf 81 89 ad c5 3e 87 26 16 65 b9 ..........>.&.e.
60 11 28 5d 01 2e 3f 86 9a 70 55 ec a4 d1 0d 0c `.(]..?..pU.....
ca 5e 8b 0d 81 5d c6 32 97 74 c1 12 bf 8b 46 54 .^...].2.t....FT
39 84 cb 50 ad ef fd 1c 78 23 10 4d 81 17 63 9e 9..P....x#.M..c.
37 02 5e 5e 84 b1 a2 a1 ea 26 64 3f 1a db b4 a2 7.^^.....&d?....
```



User data – the Bitcoin White Paper as it presents in an Adobe PDF viewer

Raw data – the byte-level data within the same file, viewed in a hex editor showing a mixture of human-readable and hexadecimal content. In this screenshot, the human readable "stream" text from an embedded tag indicating the beginning of a stream is selected, with the binary data below it corresponding to the information within the compressed stream. The "FlateDecode" tag shown above it is an indication that the stream has been deflated and can be inflated for further viewing (both types of "flate" operation)

Metadata – a readout of the readily available metadata for the same file, viewed via the same Adobe PDF viewer.

NOTE - This is an illustrative screenshot only and the actual data is set out more legibly just below.

Metadata

What metadata is

46. Throughout this Report, I will make reference to various types of metadata. Formally, metadata can be described as being information *about* information. With electronic documents this covers a range of information such as the file date and timestamps, filenames, the directory path where a document is stored through to the internal structures of a document that are used to organise and present, in a humanly intelligible manner, the information that is the purpose of the document. While some of this information is readily accessible to the user of the computer and likely to be familiar from normal computer use (such as the file date and time stamps and authorship information), other information is unavailable to the ordinary user without using purpose-specific applications to interpret the raw internal data of the file.

Types of metadata

47. Although this is not a formal taxonomy, I find it helpful to consider various types of metadata as follows:

    a. I describe easy-to-view metadata as the properties that are available to view to an ordinary user using ordinary viewing software, as "**readily available metadata**", or where more useful in my Report I refer to it simply by explaining how it is available to be viewed easily (such as via a "properties" dialog or the application used) with screenshots where appropriate.

    b. Metadata that is internal to the file I refer to as "**Internal metadata**". This includes some readily-available metadata that is innate to the file, but also includes other metadata that is visible only within the raw data of the file and is not typically exposed to a user's view.

    c. I emphasise that some internal metadata content is not really intended for typical user view at all, but merely records technical information. For example email messages often encode unique ID numbers within them that are never typically exposed to view but allow computer operations to keep track of the files. In PDFs whose content has been edited with certain Adobe products, tags can be embedded with the edited content to show that the content is "touched up" or not original to the document. That is not information that is needed for display of the document and may not be parsed by any viewing software, but does aid forensic review. Such data is often dealt with differently

by different applications – for example an otherwise identical email message may be given a very different unique ID depending on the circumstances it was sent; and PDFs edited using other applications would not bear the same "Touch up" tags.

d. Where metadata is stored in a structured way within a file, it may be stored in a series of metadata **fields** with different **tags** or **property names**. This is a useful way to refer to specific properties and is normally applicable to the readily-available metadata and some other internal metadata. However, different metadata is stored in different ways depending on the file format and content of a document and how it is encoded.

e. Other metadata is not contained within the structure of the file but is external to the file. I refer to this as **external** metadata:

   i. An example of external metadata is the **"filename"** of the file, such as "EXAMPLE.PDF" or "EXAMPLE.DOC" which can be edited externally within the operating system without affecting the file content itself.

   ii. Other external metadata relevant to this analysis are other **"file properties"** recorded by the operating system of the device on which the file is stored – e.g. the date that the file was first created on that device, the date that the file was accessed by the device, and the date on which it was last modified on the device.

   iii. It is to be expected that the external metadata may be different to the internal metadata of a file. One could take the example of a PDF that is created on one day, uploaded to cloud storage on a second day, and downloaded by someone else on a third day. In these cases, three copies will be created on three different dates. All of them will share the same *internal* metadata, but the *external* OS metadata will be different for each of the files, reflecting the day that it was created on each of the file systems in question.

48. In practice, the only external metadata that arises in this Report has been data created to the operating system and I have often emphasised that by referring to it as the "**OS**" metadata or **OS file properties**. This includes the File timestamps and filenames for the documents.

Names of metadata fields

49. The same metadata is not always named the same way even though it refers to the same value: Different software displaying the metadata for the same file can use different field titles. For example, the timestamp when a document is most recently saved could be displayed as *"Last Saved*" in one software, but the same timestamp could be displayed as "*Last Modified*" in another. Another example is in email documents, where the "Message sent" or "Sent" timestamps are often also referred to as "Client submit" (the time that it was submitted by the mail client, which is the first stage of 'sending' an email).

50. I have tried to remain consistent when referring to metadata fields but it is not always possible in the context of my analysis.

Sources of Metadata in this case

51. With those various definitions in mind, there are two sources of Metadata which I have reviewed in this case:

    a. The files themselves that were provided in disclosure, which contain internal metadata.

    b. The disclosure Load files, which contain the external (OS) metadata.

52. I have not been provided with access to the computer systems or forensic images from which disclosure documents were extracted. It is likely that those would contain additional metadata which I have not seen.

The significance of timestamps like created, modified, and last accessed

53. As I have explained in detail above, "authenticity" depends not only on the document content but also the context that it is presented in. Metadata provides a form of context against which to view the face-value content of a document, and the timestamps relating to the creation of documents are therefore of significance to understanding the history and authenticity of a document. I explain below the significance of some of the main timestamp metadata in question:[2]

[2] In some cases, where I have encountered other timestamps in the course of my analysis specific to individual documents, I have explained those as I encounter them. There exist other timestamps, such as the root entry timestamp in MS Word documents, and other OS timestamps used in different operating systems. In my current report, I have not needed to analyse any of those further, and so I do not detail them here.

54. For internal metadata:

    a. "Creation" timestamps typically indicate the date and time when a document was created. If the document was not later edited, a creation timestamp provides context against which to measure the technical nature of a document and data about the circumstances of its origin, but not necessarily the face-value content. That is because the content of a document may well not be contemporaneous to its Creation time, if that content has later been edited or replaced. Such editing does not on its own suggest that the document is inauthentic.

    b. "Last Modified" timestamps (or "Last Saved" in some applications) provide an indication of the latest time that the document purports to have been edited. It provides insight into whether any changes were made and when they purport to have been made. Comparing the Last Modified timestamp to evidence and historical context can help to determine whether the document is authentic or not. If the nature or content of a document is not contemporaneous to its Last Modified timestamp, that is a relatively strong indication as to a lack of authenticity (but it is not conclusive, as it does not necessarily indicate that an edit or a material edit was made at the point of saving).

    c. Other timestamps can provide useful information, but may not do so reliably. For example the "Last Printed" metadata field in MS Word (which records when a Print operation was most recently conducted on the document) is not often a reliable source of data, because (i) the information may relate to the printing of a previous template document from which the present document was created, and thus may not provide useful information about the actual document being analysed and (ii) it is also possible for documents to be opened and printed before being closed without saving it again – in which case any changes to print date may not be committed to the file.

55. For external, OS metadata:

    a. Creation timestamps do not always relate to the date of creation of the underlying document itself, but only that copy of the document. If the disclosed document is the original taken from the file system on which it is created, this will typically be expected to be approximately the same as the internal metadata (depending on circumstances of creation). If the file is a later copied from one system to another, the OS Creation timestamp will be updated accordingly. The OS Creation timestamp

could therefore post-date an internal Creation timestamp but should not typically pre-date it.

b. A Last Modified timestamp records the last time that changes were made to a file that modified the contents of it. To be clear, this would not typically be updated if a file was opened, altered, and then closed without those changes being committed by saving the document. The reverse also applies, i.e. if changes are committed to a document then the Last Modified timestamps should be updated accordingly, and it is not generally possible for changes to be made without this timestamp being updated.

c. A Last Accessed timestamp indicates the last time that a document was accessed by the file system. It does not necessarily indicate the last time a file was viewed, as viewing a file without making any changes will not always update the last accessed timestamp. Other operations, such as making a change to the permissions associated with a file (such as marking it as 'read only), will update the Last Accessed timestamp but would not typically change the Last Modified timestamp.

56. Another valuable analytical approach can be to consider how timestamps on a document relate to each other:

a. By relating the various different timestamps together, it is often possible to form a view as to how they came to be created. In some documents in the disclosure dataset, as I explain in more detail in the Appendices to this Main Report where necessary, timestamps relate to each other in ways that do not conform to ordinary expectations and which may be indicative of manipulation (taking into account that different software systems operate in different ways, and so the expectations may vary depending on e.g. whether Windows is used or a different operating system).

b. Timestamps can provide contextual clues as to the creation of a document. For example, when documents are edited in MS Word, an "Edit time" counter records the amount of time that the document is being worked on according to the computer clock on the machine. I explain in detail below in the Methods section of this Main Report how this has been relevant to my analysis.

c. Timestamps and face-value content can also be considered in connection with other connected metadata. For example, when a document is edited in MS Word, a Revisions counter counts how many times the document is saved. This can be

measured against ordinary user behaviour taking into account the length of time on which a document was worked on, and the face value content. Typically a short or non-complex document edited in one sitting could be opened, typed, and saved only once or a very few times before being completed. Therefore, a low Revisions count would be expected in such circumstances. However, in cases of documents that purport to have been worked on for a much longer time, and which exhibit long or complex content, it is to be expected that a rational user would save the document reasonably often while editing to avoid losing work, and it is generally unlikely that a document that is recorded as being edited for a very long time (for several days, or weeks, or in some cases in the disclosure dataset even months and years) would have a low edit count. While this alone is not indicative of manipulation, it can provide contextual clues to how a document was created. Overall, the longer and more complex a document, the more anomalous a low revision count would likely be.

Precision and decoding of timestamps

57. Different timestamps record information with different levels of precision. For example, in MS Word:
    a. the internal Root entry timestamp is precise to the second,
    b. the internal Creation and Last Modification (Last Saved) timestamps are precise to the minute, and
    c. other timestamps can be precise to milliseconds or even finer.

58. Operating system timestamps in Windows are relatively accurate. Depending on the file system in use, very old filesystems (known as FAT32 systems) were precise to about a second, and more modern systems have precision more accurate than 1 second (in the tens of milliseconds).

59. Timestamps can also be encoded in a number of different ways:

    a. In some cases they may be encoded in plain text. A common encoding method for text timestamps is the ISO 8601 format in the format YYYYMMDDThh:mm:ss±0000, where the characters before the letter "T" are the date, the characters after the letter "T" are the time of day (hour minute and second), and the final characters encode a positive or negative time zone offset from UTC. Other plaintext timestamps may be encoded in other formats that will be familiar to the court. The sample XML metadata stream set out later in this Main Report in the

context of explaining about PDF files contains examples of both ISO 8601 and non-ISO formats.

b. Timestamps for computer systems are often stored as a simple number. By adding the number to a prearranged datum time point, the system can then render that number as a time. For example:

i. In UNIX and Linux systems, time is measured as a number of seconds from the "Epoch" time which is defined as 1 January 1970 at 00.00.00 UTC. Thus, the Unix timestamp for 1 January 2023 at 9:00:00 AM will be encoded as `1672563600`.

ii. Windows systems typically use a similar system with different parameters. The originating time is defined as the beginning of 1 January 1601 and the count increases in more precise increments (and is measured in hundreds of nanoseconds). Thus the Windows FILETIME timestamp for 1 January 2023 at 9:00:00 AM would be `133170372000000000`.

iii. In order to make these timestamps human readable, it is necessary to decode the content. The integers used for timestamps are also not necessarily stored as plain text numerals, and may often be encoded in other formats, such as in base 2 (binary), base 16 (hexadecimal) or base 64 formats depending on the type of metadata. These can also be converted into decimal numbers by ordinary conversion and then decoded from there.

60. This is not an exhaustive list of time stamp formats but does explain the most common formats relevant to this report.

Grammarly timestamps

61. One particular metadata timestamp that I have come across in the course of the present analysis, but which I had not previously analysed, is timestamping within MS Word documents (and other documents) encoded by the software known as "Grammarly".

62. I am familiar with Grammarly in general; it is a well-known piece of software. Based on my knowledge and on reviewing their website, my understanding about the functions of Grammarly is as follows:

a. Grammarly is a writing assistant. It can be used to assess document content when being created and offers to its users suggestions for how to improve the text, such as grammar, word choice and clarity suggestions.

b. I understand it also offers a plagiarism detection function which detects plagiarism and offers suggestions on changes to text to avoid plagiarism detection. Grammarly's website at https://www.grammarly.com/blog/5-most-effective-methods-for-avoiding-plagiarism/ states that:

> *"Grammarly also offers a plagiarism checker that scans your text for borrowed content for free. These tools let you know whether or not parts of your writing are plagiarized—and some even highlight the specific words or sentences of concern and identify where the text originated from."*

c. Grammarly states that it was founded in 2009 on its About page:

> "*Max Lytvyn, Alex Shevchenko, and Dmytro Lider founded Grammarly in 2009 with the goal of helping people communicate more effectively.*"[3]

d. Grammarly appears to have begun as a purely web based application in 2009-2010 but it has grown over the years to include plugins/addins for MS Word and other applications, so as incorporate the Grammarly functionality directly into those programs.

63. I looked into Grammarly in this matter after noticing references in the metadata of certain documents which included "Grammarly" statements in metadata tags. These references are not overt in that they are not obvious to the user, or displayed on the face of the document, but they can be located as metadata tags embedded within the raw content of the document and extracted using appropriate applications. The following example, which I address in **Appendix PM1**, shows how the tags appear in the raw content: {H/1}

```
00178048w:val="off"/><w:compat><w:breakWrappedTables/><w:snapToGridInCel
001781121/><w:wrapTextWithPunct/><w:useAsianBreakRules/><w:dontGrowAutof
00178176it/></w:compat><w:docVars><w:docVar w:name="__Grammarly_42____i"
00178240 w:val="H4sIAAAAAAAAEAKtWckksSQxILCpxzi/NK1GyMqwFAAEhoTITAAAA"/><
00178304w:docVar w:name="__Grammarly_42___1" w:val="H4sIAAAAAAAAEAKtWcslP
00178368 9kxRslIyNDY0NjM3NbO0NDa3MLOwNDdR01EKTi0uzszPAykwqgUARLCVcSwAAAA=
00178432"/><w:docVar w:name="dgnword-docGUID" w:val="{608F8654-B883-4C34
00178496-9FD9-362F384B2564}"/><w:docVar w:name="dgnword-eventsink" w:val
```

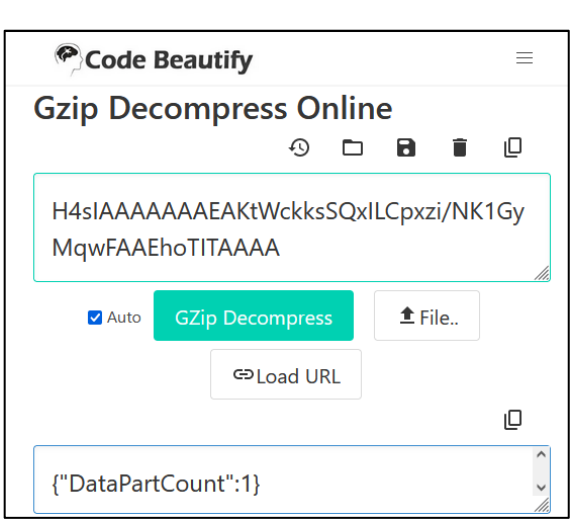
[3] https://www.grammarly.com/about

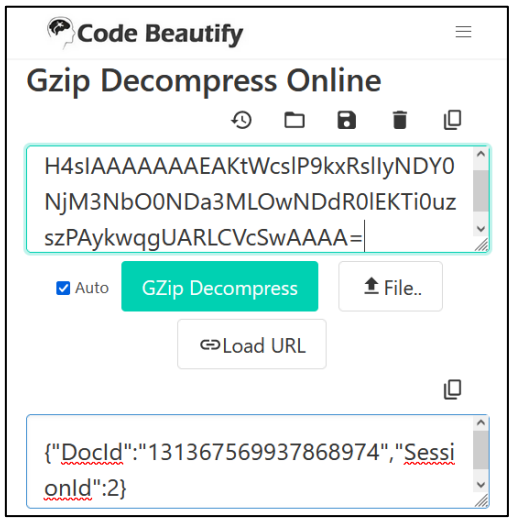
64. Taking this as an example, the highlighted tags are as follows, which I have colour coded for ease of reference:

```
<w:docVar w:name="__Grammarly_42____i"
w:val="H4sIAAAAAAAAEAKtWckksSQxILCpxzi/NK1GyMqwFAAEhoTITAAAA"/>
<w:docVar w:name="__Grammarly_42___1"
w:val="H4sIAAAAAAAAEAKtWcslP9kxRslIyNDY0NjM3NbO0NDa3MLOwNDdR0lEKTi0uzszPAykwqgUARLCVcSwAAAA="/>
```

65. The tags specify "Grammarly" within the name of the tag and each tag includes a long string, coloured above. The strings appear to be formatted in base64, and are not directly human-readable. There are various ways of decoding base64 data, but in this case the encoding is indicated by the first characters, which are "`H4sIA`" in both cases. These characters encode the digits `1f8b` (in hexadecimal) which are **Magic Bytes**. Magic Bytes, also known as file signatures, are characters at the beginning of a file which provide an indication of their file type.

66. In this case the Magic Bytes correspond to the **gZip** file format, which is a compression algorithm broadly similar to the zip compression which will be familiar to most users. This therefore indicates that the strings are in fact gZip compressed information which can be decoded by decompressing them in a standard way. There are many tools to decompress gZip, which is an open and widely used format, but it is convenient to use an online graphical tool[4] to show the conversion for each of the two strings above, as follows:





[4] https://codebeautify.org/gzip-decompress-online

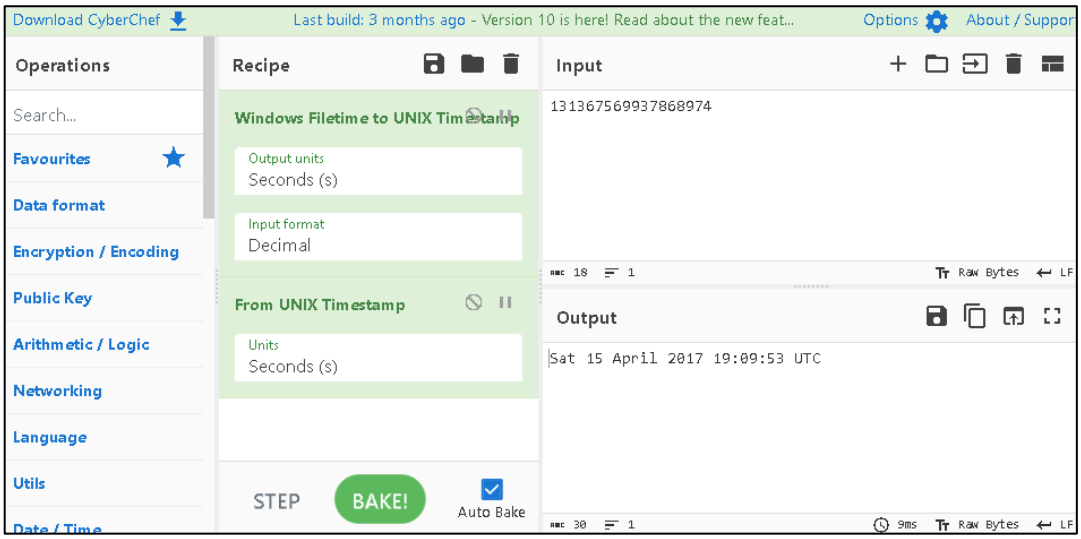
67. Thus the encoded data is revealed to be as follows:

H4sIAAAAAAAAEAKtWckksSQxILCpxzi/NK1GyMqwFAAEhoTITAAAA	{“DataPartCount”:1}
H4sIAAAAAAAAEAKtWcslP9kxRslIyNDY0NjM3NbO0NDa3MLOwNDdR0lEKTi0uzszPAykwqgUARLCVcSwAAAA="/>	{"DocId":"131367569937868974","SessionId":2}

68. Referring to the explanation of timestamp formatting above, the decimal number in the green text can be recognised as corresponding to a Windows FILETIME format timestamp. Converting the timestamp using the inbuilt function within MS Windows indicates that it decodes to 15 April 2017 at 19:09:53 UTC:

```
PS C:\> [datetime]::FromFileTimeUtc(131367569937868974)
15 April 2017 19:09:53
```

The same function can conveniently be conducted online via a graphical interface, at the following URL, as shown below with the same output:
https://gchq.github.io/CyberChef/#recipe=Windows_Filetime_to_UNIX_Timestamp('Nanoseconds%20(ns)','Hex%20(big%20endian)')From_UNIX_Timestamp('Nanoseconds%20(ns)')

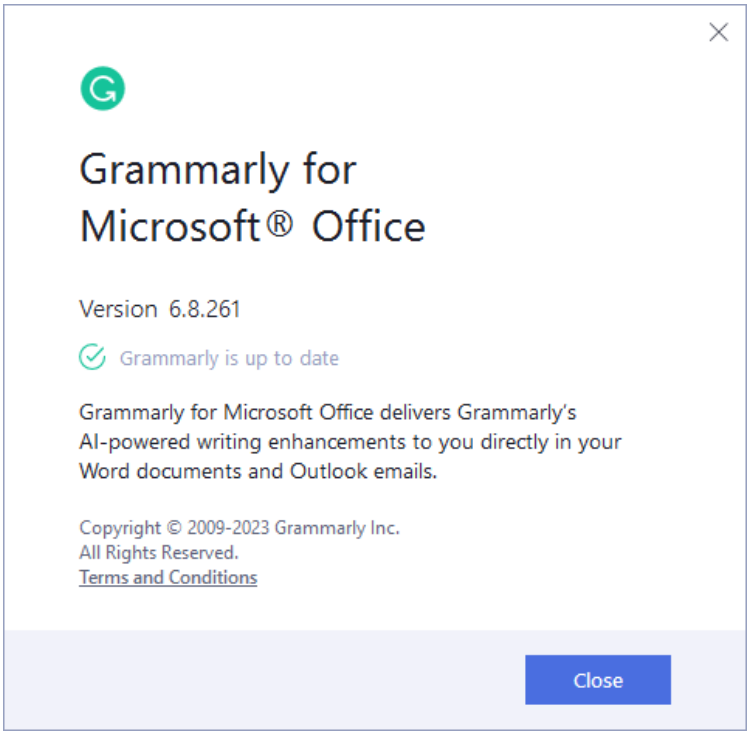


69. The resulting timestamps have cast into doubt the authenticity of several documents in the disclosure dataset. I have therefore also investigated the functionality of Grammarly that leads to the creation of this timestamp, to ensure I fully understood it.

70. I created a virtual machine on which to operate a Grammarly instance. As I did this in 2023, I attempted to obtain a more contemporaneous version of the Office addin for Grammarly from which this timestamp appears to relate. As Grammarly was only founded in 2009 and its Office addin was not released at that time, it was not possible to download a Grammarly installation contemporaneous to the purported creation date of the documents. However, I did obtain several historic Grammarly installation files from Wayback Machine archive snapshots of the Grammarly website ranging from 2011 through to 2020.

71. It was not possible to utilise the majority of the older releases because of the way that Grammarly works. While the software would install, it was not possible to operate it fully without authentication, or without an internet connection. I observed that Grammarly works by uploading a copy of the information being processed to a Grammarly server and responses sent back to the local computer. Although I was able to register valid credentials, older versions of the software did not work and reported a failure to connect to Grammarly's server, even when an internet connection was enabled. This is consistent with the older versions of the software attempting to connect to servers which are out of use and have been replaced.

72. The earliest version that I obtained which would successfully authenticate was from a Web Archive snapshot dated 09/12/2020 at https://web.archive.org/web/20201209065448/https://download-windows.grammarly.com/bundle/GrammarlyBundleSetup.exe . The installation file "GrammarlyBundleSetup.exe" computes an MD5 hash value of eed1b7336536dce0acbb35f127ee7125. Using a virtual machine configured with a basic installation of Microsoft Windows 10 and Microsoft Office 2019 Pro plus, I undertook a process of creating multiple documents and auditing the results. After completing the installation of Grammarly, it was necessary to log into the Grammarly service using a set of valid Grammarly credentials. Upon authentication the software updated automatically to the latest release:



73. It has therefore not been possible to conduct the testing with an earlier version of the Grammarly software. I proceeded with the latest version of the software as listed above. Observing how Grammarly interacted with documents, I learned the following characteristics of its behaviour:

    a. The software did not trawl automatically through pre-existing files and folders and automatically insert its record into user documents. It only interacted with documents at the user's request.

    b. The Grammarly addin for MS Office would launch shortly after a document is opened in MS Word, but it would not actively scan the document until specifically clicked on.

    c. If the user does not specifically save the changes made to the document after operating the Grammarly software, no changes would be committed to the document, and neither would the Grammarly DocID and associated information be added to the file.

    d. Therefore, I was not able to find any way to cause similar tags to be embedded into a document without both (A) instructing Grammarly to review the content of the

document and (B) afterwards, saving the document to commit the changes to the file.

e. The Grammarly timestamp appears to correspond to the time that the Grammarly addin is launched within MS Word. The timestamp is generated by the local computer clock. If the local clock is inaccurate, then the timestamp would be incorrect.

f. However, the tag is not later updated if a document is opened again. Once committed to the document file, the Grammarly timestamp remains constant. Future document editing or review with Grammarly does not update this part of DocID field.

g. The recorded session ID number however does increment with subsequent uses of Grammarly.

h. If a Save As was performed to generate a new document, this would inherit the timestamp from the original document.

74. These findings were consistent also with my experience of how files are authored and saved within MS Word and with the face-value content of the Grammarly tags.

75. When analysing Grammarly timestamps in the disclosure dataset, I proceeded on the basis of these observations and findings.

Time zone offsets

76. It is important to bear in mind the time zone in which a timestamp is recorded.

77. It is not always possible to determine with confidence what the clock (on the computer used to interact with the time zone) would have displayed at the time the timestamp was recorded. The manner in which a time zone is displayed can also add complications:

a. In some cases, especially when viewing the raw text metadata in a file, the timestamp will normally reflect the time shown on the clock of the computer used to create a document. It may or may not record a time zone offset. Some metadata fields will simply default to Universal time without indicating any time zone information.

b. When a timestamp is not being viewed directly in the raw internal data but is being viewed an application or interpreter, timestamps may sometimes be 'interpreted' and displayed to the user in their own current local time. This depends on the software in question.

c. Since some programs (and file formats) do record time zone offsets and other programs (and file formats) do not, it is not always possible to know the time zone in which the timestamp has been recorded at all.

78. In general, I always keep in mind the relevance of time zones when considering timestamps, and have done so in the present case, and I am alert to the possibility for time zone offsets to explain what might otherwise seem like a suspicious irregularity in a document. It is often easy to spot such instances because time zones are almost always offset in whole-hour increments (with some limited exceptions where offsets include 30 minute increments), such that examining the minutes part of the time stamp can help to resolve differences. In the relatively rare cases where the time zone offset is or could be of relevance to the analysis in this way, I have endeavoured to make a note of it when it comes up. My localisation settings have been set locally to the UK throughout my analysis (other than in one instance where localisation settings were relevant), and the prevailing time zone has been BST which is UTC+01. Often when viewing readily-available metadata through an application, it will be displayed in BST as a result of this.

79. In other circumstances, the time zone does not make any difference:

a. In many cases, the time is not important; for example if it is only the date or year which is important to the analysis then the time (and thus the time zone offset) can be ignored.

b. In other cases, what matters is not the precise time shown on the clock, but how different timestamps within a document (or in different documents) compare to each other. In those cases, the specific time zone offset does not matter so long as it is possible to ensure that the timestamps being compared are sourced in the same or sufficiently similar way, such that they do provide an accurate point of comparison.

80. Where time zones do not matter to my analysis, I have tended to avoid mentioning them in my Report to avoid adding additional complexity to what is already a long and detailed set of Appendices.

The Load file metadata

81. As I have explained above, the external metadata is not carried with the document when copies are made. The primary source of external metadata would be the computer systems and forensic images from which the document was extracted. In this case, since those were not provided, the external metadata was provided in a schedule listing the properties for each document. This schedule is called a "load file" and is typically created to be ingested with the disclosed documents in order to re-associate the file properties with the source documents.

The load file and problems with the data provided

82. In the present case I have been provided with one load file for each of Vol001, Vol002 and Vol003.

83. I have found that the manner in which the data has been provided is not akin to a logical forensic image (which would retain all of the internal metadata intact with the dataset in an auditable tamper evident file), and does not include a separate catalogue of the external file properties and provenance of each file.

84. The form of the load file is typical format of an export from an e-disclosure database, through which the disclosure data has been processed, and in my opinion does not faithfully record all the original file metadata. I observe that the files are provided in a non-original folder structure, and have been stripped of their original filenames, file property timestamps, and any indication of folder path or from where they were sourced.

85. The load files provided contained a partial record of that information, but many relevant pieces of information were not provided at all. The data that was provided itself had many deficiencies and problems that impeded my analysis. Furthermore, each schedule was constructed in a different manner, with no conformity between the datasets, complicating my review from one set to the other.

Precision

86. As observed above, OS file timestamps are generally very precise, to about a second for older file systems and at the millisecond level for more modern systems.

87. However, the file timestamps in the load file are accurate only to the minute level. This indicates to me that they are not the original time stamps, but have undergone some process of conversion which has led to them becoming less precise, and data therefore becoming lost. I do not have any method to verify the process by which these timestamps were created and do not know how the precision was selected and what rounding was applied.

Discarded time zone information, and trying to interpret the metadata provided

88. I was able to identify that the load file timestamps were not original and had apparently been separated out into different partial timestamps and were required to be reconstituted before they could be understood and analysed.

89. For example, I set out below a partial set of the load file information that was provided for documents ID_000757 and ID_000835:

{ID_000757}
{ID_000835}

Production Begin Bates	Date - Created - Date	Date - Created - Time	Date - OS Created - Date	Date - OS Created - Time	Date - Time - Created
ID_000757	06/08/2009	27/10/2022 06:15	09/11/2015	27/10/2022 09:31	06/08/2009 06:15
ID_000835	24/10/2008	27/10/2022 20:24	24/10/2008	27/10/2022 21:24	24/10/2008 20:24

90. As can be seen there were several fields with very similar labels (including the same words in different orders) and no explanation of what was meant by each of them was provided.

91. I approached this first by verifying the metadata timestamps for each document by manually inspecting both for their internal 'Created' timestamps. After confirming that the internal timestamps matched the last column "*Date - Time – Created*" I used this as a baseline to establish how the other columns were constructed.

92. I found that for the two columns "*Date - Created - Time*" and "*Date - OS Created - Time*", the dates in these fields (highlighted in red above for ease of reference) did not appear to be reliable. However, the hours and minutes within those fields matched the *Date-Time-Created* field. I therefore took this field to consist of *date* data which was not relevant but *time* data which was said to be relevant. It was therefore necessary to strip out the parts marked in red above.

93. Having isolated the time part of the metadata field, these resulting times are apparently to be applied to the "*Date-Created-Date*" field and reconstituted to form a single timestamp.

94. It can be seen however that these do not always align perfectly and time zone offsets have not {ID_000835} been applied to the OS timestamps – this being clear from the example of ID_000835. After reconstituting the *"Date-OS Created-Date"* and *"Date-OS Created Time"* fields, the resulting timestamp shows the Created Date as being exactly one hour later than actual date of creation of the file (21:24 instead of 20:24).

95. It is therefore my understanding that the OS file metadata provided may not have reliably taken into account the time zone of creation of the document, but has instead applied the timestamps across all documents without accounting for time zone offsets. This often leads to apparently contradictory metadata as indicated in the example above, which initially looks irregular and can only be explained by making assumptions about discarded time zone information. Where the time zone offset has been lost at the source device, it would normally be possible to confirm this by reverting to the forensic image collected from the device. Without access to these forensic images, I cannot independently verify the accuracy of the provided information.

96. The example above is an illustrative example where the offset is very clear. However it is most often not so simple to form a view about what the data is intended to indicate: in the case of {ID_000757} ID_000757 for example, the OS Created timestamp is greatly different from the metadata timestamp, even after the reconstituting process I mentioned above.

97. The fact that other timestamps are plainly not correct leads me to doubt the accuracy of this {ID_000757} information, but in cases such as ID_000757 I have no method to validate whether or not these are correct and have not therefore been able to assess the timestamps in such cases as part of my review.

Later schedules provided via Travers Smith

98. I was later provided with two additional supplementary schedules of data that I understand were provided by Travers Smith to address some of these deficiencies. However, these supplements were also inconsistent. For example, it can be seen in the below screenshot extract that the dates have been formatted in different ways in adjacent rows, and the times alternate between 12 and 24 hour clock configurations:

11/11/2015 18:26	11/11/2015		6:26:32 PM		
07/11/2015 23:08	07/11/2015		11:08:42 PM		
6/15/2011 11:54	6/15/2011		11:54:11 AM		
04/02/2015 03:16	04/02/2015	12/02/2015	03:16:00	01:40:12	
04/02/2015 03:21	04/02/2015	12/02/2015	03:21:00	01:40:12	
04/02/2015 03:37	04/02/2015	12/02/2015	03:37:00	01:39:54	
04/02/2015 03:39	04/02/2015	12/02/2015	03:39:00	01:39:51	
04/02/2015 04:12	04/02/2015	12/02/2015	04:12:00	01:39:39	
04/02/2015 04:14	04/02/2015	12/02/2015	04:14:00	01:39:35	
12/10/2015 14:51	12/10/2015		2:51:00 PM		
8/31/2015 15:12	8/31/2015		3:12:48 PM		
10/24/2010 7:32	10/24/2010		7:32:10 AM		
08/09/2015 12:03	08/09/2015		12:03:43 PM		
5/27/2014 8:32	5/27/2014		8:32:32 AM		
10/13/2015 9:32	10/13/2015		9:32:39 AM		
3/24/2015 16:04	3/24/2015		4:04:28 PM		
11/13/2015 8:13	11/13/2015		8:13:54 AM		
10/01/2014 14:28	10/01/2014		2:28:00 PM		

99. While it is possible, with some effort, to reformat these into a more consistent formatting (and I have attempted to do so where the point has arisen in relation to a document I have analysed) I cannot however be sure that I have always correctly applied a 12 or 24 hour clock and I am not able to rely on this information fully. It also has the same problems with time zones that I have explained above. As such, I have no method by which I can validate that the information provided to me is accurate and has not already been muddled or mixed up.

100. Overall, the accompanying schedules give the impression that the dataset has been exported from multiple different e-disclosure platform sources and combined in a way that was not consistent and which appears to have discarded important information.

101. I have therefore been able to rely upon this provided metadata only very little. In cases where I have needed to consult the provided metadata to aid my analysis, I have either taken the metadata provided at face value (which should be understood as a limitation in the

circumstances), or I have done my best to try to account for time zone variations. In doing so, I am informed by Bird & Bird that relevant time zones may include Australian time zones (as Dr Wright was previously resident there), which I have taken into account, but for many of the document types the relevant time zone at the time of authorship cannot be reliably determined and therefore cannot be reliably applied.

Different approaches to metadata used by various technologies

102. When discussing Internal Metadata, it is important to note that different file types record and store metadata in different ways. They also store data in different ways, therefore when converting between formats it is not unusual to incur defects or changes to the way that the content was displayed or organised on the screen.

103. Even within the same File format, there are often different ways in which the content can be structured and stored. Similarly, MS Word .DOC and .DOCX files, though related, are in fact entirely different formats as I explain below, and the PDF file format can incorporate a variety of different encodings into a single document. I have set out below some of the differences relating to each of the main file formats of relevance and some of the main metadata fields associated with each type of document.

104. Different products, and even different versions of the same product can also behave differently when creating content or applying changes to a document. For example, the behaviour of the metadata timestamps and Edit Time count of an MS Word document behave differently to those of an MS Excel spreadsheet or MS Powerpoint presentation.

105. Other documents contain very little metadata at all, such as plain text (.TXT) files, some image documents, and documents which are scanned from hard copies.

FILE FORMATS / TYPES OF DATA FILES PROVIDED FOR ANALYSIS

File formats - overview

106. When I refer to ‘file formats’ and “file types”, I mean the way in which data in a document file is encoded. Different types of file encode data in different ways, and require the use of different software applications for the data to be viewed. For example, PDF files are typically viewed in a PDF viewer like Adobe Acrobat or Adobe Reader, and .DOC and .DOCX files are typically associated with versions of MS Word. The file type is conventionally indicated by a suffix in

the file name called a file extension, such as ".pdf" or ".doc" but this is not always the case, and the file extension can be edited by a user.

107. During this investigation, I have identified and analysed a number of different file formats. The majority of the documents I have examined in detail are common formats including MS Word documents, and PDF files. There are also a number of email files that have been produced as individual MSG format files.

108. I describe below the properties of these most common files. There have been other formats that I have analysed that are somewhat less common, such as TIFF image files, a proprietary accounting database format called "MYOX", which are not as broadly applicable to my overall analysis and therefore which I address as they come up in the individual Appendices.

PDF file format

109. The PDF (Portable Document Format) file is a common format used for the publication of a variety of documents and PDF files can be used to preserve and present a multitude of data types. It is a very flexible format, and a PDF file can be used to present anything from a written text document, a scan of a multi-paged hardcopy document, a multimedia presentation, or an interactive form where the user is required to complete certain sections and electronically sign.

110. It is commonly used to share completed documents such as reports, contracts or invoices where there is an intention that a document is intended as the final version. They are not as readily editable as other file types and are typically used where there is an expectation that the content of the document will not need to be changed or edited. Rather than being designed primarily for editing, the format is also designed to appear in the same way on a range of different devices, which is the reason for its name "portable", so it is well suited for publishing of documents that could be viewed on a number of devices such as newsletters, contracts, or reports for example.

111. There are many different methods by which a PDF file can be made. They can often be exported directly from an editing application, and recent MS Office applications can create them directly by conducting a "Save As" and selecting PDF as the output file format. There are also applications available that install as a virtual printer on the computer, and the PDF is effectively 'printed' from any application that supports printing, allowing it to be created from

a wide range of inputs. There are also a raft of utilities available that can mark-up or alter the content of a PDF file. The Adobe Acrobat[5] suite includes such a function.

How data is stored within PDF files

112. Data within PDF files is stored within "streams" of content. Each encoded stream describes part of the content of the pages of the PDF, such as the text, pictures, drawings and any other content, describing (among other things) what the content is and the position it should be displayed on the page. When viewing a PDF, the viewing program interprets each stream and displays it in accordance with the encoded instructions. Simple PDF files may have just one stream to encode all the content in the document, but this is not a requirement, and depending on how the PDF is created different streams may be used for each page, or each part of a page, or any other variation. When the content of a PDF is edited, this may create a new stream for the edited content, or a pre-existing stream may be directly edited.

113. Content within a PDF is not typically stored in human-readable plaintext data but is usually encoded. This will often be via encoded numerical strings to represent each character. Those strings can be decoded via a character map table or "CMAP" which functions as an embedded codebook that pairs up numerical strings to each character they are intended to represent. There are often multiple cmaps in a document, one for each stream.

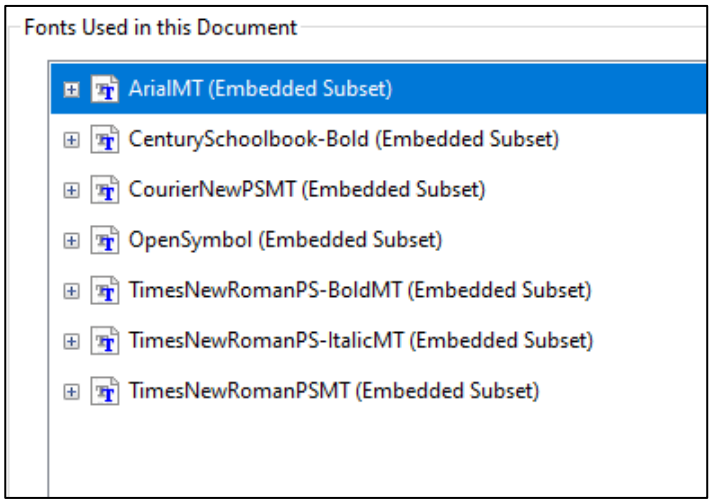
114. Streams are typically also compressed, or "deflated", within the PDF's raw data, and need to be expanded before they can be viewed. Inflating and deflating the streams is a simple compression and decompression process that does not alter their informational content, only the format in which it is displayed (and the amount of disk space it takes up when the file is saved). Inflating the streams of a PDF usually results in the encoding discussed above becoming viewable for text streams for example, in the format of numerical strings which can then be decoded further.

Embedded files within PDF files

115. A PDF may also contain other files embedded within it (a different sense of "embedded" to that used above), whereby a whole additional file is encapsulated within the PDF, retaining its own encoding. This functionality is very commonly used to embed font files within a PDF, allowing the PDF to display in the same way using the author's chosen font, even if that font is not installed on the viewer's machine. Instead of reading the font information from the font

[5] Adobe Acrobat is the paid for equivalent to the Free Adobe Acrobat Reader application that is free to use.

directory on the viewing user's machine, the PDF viewer will read font information from those embedded files. This is illustrated in the following screenshot:



Partial screenshot of Adobe Acrobat document properties for the Bitcoin White Paper showing embedded fonts. As well as common fonts like Arial and Times New Roman, somewhat less common fonts such as Century Schoolbook and OpenSymbol are embedded, allowing content in those fonts to display in the same way on any machine regardless of whether the viewing user has those fonts installed.

Types of metadata in PDF files

116. The following types of user-visible metadata are typical of PDF files:

a. The following **timestamps**, which are created depending on the date that is set on the local computer at the times the relevant events took place:

i. **Created:** the date that the PDF was created,

ii. **Modified** or **Last Modified**: when the document was modified (in some circumstances);

iii. **Metadata Date**, to indicate the date on which the internal metadata fields were written

b. Optional user-set document information about the document, such as the **Title**, **Subject,** relevant **Keywords**;

c. **Application** information, indicating the application software used to create the PDF (also called "Creator" or "CreatorTool");

d. **Producer** information, indicating the underlying software library that was used by the Application to encode the PDF;

e. **Document ID or UUID (universally unique ID)**, a long string of characters intended to uniquely identify the PDF file in question;

f. Digital signature, and related information such as the signature date if a PDF is signed digitally;

g. **Bookmarks**, used for navigating the PDF file.

117. This is not rigid. Not all of these are required, and different PDFs will contain different fields. In addition, there is the possibility for inclusion of custom fields.

XMP within PDF files

118. Metadata in a PDF is typically encoded in "XMP" form. XMP stands for (eXtensible Metadata Platform) and is a protocol for encoding serialised metadata in digital documents in the form of XML (eXtensible Markup Language) tags, originally created by Adobe and since ISO standardised.

119. Different versions of the XMP specifications have been released over the years, and different versions of the software toolkit used to create XMP data are known as the "XMP Core". Each new XMP Core version is assigned a new version number, and a timestamp relating to its date of its creation.

120. To illustrate how XMP metadata presents, I show below the XMP metadata stream of one of my exhibits which happens to be a PDF (in this case I have chosen **Exhibit PM1.13**, simply because it is a PDF file containing a useful metadata stream to illustrate how it looks without the content being important). I have removed extraneous whitespace to aid review and added highlighting for parts I comment on directly below: {H/14}

```
<?xpacket begin="ï»¿" id="W5M0MpCehiHzreSzNTczkc9d"?>
<x:xmpmeta xmlns:x="adobe:ns:meta/" x:xmptk="Adobe XMP Core 5.6-c018
91.98c2f96, 2021/06/15-20:39:32        ">
   <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
      <rdf:Description rdf:about=""
            xmlns:xmp="http://ns.adobe.com/xap/1.0/"
            xmlns:xmpMM="http://ns.adobe.com/xap/1.0/mm/"
            xmlns:dc="http://purl.org/dc/elements/1.1/"
            xmlns:pdf="http://ns.adobe.com/pdf/1.3/"
            xmlns:pdfx="http://ns.adobe.com/pdfx/1.3/">
         <xmp:ModifyDate>2023-07-26T17:51:06+01:00</xmp:ModifyDate>
         <xmp:CreateDate>2023-07-26T17:51:04+01:00</xmp:CreateDate>
         <xmp:MetadataDate>2023-07-26T17:51:06+01:00</xmp:MetadataDate>
         <xmp:CreatorTool>Acrobat PDFMaker 20 for Word</xmp:CreatorTool>
         <xmpMM:DocumentID>uuid:15ade5d0-00a2-439a-8972-
9d14df8786f6</xmpMM:DocumentID>
         <xmpMM:InstanceID>uuid:a91e0b41-6ebe-45e1-8f07-
aff0cbbfbab5</xmpMM:InstanceID>
         <xmpMM:subject>
            <rdf:Seq>
               <rdf:li>4</rdf:li>
            </rdf:Seq>
         </xmpMM:subject>
```

```
            <dc:format>application/pdf</dc:format>
            <dc:title>
               <rdf:Alt>
                  <rdf:li xml:lang="x-default"/>
               </rdf:Alt>
            </dc:title>
            <dc:description>
               <rdf:Alt>
                  <rdf:li xml:lang="x-default"/>
               </rdf:Alt>
            </dc:description>
            <dc:creator>
               <rdf:Seq>
                  <rdf:li/>
               </rdf:Seq>
            </dc:creator>
            <pdf:Producer>Adobe PDF Library 20.5.233</pdf:Producer>
            <pdf:Keywords/>
            <pdfx:SourceModified>D:20230726163000</pdfx:SourceModified>
            <pdfx:Company/>
            <pdfx:Comments/>
         </rdf:Description>
      </rdf:RDF>
   </x:xmpmeta>
   <?xpacket end="w"?>
```

121. In the highlighted sections above it is possible to see the XMP Core version used, as well as the basic metadata properties of creation date and modified date for the document itself, the "Metadata" date (the timestamp that the metadata itself was created, which matches the native Modify Date), and the software tool used to create the document (in this case, PDFMaker for Word 2.0). It can also be seen that the first three highlighted timestamps are encoded with a time zone offset of UTC+1, which is British Summer Time (the prevailing time zone in which I am located at the time of writing this Report).

122. In the last highlight, the SourceModified tag, it can be seen that the PDF was created from an underlying source which was itself modified at 16.30 on the same date. However, that timestamp is encoded differently in Universal time, and so does not include the UTC offset. Therefore that timestamp (16.30) is around 21 minutes before the Creation / Last Modified timestamps, indicating that the exhibit was created in MS Word first, around 21 minutes before being exported to PDF.

123. PDF Metadata can also be included in simple properties tags which do not follow the XMP specification, which is true of the example of the Bitcoin White Paper ID_000865, which simply lists its properties in tags as follows (which is interpreted from the byte-level content of the file and which I have formatted for ease of reading):

```
)  <<

      /Creator: .W.r.i.t.e.r
```

```
      /Producer: .O.p.e.n.O.f.f.i.c.e...o.r.g. .2...4
      /CreationDate(D:20090324113315-06'00')

 >>
```

DOC/DOCX file formats

124. MS[6] Word is one of the most commonly used document editing programs. It mainly uses the .DOC or .DOCX file formats.

125. The older .DOC file format was the typical format used by MS Word for saving documents before 2007. Compared with the newer .DOCX format I describe below, the .DOC format was inefficient and cumbersome. It has been identified many times as retaining unnecessary (and often embarrassing and unintended) remnants of information hidden with redundant space[7] of the file. This additional information often provides useful information for forensic review.

126. In 2007, the more recent .DOCX format was introduced. This format was very different to the .DOC format that preceded it and was more reliable and streamlined. The "X" in DOCX refers to XML. This is the same markup language also used in XMP Core that I describe above. In the case of DOCX files, a great deal of content is encoded in XML format. .DOC files did not use XML, and saving a pre-existing .DOC as a .DOCX file will alter the layout and formatting of the raw data within it to match the newer file format. Such a conversion will also affect the face value presentation of the document.

127. Both file formats .DOC and .DOCX are typically used to author written works and can contain a number of different media types within them. They are commonly used to prepare documents that are intended to be printed on paper, or to be further edited, or to be exported into PDF files for electronic dissemination. They are somewhat less suitable for publishing documents, because they are readily editable and the content of them can look different depending on the viewer's computer system (e.g. depending on whether they use Mac or Windows operating systems, or whether they have certain fonts installed).

[6] "MS" standing for Microsoft

[7] I use the expression **redundant** space to cover areas of a file that do not contribute to the active and current face of the content of the file. It is not limited to Word documents - similarly PDF files can sometimes retain redundant data streams where the content or part of the content is no longer part of the printed face of the document.

128. .DOC and .DOCX are not the only file formats used by MS Word. Other formats that it uses or that more modern versions of MS Word are capable of using include:

a. .DOT and .DOTX, the file format of "templates", which can be used to create a document with pre-populated content, formatting, and content fields but which are otherwise blank and can be edited and saved as a new file (without changing the content of the underlying .DOT or .DOTX template),

b. .ODT, the OpenOffice.org open format which I describe below and which is also editable in MS Word,

c. .PDF, which can be imported into MS Word and edited directly within it, albeit that this involves a decompression and conversion process which is often imperfect for anything more than the most simple documents, and

d. Various other text-based file formats, like plain text (.TXT), Rich Text format (.RTF) and web page editing (.HTML).

Versions of MS Word

129. There have been different versions of MS Word released over time. Each version has a product name, such as "Microsoft Word 2003" or "Microsoft Word 2007", which are displayed to the user of the program and which would appear for example in the "About" dialog if viewed within the viewer.

130. Somewhat less widely known and somewhat less visible to the user is that each version of MS Word also has an internal version number such as "11" (in the case of Microsoft Word 2003) and "12" (in the case of Microsoft Word 2007).

131. Over the course of its lifetime, each version of Microsoft Word was updated. In the case of Microsoft Word 2003 (which is the only case in which updates to the program are relevant for my Report), this was done in a series of Service Packs, called SP1, SP2 and SP3. The version number of MS Word 2003 updated with each service pack, and the internal version number for Word 2003 SP3 was "11.9999". Service Pack 3 (SP3) for MS Word 2003 was released to coincide with the release of MS Word 2007 and was therefore the final service pack for MS Word 2003 (as indicated by the number itself).

132. The version numbering for more recent versions, from MS Word 2007 (Internal version "12"), is less granular being that even with regular updates to the program, the metadata still records the version number as a whole number.

Metadata within MS Word documents

133. Metadata within MS Word documents is typically encoded:

a. in XML for DOCX format, similar to the example shown above in respect of PDFs, and

b. in a different format, OLE, for DOC formats. The details of OLE encoding are not important for this Report save to say that they are very different to DOCX.

134. MS Word documents typically include internal metadata including the following:

a. The following **Timestamps,** which are created depending on the date that is set on the local computer clock at the times the relevant events took place:

i. **Creation date**: when the document was originally created (or if created via a 'save as' operation from a precursor document, the date of Creation of the precursor document) .

ii. **Last Save date (also referred to as Last Modified date)**: when the document was Last Modified and modifications committed to a file by performing a "save" operation. Each time the document is saved, the Last Save timestamp will be updated.[8]

iii. **Revision number**: The number of times the document has been saved since Creation. This will typically begin at 1, but in the case of a "save as" operation being used to create a document, it will typically begin at 2.

b. The following **Account Information**, which is read from the name associated with the user account ascribed to MS Word that interacted with the document. These fields are

[8] Other operations can cause the Last Save date to be updated, for example, if consecutive save operations are conducted within a minute of each other, in some cases it can cause the last saved property to be incremented by 1 minute. This sort of functionality does not appear to be directly relevant to my analysis, but I mention it for clarity. The accounts of the properties here are given by way of technical background and it is not possible to address every detail of their operation, particularly where it does not appear to be relevant.

not directly editable, but the setting from which they are read is editable. This is because the name associated with the user account is typically supplied by the user typing it in, and can be altered easily by simply editing the account settings:

i. **Author**: The name setting associated with the user account at the time the document was created.

ii. **Last Saved By**: The name setting associated with the user account at the time the document was last saved.

c. The following **user-editable content fields,** optional fields which can be typed in by the user and describe or annotate the document content:

i. **Title**: The title of the document, which is user-editable.

ii. **Subject**: A brief description of the document's content.

iii. **Keywords**: A list of keywords or phrases that describe the document's content.

iv. **Category**: The type of document, such as a report or memo.

v. **Comments**: Any comments or notes added by the author or other reviewers.

d. In addition, MS Word records an Edit Time during which the document has been edited, as set out below.

Microsoft Word Edit time

135. While a Word document in Microsoft Word is open, the "**Edit Time**" metadata counter increments each minute that the document is being edited.

136. A document counts as being "edited" for this purpose if it is the primary document open in MS Word. More specifically, for the timer to increment:

a. The document must be open in the last Microsoft Word window that was active and in focus. A window that is "**Active**" is the one that is currently being

interacted with by the user and which is receiving input. The term "**In focus**" refers to the window that is currently receiving keyboard input (and is invariably the same as the Active window).

b. It does not matter whether a user actually types into or makes changes to the document, as the Edit Time will increase regardless.

c. If a second document open in MS Word is switched to, then the Edit Time of that second document will increase instead of the first, until switched back. "**Switching to**" a window is the process of changing the active window, usually by clicking or pressing Alt-Tab.

d. I emphasise that the document Edit Time increases depending on the window that was last active / in focus in MS Word. It is not affected by using different applications alongside it, so for example if a web browser is switched to and typed into while a document is open in MS Word in the background, the Edit Time property for that MS Word document will continue to increase until the document is closed or until another different MS Word document is switched to.

e. I have referred to this in my report as "exclusive use" for the purpose of document editing in Word.

137. Certain other activity can cause the Edit Time to increase:

a. If a document is saved and then re-saved, it can cause the Edit Time property to increase by 1 minute even if the time between them is less than 1 minute.

b. If the operating system clock suddenly jumps from one time to another between saves while an MS Word document is the most recently active document, that will usually cause the Edit Time property to jump with it, and record as if the whole intervening time had been spent editing. Sudden jumps in clock time can occur if a computer is hibernated (a method of turning off a computer so that the user session is saved, and then restored at the point it is turned back on). They can also occur if a computer clock setting is deliberately edited, as I explain in detail later on in this Main Report.

138. It should therefore be impossible for two MS Word documents to record Edit Times that overlap with one another, if they are edited on the same device. In the present case, I have not been provided with detailed information about the source devices of any documents (although I understand that this was requested), so have proceeded on the basis that documents that are similar in character and bear similar user data and other metadata are likely to have been created or edited on the same machine.

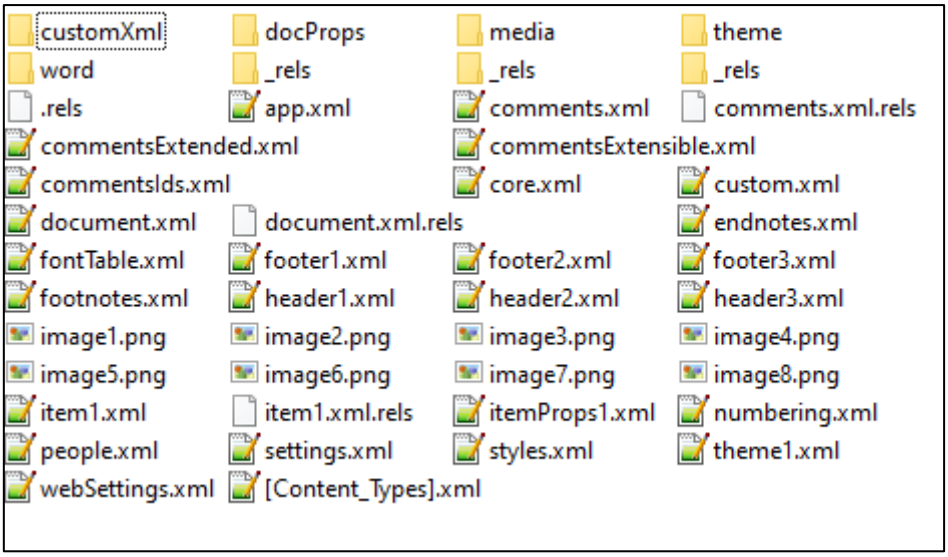
139. I caution that while other documents such as MS Powerpoint have similar-looking Edit Time metadata fields, the way they record metadata can be very different and this analysis should not be applied to other similar looking fields outside MS Word.

DOC and DOCX as "compound files"

140. Though saved as single files, both DOC and DOCX files are in fact "compound files" made up of various constituent parts (as are many other file types). This is similar in concept to a ZIP file, whereby one single file is actually a container for copies of multiple other parts within it.

141. When a DOC or DOCX file is opened, it is actually opening a container that holds the different parts. These can include parts relating to the main document content, formatting information, images and objects embedded within the document, and components relating to metadata information alone.

142. To illustrate the various files and directories within a DOCX compound file, the view below shows the view of this report document itself when the DOCX is opened in an archive viewer (in this case the widely-used viewer 7-ZIP, though other viewers such as WinZip could also be used):



The various constituent parts of this report document, including XML images, directories and other files, as viewed within an archive viewer

143. In some cases, MS Word documents can also contain old, partial embedded files from previous drafts of the document, such as ZIP-encoded files which themselves contain additional data not readily viewable. When it is possible to identify and isolate these, viewing those files can provide insights into the history of the document's creation.

OpenOffice

144. OpenOffice.org ("**OOo**") is a suite of office software very similar in functionality to MS Office, but which is open-source and freely distributed. It contains various applications within it, such as "OOo Writer" (equivalent to MS Word), "OOo Calc" (equivalent to MS Excel), "OOo Math" (a tool for creating mathematical formulas) and "OOo Impress" (equivalent to MS Powerpoint).

145. Of these, only the OOo Writer application is relevant to my analysis so I do not address other parts of the OOo suite. The file format mainly used by OOo Writer is "ODT" (standing for "Open Document Text"), which is equivalent to MS Word .DOC or .DOCX files. OOo Writer is also compatible with use and editing of .DOC and .DOCX files.

146. OpenOffice ODT files are also compound documents, very similarly to the examples shown above, though they use different constituent parts.

{ID_000254} {ID_000260} 147. I have analysed two specific ODT documents from the disclosure dataset, ID_000254 and ID_000260, and address the above topics in more detail in the course of the relevant Appendices.

Emails and different formats

148. There have been many email files provided within the disclosure dataset which I have analysed.

149. Various file formats used for storing emails include MBOX, EML, PST, and MSG. These all relate to emails, but they differ in their structure and the software applications that support them:

a. **MBOX** is a file format used for storing multiple email messages in a single file. Multiple messages are stored sequentially, separated by blank lines, with metadata and header information included alongside each email. MBOX was originally developed for Unix and Linux systems and is an open format, which has since become much more widely supported. Attachments may be embedded within MBOX files.

b. **EML** is a file format used for storing email messages as individual files. It is different to MBOX because each message is stored as a separate file, usually with a .EML file extension. EML files contain the entire message, including the header information. Attachments may be embedded in each file.

c. **PST** (also OST) is an Outlook Data File, a file format used by Microsoft Outlook to store email messages, contacts, calendar items, tasks, notes and other data. Within Outlook, emails are typically stored together in a single PST file relating to each account in use, which may be a very large for accounts containing a lot of information. This is similar in concept to the way MBOX files store multiple emails together, but includes a larger range of different types of information. However, the Outlook PST format is proprietary to Microsoft and more complex than MBOX. (If stored offline, a PST file can also be named “OST”, which is a synchronised copy of the PST and is not materially different to a PST.)

d. **MSG** is a file format used by Microsoft Outlook to encode individual email messages. Similarly to EML, it stores emails as different files, but it is a different format. The format of storage is OLE file format similar to that used in the older MS Word .DOC format. Unlike PST files, MSG files contain only a single message and any attachments,

rather than a collection of messages. However, emails are typically not natively stored within Outlook as MSG files, but as part of the aggregated PST/OST. Typically, MSG files are only created from Outlook at the point that they are exported, for example by a "save as" operation or by dragging and dropping the email from outlook into a file folder. At the point that an email is exported in this way, it undergoes a conversion process which does not always preserve the metadata associated with the email.

150. PST and MSG are therefore file formats used specifically by Microsoft Outlook for organizing and archiving email messages and for exporting or converting messages when using a Microsoft Exchange email account. MBOX and EML file formats are used for storing email messages in a more portable and standardised way, and in particular are used by Gmail accounts.

Loss of metadata when converting emails

151. I have set out the different types of email files as this is important for present purposes. All but 6 of the email files that are included in the disclosure dataset have been presented in .MSG format. There are indications that in many cases this was not the original format in which they were sent or received. Rather, they seem to have been created by the process of conversion and export which has reduced the metadata within them that I am able to review, or not captured data that would be available at source. As I understand the Disclosure Certificate and Disclosure Review Document, the original native formats of emails were taken from Google Gmail accounts by its native Takeout function, which would have been in MBOX or EML format, but they have not been disclosed in that native format. In my opinion, proper disclosure of these emails should have been given by obtaining them natively in EML format as directly as possible from their source.

152. When Outlook is used as a mail application to interact with mail received via a Gmail account, Outlook does not save an original native copy of the emails being read, but instead converts the email data and incorporates it into its own PST format.

153. This matters because MBOX/EML and PST/MSG have similar metadata fields, but the fields and some content is handled differently, and they do not always perfectly overlap. Because of the differences in how metadata is handled, the metadata does not always get converted perfectly, and certain information is often lost. In particular, there is a risk of losing certain data relating to the IMAP protocol, which is a protocol for accessing a remote mail server. This will

include loss of IMAP timestamps which are typically present in Gmail-formatted emails but will not be carried over into the copies converted and incorporated into a PST.

154. When emails are taken from Gmail to Outlook PST, and then converted again into MSG format, there is a double conversion process which can also compound matters. In that situation, various timestamps such as the PR_Created and PR_Modified dates are typically recorded to the date of creation of the conversion to MSG, in place of the original values that were there previously and therefore removing the underlying metadata that would otherwise inform forensic analysis. I note however that the specifics of this conversion process would depend on the specific version of Outlook, or other software, that was used to manage the files.

155. While this approach is not a problem for basic e-disclosure in most cases (where typically the face value content of the document is important), in cases where documents are to be subject to forensic review, it is not sufficient. As a result, the converted or double-converted emails do not appear to have been the original native formats, but downstream copies. Their production in this manner has impaired some of my analysis.

Other conversions of MSG files

156. In other cases, emails have been supplied in a way that indicates different conversions were used. Examples of this include conversion of messages to PDF form, and in the case of 6 emails, apparent conversion from PDF into EML format. It cannot be expected that these documents would capture any digital metadata in respect of their original creation.

Metadata in emails

157. Metadata fields commonly used in emails include the following types. In some cases, these typically originate from the sending or receiving application. In other cases, they are encoded within email headers:

    a. The following fields which will be familiar to most users from sending and receiving emails:
        i. **From**: A field indicating the identity of the sender of the email. In some cases, an email can be sent by one account 'on behalf of' another account, in which case both senders will appear.
        ii. **To**: The email addresses or names of the recipient or recipients.
        iii. **Cc** (Carbon Copy): The email addresses or names of the copied-in recipients of the message.

iv. **Bcc** (Blind Carbon Copy): The email addresses or names of the recipients who receive a 'blind' copy of the message. This is likely to be preserved on emails captured from a sender's "Sent Items", but would be hidden on copies of the email that were sent to other recipients and so would not appear in received copies.

v. **Subject**: The Subject field of the email.

vi. **Various timestamps**: Metadata fields relating to the dates and times the email was Created, Sent, Received by its recipients.

vii. **Attachments**: Information about any files or documents attached to the email (and whether the email does or does not have attachments).

b. The following information which may be less familiar from ordinary email use:

i. **Reply-To**: The email address or addresses to which replies should be sent when pressing "Reply". When this is absent, replies will usually be automatically directed back to the "From" email address, but when present this can direct replies back to a different address other than that from which it was recorded as being sent.

ii. **Message ID**: A unique identifier assigned to the email for tracking and referencing purposes.

iii. **In-Reply-To**: The Message ID reference of the email to which the current email is replying, enabling that email message to be replied to.

iv. **Other references**: A list of Message IDs of previous emails in the thread of emails being replied to, enabling the emails in question to be identified.

c. Among others, the following types of information may also be included in a Transmission header when an email has been sent over the internet or a network, recording details relating to the email's origin, routing, and delivery:

i. **Routing information:** entries indicating the servers through which the email passed on its route from sender to recipient, which often includes timestamps, server names, and IP addresses.

ii. **Signature and authentication information**: information relating to any digital signatures used to verify the authenticity and integrity of the email, plus results about any authentication used during the email's transmission.

158. Addressing Outlook MSG files specifically, as that is the format in which most of the emails have been produced, the metadata tracked by MS Outlook includes the following fields. Many of the fields include a "**PR_**" property tag. The suffix **_W** is used to indicate that they are stored in Unicode format. The content of these fields is typically imported from other sources

such as the Transmission header, or populated by MS Outlook itself for emails created within MS Outlook. The actual metadata recorded varies widely depending on the nature of the MSG itself and it is not possible or useful to try to set out an exhaustive list, but some of the relevant metadata includes the following PR_ fields (with a short description of their use):

PR_CREATION_TIME The creation time of the file or message
PR_DISPLAY_[CC/BCC/NAME/TO]_W Fields relating to the "display name" – the contact names associated with an email addresses – with possible tags for each CC, BCC, Sender and Recipients
PR_HASATTACH Whether or not an email has an attachment
PR_ATTACH_EXTENSION_W The extension of any attachment
PR_ATTACH_FILENAME_W The filename of any attachment
PR_ATTACH_SIZE The size of any attachment
PR_BODY_W The main body text of the email – normally the main user data content
PR_CLIENT_SUBMIT_TIME The time when the email was submitted by the client – according to the local clock setting on the machine used to send the email
PR_IMPORTANCE A setting relating to the importance or priority of the email

PR_INTERNET_MESSAGE_ID_W The unique identifier of the message for tracking and referencing
PR_INTERNET_ORGANIZATION_W Any organisation or company name associated with the sender, if an organisation is associated with the email address of the sender
PR_LAST_MODIFICATION_TIME The date of last modification of the message
PR_MESSAGE_CLASS_W The class or type of the message (e.g. email, appointment request, delivery receipt)
PR_MESSAGE_DELIVERY_TIME The time when the message was delivered to the recipient
PR_SENDER_EMAIL_ADDRESS_W The email address of the sender
PR_SENDER_NAME_W The name of the sender
PR_SENT_REPRESENTING_ EMAIL_ADDRESS_W If an email was sent by one person on behalf of another, this indicates the email address of the person on whose behalf the email is sent
PR_TRANSPORT_MESSAGE_ HEADERS_W The set of email headers received during transport of the email

159. This is not an exhaustive list but does provide a useful overview of the types of metadata expected to be present within emails.

EXIF metadata

160. Certain types of files, such as image and video files, contain metadata in "EXIF" format which stands for EXchangeable Image File format. This is internal to the file and can contain basic

details (such as the size and encoding of an image) as well as quite rich details including the geolocation at which the image was taken (if a photograph) and information about the device on which it was taken, copyright and tags relating to image editing. It is typical for photographs taken on smartphones to contain embedded EXIF metadata.

Fonts and Typefaces

161. When displaying or printing text or numbers, the computer system will use fonts and typefaces. Fonts and typefaces are used to present written content in a particular way, by interpreting character information (such as encoded letters) and outputting visual copies of those letters on screen according to the designs specified in that font.

162. While the terms are often used interchangeably, it is helpful to consider a distinction between a **typeface** and a **font**. Simply put:
    a. A typeface is the design of the letters and characters themselves;
    b. A font is the actual file that contains the information necessary to draw or render the document's content in the specified typeface.

163. Different operating systems come with different fonts installed, and users can install additional fonts by downloading the font files from the internet and installing them. Fonts may also be uninstalled. It is therefore common for some file formats, such as a PDF document to embed the content of a font file into the structure of the document to ensure that the text presents as intended, which I have explained above.

164. Fonts tend to contain metadata and the releases of popular fonts are also often well-documented online. This can be a useful forensic resource for discerning the true authorship time period for a document, because (a) fonts can contain metadata indications of their date and (b) the dates of their content and publication can also be used to narrow the time window of creation of a document.

165. There are five specific fonts that I have analysed in particular in my reports and it is useful to summarise them here:

    c. **Times New Roman**: This is a very standard font that will be familiar. It has existed for a long time as a computer font and has come pre-installed with Windows systems since the early 1990s. However, the font has been revised and extended since that time, such as to extend it to be suitable for additional

non-Latin alphabets, and different versions of it are available. These can often be identified by the date of the copyright statement embedded within the font.

d. **OpenSymbol**: This is a font which appears in the Bitcoin White Paper and is used to encode symbols within equations.

e. **System UI:** This font appears in one document relating to the Bitcoin White Paper and I explain it in detail when it occurs in my analysis, in conjunction with OpenSymbol.

f. **Calibri Light**: I observed that various documents in the disclosure dataset contained references to the font Calibri Light. While to my knowledge the font Calibri itself is a Microsoft default font that has existed since 2007, the variation Calibri Light was not released by Microsoft until much later, in connection with versions of MS Office in 2012-2013[9]. I raised this with Bird & Bird, who I understand asked the designer of that font to indicate when it was designed. I have been provided with a copy of a letter from the designer, Lucas de Groot, which confirmed the view I had reached independently.

g. **Nirmala UI:** I observed that some documents in the disclosure dataset contain references to a wide range of fonts whose names I was not familiar with, which appeared to relate to various different non-Latin scripts. Choosing one of these at random, I searched online to see when Nirmala UI was released, and noticed that it appeared to post-date the date of certain documents that referred to it. Bird & Bird then contacted the designer. I was later provided with a copy of the witness statement of John Hudson which confirmed the results of my own research.

Types of documents that have not featured significantly in my review

166. I have not been able to assess all documents within the disclosure dataset. Documents included in the disclosure dataset which have not featured significantly in my analysis include:

[9] Microsoft's own documentation about the release of Calibri Light in a software update dated September 2012 is available at the following URL - https://support.microsoft.com/en-us/topic/an-update-is-available-to-add-the-calibri-light-and-calibri-light-italic-fonts-to-windows-7-and-windows-server-2008-r2-717a0bbe-0610-bd3c-3cc1-bad9b7809e55

a. **Computer source code documents such as those with the suffix .CPP and .H., and other plain text files.** Source code is predominantly stored in plain text files, which is a very lightweight format that does not carry much or any internal metadata. While I have included a few plain text documents in the course of my analysis where I have encountered them, unless there is a comparative analysis to be conducted between documents it is typically difficult to conduct a meaningful forensic examination of these. While a comparative review between code files would be possible, computer code is not written in the same way as normal language and a meaningful review of the code would require a level of programming / development expertise that I do not have.

b. **Executable files**. Although Executable files did feature in the dataset, they did not originally form part of my analysis – because in my experience they do not typically carry document data susceptible to forensic review, but are files designed for execution on a computer system. Towards the end of my analysis, Bird & Bird asked me to look at executables in the disclosure dataset from a perspective of examining them as any other disclosure document for any information contained within them, and this also led me to analyse certain plain text 'log' files that appeared to be related. I have reported on that analysis in **Appendix PM11** relating to the log files and **Appendix PM12** relating to an executable file.

{H/64} {H/68}

c. **Scans of hard copy documents.** Many disclosure documents have been provided as scans or photographs of hardcopy documents and also some handwritten notes. There is, in most cases, little I can discern by digital analysis of such scanned documents beyond the properties and metadata that relate to their photographing or scanning. In several cases, digital documents have been printed and then re-scanned, and only the scanned copies disclosed, effectively leading to the complete loss of underlying metadata relating to the original copy. In a few cases, I was able to draw conclusions based on a face-value review, where the user data of the file happened to provide indications that were pertinent to forensic review (such as timestamps, irregularities with fonts, and the ability to cross refer to comparative documents sourced from outside the disclosure dataset). In one case, covered in **Appendix PM5**, the disclosed document itself did not provide a useful basis for forensic analysis, but I was able to conduct a comparative analysis against a native digital document provided to me by Bird & Bird.

{H/31}

d. **Disclosure documents with little metadata.** I have observed in many cases that there are several documents that appear to exist in multiple different forms, with different

forms bearing (or likely to bear) different metadata. I have observed several occasions where the list of Reliance Documents refers to documents that are very light on metadata; meanwhile other documents which would have carried more useful metadata for analysis appear to have existed, but have not been disclosed. In other cases, other related documents have been included within the disclosure dataset but were not simple to connect. To provide an example of why this matters, I refer to Reliance Documents ID_0004077, ID_0004078, and ID_004079, which consist of screenshots that have been converted from image files into PDFs dated within the period 2009-2010:

- These Reliance Documents contained no useful metadata for analysis.
- It was only by conducting detailed investigation online for fragments of the content visible on the face of those documents and browsing technical documentation that I was able to establish the likely application from which they were taken.
- Having done that, I noticed an apparently relevant file that was included in the disclosure dataset, but that file was not disclosed in its own right nor was it assigned with its own ID number.
- Rather, there was an email whose body content was blank, which contained a zip attachment. That zip attachment within the email contained various files inside it which included another zip file.
- Within that second zip file were various files including one with a file extension that appeared to relate to the application software I had established as a possible source of the Reliance Document screenshots.
- It was only by investigating that file, which was three layers removed, that I was able to establish the origin of the Reliance Documents themselves, as it was a database file containing related entries to those shown in the Reliance Documents.
- At length, I was able to form a conclusion that the content shown in the Reliance Documents did not date from 2009-2010 but was in my opinion much more recent, apparently dating from 6-7 March 2020.
- The database was the actual source of the information in those three Reliance Documents.
- As can be seen, the analysis was very convoluted because of the way the Reliance Documents had been produced as screenshots without context, and the original content file was hard to find. Had the database file been given its own ID number

and relied upon directly, the analysis would have been significantly swifter and more direct.

- Had I not been able to establish the link between those files, the only contextual information I would have had about the Reliance Documents would have been the date shown on their face and I would not have been able to meaningfully investigate the authenticity of the documents.

{H/47} The full analysis of these documents is set out in **Appendix PM7**. The purpose of giving this example, which is at the more extreme end (it is not given as representative), is to show how the analysis can depends on being able to put the disclosure files in their context, and this is difficult to do and cannot be done reliably simply by searching the database electronically. There are very likely to be other circumstances where relating documents together was not possible, an example being given in the Email Forwards paragraph below. In those cases, I have not been able to meaningfully examine the documents of this kind.

e. **Email Forwards.** Forwarding an email typically involves authoring a new message, with the content of the original email shown below it in text form. The original content is, at that point, freely editable and manipulable. Therefore, an email forward provides very little meaningful information for a forensic analysis. If the original email that has been forwarded was disclosed, it may be possible to relate them together and to form a view as to authenticity by examining the header content of the forward, but if the relevant headers are not recorded or have been altered then it is simply not possible.

f. **Other Conversions.** Several documents have been provided which have been converted digitally between formats, in ways that have led to the loss of some or much relevant metadata – such as digital conversions of PDFs into image files, and digital conversions of image files into PDFs. In most cases, the underlying files from which conversion has taken place have not been provided in the disclosure dataset. In one case addressed in detail in my report in **Appendix PM6**, I observed a PDF which was apparently created by combining two different PDFs together: neither of the underlying PDFs appeared in the disclosure dataset.

{H/40}

167. Finally, I emphasise that in my report I have been instructed to report on documents on which I am able to provide useful observations, and not to report on each and every document in the dataset.

TOOLS USED IN MY ANALYSIS

168. The purpose of this section is to briefly explain some of the tools I have used in the course of my analysis and what they are for. Although I try to be as brief as possible in this section, the dataset I have been asked to examine is very wide ranging and the evidence of manipulation I have seen spans a wide range of different techniques and documents, so there are correspondingly more tools to describe.

Hardware and Virtual machines

169. It is my practice to conduct my analysis on a dedicated computer that is not connected to the internet. The purpose of isolating the computer from the internet includes security reasons, as well as being certain to avoid contamination of any disclosure documents. It is also my practice to conduct my analysis using new, clean installations of MS Windows.

170. The use of Virtual machines is another common tool for forensic analysis. Virtual machines are software emulations of a physical computer, which simulate a physical computer, and which can be configured according to the needs of my analysis at the time. This can include the use of different operating systems to be run simultaneously, each on its own virtual machine, but without interacting with each other and without interacting with the data on my main machine itself. Therefore where my analysis requires inspecting the behaviour of applications in their original context (such as on a legacy version of Windows like Windows XP installation), it can be done without requiring a dedicated physical laptop.

Dedicated forensic software

171. I have used a variety of specific forensic applications and techniques to analyse the content of the disclosure dataset:

a. **FTK Imager, version 4.7.1.2**[10]. FTK Imager is a free utility that can be used to create full disk and logical forensic images. While it is a lightweight utility, it includes some useful capabilities such as quickly viewing the content of a file or document in plain text or hexadecimal format.

b. **Encase Forensic, version 6.19.7 and version 22.1**[11]. Encase Forensic is an extensive application that handles a wide range of tasks useful to a forensic examiner, including

[10] https://www.exterro.com/ftk-imager
[11] https://www.opentext.com/products/encase-forensic

a number of utilities and functions to dissect and analyse electronic documents. This can be used for example to view files or parts of files, conduct advanced searches on their internal content, and extract relevant data useful to analysis.

More standard user applications

172. I have also used a variety of standard user applications relevant to the source of some of the information. Some of them have overlaps in functionality as they complete certain tasks differently in ways that are more useful in some cases than others, and are therefore each better suited to certain tasks. Where possible I have used either free open-source, or common readily-available software to permit ease of access to the analysis and findings.

173. The software used includes, but is not limited to:

a. **Hex Editor Neo[12].** Hex Editor Neo is an application that is useful for interrogating the content of files allowing the file content to be viewed at a binary level. It includes a number of features such as a structure viewer for recognised file structures and the ability to compare two digital files side by side. It also includes a number of encode/ decode functions and a base convertor.

b. **Textpad[13] version 6.5 and version 8.2.** Textpad is a simpler text editor that includes a number of useful search features, and abilities.

c. **Notepad++**[14] is another text/hex editor that offers a range of utilities for searching and altering text files and also provides highlighting for various forms of code, which can allow for easier reading of that data.

d. **MS Office**[15] (various versions). Microsoft Office has been used to create a number of the documents from the disclosure dataset as discussed in this Report. Where possible, I have conducted testing using the appropriate versions contemporary to the documents in question. I have also used it in the preparation of this Reportand other aspects of case management.

[12] https://www.hhdsoftware.com/free-hex-editor
[13] https://www.textpad.com/
[14] https://notepad-plus-plus.org/
[15] https://www.office.com/

e. **OpenOffice.org suite[16] (various versions)**. OpenOffice has been used to create a number of the documents discussed in this Report. Where possible I have conducted testing using the appropriate versions thereof.

f. **Microsoft Windows[17]** (various versions). Different versions of the Microsoft Windows Operating system have featured in this analysis. I have conducted analysis regarding several versions that would have been contemporaneous at the time.

g. **QPDF**[18], a command-line tool that performs content-preserving transformations on PDF files. It can be used to expand (inflate) compressed portions of a PDF file to make them easier to analyse.

h. **PDF Stream Dumper[19]**, a useful application for exploring the content and structure of streams within PDF files.

i. **Winking PDF Analyser[20]**, another PDF analysis utility that can be used to explore and analyse the structure and working of a PDF document.

j. **Paint.net[21]**, a graphics editing suite useful for creating screenshots and diagrams.

k. **OutlookSpy**, a tool for analysing the metadata and content of email message files.

Hex editor functionality

174. Many of the functions above will be familiar or will be clear from the way in which they are presented but one which may not be familiar and is useful to explain is the functionality of a hex editor.

175. A hex editor, or "Hexadecimal" editor, operates in a similar manner to a text editor, but it is more advanced and includes additional functionality. A hex editor typically allows a file to be viewed in two ways at the same time:

[16] https://www.openoffice.org/
[17] https://www.microsoft.com/en-gb/windows
[18] https://qpdf.sourceforge.io/
[19] http://sandsprite.com/tools.php
[20] https://www.winking.be/en/products/pdfanalyzer
[21] https://www.getpaint.net/

a. On the left-hand side, it shows the raw content of the individual bytes of the file. A byte (which consists of 8 bits) can range in value from 00000000 to 11111111 rendered in binary (base-2) format. When rendered in hexadecimal (base 16) format, these same numbers range from 00 to FF, and can be shown in two-character pairs. A hex editor is referred to as such because it renders the bytes of the file in hexadecimal format in this way, allowing individual byte-level editing of a file.

b. Byte-level data is not usually human readable, and so on the right-hand side a hex editor will typically display the same data interpreted as text. This allows text-encoded parts of a file to be rendered as human-readable text where appropriate. Other parts of the file which are not human readable will display as null characters (such as a series of dots ....) or as nonsense.

c. The two panels therefore show the same data in two ways, and scroll alongside each other. Making a change to one side will affect both sides, since they are both showing the same file in two views.

d. A hex editor can be used to alter the raw hexadecimal or byte level content of any filetype. They can display or represent effectively any character, and can save any character too.

Hashes and checksums: MD5 and SHA256

176. A useful tool in forensic analysis is the use of common hashing functions. These are cryptographic hash functions which take as their input the data encoded within a file, and which output a string of characters that represents the data content. They can be thought of as showing the 'fingerprint' of a file, such that these algorithms can be quickly used to determine whether or not two files are strictly identical (i.e. whether there exist any differences at all in the internal byte-level data). If the hashes of two files match, then the chance that any changes have been made is negligible.

177. The hash is applied to the internal content of the file, and does not include the file name or file (OS) timestamps which are stored separately to the files. This means that two files with the same content will match even if they have different filenames.

178. Hashing functions are also sometimes referred to as 'checksums' because they provide a checksum functionality to check whether a document has been corrupted or modified. For

example if a file is hosted online for downloading, the web page may also include the hash of that file as a way of checking the download integrity. After downloading, a user can conduct a quick hash operation to ensure that the file has downloaded properly, and to ensure that the file they have received is the file that was intended.

179. Two widespread hashing functions for this purpose are known as **SHA256** (Secure Hash Algorithm – 256) and **MD5** (Message Digest 5). These perform essentially the same function, though SHA256 provides a 256-bit output and MD5 provides a 128-bit output.

180. To illustrate, the hashes of the Bitcoin White Paper (ID_000865) are as follows:

MD5: `d56d71ecadf2137be09d8b1d35c6c042`

SHA256: `b1674191a88ec5cdd733e4240a81803105dc412d6c6708d53ab94fc248f4f553`

WHOIS Internet Domain registration records

181. As I have indicated above, a forensic analysis includes assessing the whole context of the document and its history, and in connection with this I occasionally make reference to the domain registration of internet domains to establish when they were registered, as comparators for the historical context of a document which makes reference to them.

182. The protocol known as WHOIS is a standardised way to query internet domain registration databases to return results relating to the registrar, IP address, registration dates and recorded owner and contact details of a website. The protocol is specified in internet standard RFC 3912 which is available at https://datatracker.ietf.org/doc/html/rfc3912.

183. This provides a good indication as to when the current owner of a domain name registered that ownership. It is not infallible because it usually relates only to the currently-recorded registration of a URL. It is possible for a URL to be registered earlier than indicated in the current records, as the records will be replaced when they are re-registered, even if the re-registration is by the same person. Re-registration is relatively rare (domains that are in use tend to be renewed rather than allowed to lapse).

184. Throughout my analysis I have used the *Domaintools.com* WHOIS lookup service for this function, which provides a quick and convenient lookup service using the same standardised protocol.

Internet archive and Wayback Machine

185. As mentioned above, I have also used the archives of the Wayback Machine at *archive.org* as a tool to provide historical context relating to the content and availability of historical internet information.

186. I understand from Bird & Bird that the Wayback Machine is well known in this Court and that it has also been used in other evidence and that I can take it to be understood without needing to explain it further here.

ELECTRONIC DOCUMENT CREATION / MANIPULATION

187. The purpose of this section is to give an understanding of how electronic documents can be manipulated and whether and how such manipulation can be detected. It is not directed at the documents in the disclosure dataset specifically, and my opinions in relation to those documents are set out in the relevant Appendices. It will however assist to understand that analysis, and I have included specific explanations of techniques that do feature in my analysis at several points.

Overview and the possibility of a perfect forgery

188. Through the span of my career conducting computer forensic examinations I have identified numerous ways in which electronic documents have been altered or manipulated to either misrepresent what was originally communicated/recorded or to generate entirely new documents that purport to be contemporaneous to an earlier time period. I am also regularly instructed on matters in relation to the misappropriation of intellectual property or confidential information. My investigations have helped courts to identify documents that were manipulated or altered, as well to establish the authenticity of others.

189. From my experience, I have however reached an understanding that it is possible to create a perfect forgery of an electronic document. This is to say that, with sufficient skill and resources, it is possible to create an electronic document forgery that is indiscernible from a genuine document, especially when the document is inspected in isolation without access to the underlying computer equipment or forensic images from which the document is sourced. It is also possible for a forgery to be imperfect, but still to be overlooked (again, especially when it is provided without the necessary context).

190. Correspondingly, it can also be very difficult to establish a document to be genuine with 100% certainty, especially without access to the context in which it was created. It is however possible to scrutinise a document carefully with a view to exposing any irregularities and to do so in combination with scrutiny of other similar comparator documents, and if appropriate, so to form an opinion that it does not exhibit any characteristics of manipulation.

Controlling, Editing and removing metadata

Donor, precursor, and intermediate documents

191. In theory, a perfect forgery of any document could be created by authoring the document at a byte level without the use of a native application. This would be very difficult however and would require a very deep understanding of both the file format in question and the encoding of data within it, and in the case of complex documents is not practically possible.

192. Therefore when creating a document artificially, rather than starting from scratch, it is often easier to take a pre-existing document and alter the content of it, so that it presents as something different. In this way a pre-existing document can be used as a template. It can also be easier to author the intended content separately, before conducting a process of incorporating that content into a document. During the process of creating an inauthentic document, (depending on the method used) it can also be necessary to create copies of documents at different stages.

193. When referring to such documents I refer to:

a. "**Donor**" documents and donor content, to mean any document whose content has been carried across into the document being examined;

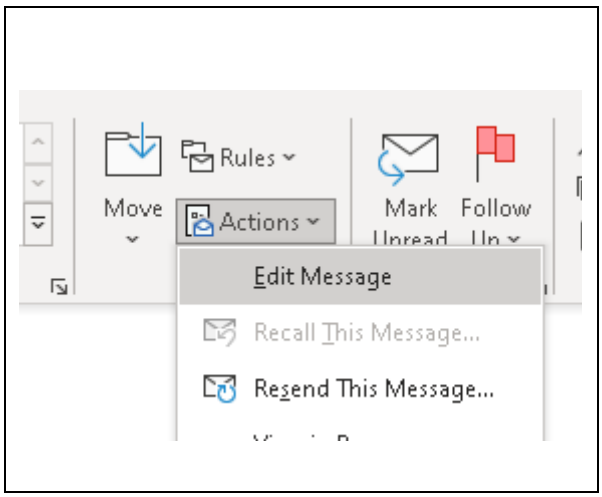
b. "**Precursor**" documents, usually to mean the pre-existing files that have been edited. When used as a template, some content from the precursor document (even if just its overall structure) will usually also be donor content; and

c. "**Intermediate**" documents, to mean documents created between the stage of the precursor and the final document.

194. To give an example of this, I use the case of a pre-existing email document. Email documents contain relatively complicated header and other metadata information that is much easier to

take from a template. Taking an email as a template, changing the sender/recipient information and the message body can be easier than trying to generate an entire exchange from scratch.

195. Many applications natively include a function to edit or alter the content of pre-existing content. MS Outlook includes such a function whereby a recipient can "edit" the content of a sent or received message. This process can be accessed through a user menu as shown in the screenshot below:



196. The capabilities of this tool are limited to certain fields of the document, however there are other applications and methods whereby the entire content can be manipulated. I address the editing of emails in this way as they come up in the course of my Appendices.

197. An even easier approach to manipulating the content of emails is not to produce native documents, but to create content when forwarding or replying to an email. At that point, the text of the email below is freely editable simply by typing. Therefore, without the complete original email message being replied to, in its native format, such content cannot be relied upon as original in a forensic analysis.

Exercising control over metadata which is generated

198. First, it is helpful to understand the factors that cause metadata to be generated when a file is created, because exercising careful control over those circumstances can allow the author to manipulate the metadata content, either directly or indirectly.

199. While some metadata is created automatically by computing systems alone (and is not intended to be editable), others are created from purely user-editable fields. Somewhat in between these

two categories is metadata which are created automatically by the software in question, but where the inputs to that software are within the control of the user:

a. An example of **system-created metadata** is a unique ID of a PDF file, the server routing information in the header of an email. These are typically created in the routine operation of the document's creation without user input.

b. An example of **user-editable metadata** is the filename of the document (external metadata) and the "title", "keywords" properties of an MS Word document. These can simply be typed in by an author.

c. Examples of the kind **in between**, system created metadata into which the user has an input, include for example:

    i. Internal and external metadata timestamps – that are typically set by the software consulting the clock setting on the computer, which is a user-controlled setting as I explain further below.

    ii. Author information – typically set by consulting the relevant username or account details associated with the software. However, user information is typically set by the user, so is itself an editable input.

    iii. Information about the software used to create a document - where such information exists, it is typically automatically generated at the time of creation of the document. However, a user has control over which tools are used to generate a document. It is typical that older versions of those tools will be retained on older systems, or available for download on newer systems, and can be selected depending on the user's needs.

200. Therefore, at the time of creation of a document, much metadata can be subject to user control. While this is not typically a concern for most users creating documents, the careful control of the circumstances of document generation is a tool used in forgery.

Manual manipulation of metadata

201. After a document is created, metadata can still be subject to change. In addition, it cannot be guaranteed that any metadata visible in a document is the same at the time of analysis as it was when the document was first created.

202. Metadata can be altered by manual manipulation of a document, such as by the use of a hex editor (to make byte-level changes to embedded metadata) or by the use of a plain text editor to edit plain text content in a way that may be difficult to detect.

203. It is rare that the internal content of a file is not editable in at least some way. I have explained this in more detail in my Report as it arises.

204. This is all to say that the recorded metadata can be manipulated by various methods to create an electronic document that, when looked at in isolation, appears to be authentic contemporaneous.

Removing metadata from documents

205. Metadata can also be removed from documents in a number of ways:

a. There are many applications available, including online and for download from the internet, that can be used to strip or alter the recorded or structural metadata of a document (AttributeMagic[22] being one such example).

b. Stripping metadata is often done by professional firms such as lawyers sending digital documents by email, whereby the metadata is deliberately stripped from the file at the point of emailing it to avoid disclosing unnecessary or privileged information. I use this example to explain that stripping and alteration of metadata deliberately can be done for legitimate reasons.

c. Metadata can also be removed incidentally, such as in a process of printing a document and re-scanning it which does not retain data of the original source.

d. Metadata can also be lost or altered as a result of poor handling (such as creating copies in different file systems).

[22] https://www.elwinsoft.com/attributemagic-free.html

e. It may also be lost as the result of a conversion between file formats or by printing and scanning the document.

The use of clock manipulation techniques

206. Almost invariably, modern electronic devices such as laptops, computers, mobile phones, and handheld tablet devices include an internal clock that tracks the date and time. Understanding the way this is done is important as it is usually the source of metadata information provided.

Meaning of a computer's 'clock'

207. The "Clock time" on the computer can in refer to multiple different types of timekeeping:

a. It can refer to the timekeeping associated with the hardware of a device (such as in the processor or motherboard). Hardware timekeeping information is not usually presented to the user.

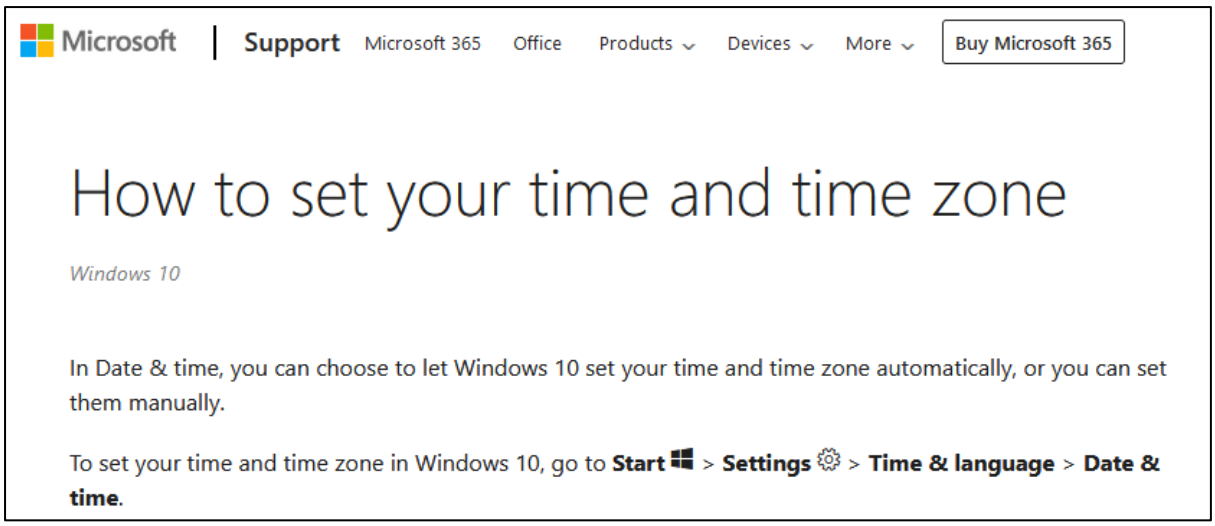
b. The 'clock' or 'clock setting' of a computer can refer to the software timekeeping that is presented at a user-level via the operating system of a computer. For example, the clock readout on a Windows 10 system typically looks like the following screenshot[23], which is likely to be familiar to any users of Windows 10 or similar operating systems:

[23] This screenshot is taken from a Google Image search result for an image at https://softkeys.uk/blogs/blog/how-to-add-uk-clock-on-desktop-windows-10, and is purely illustrative.

c. When I refer to the computer's clock, I am not referring to any internal hardware features, but typically to the operating system clock such as the example above from Windows 10.

Automatic and manual setting of the clock and effects on metadata

208. Often devices are configured to set the time automatically, whereby they routinely synchronise their date and time from a central time reference server, the intention being to maintain the local clock accuracy and account for time zone variations. However, an automatic clock setting is not the only option, and on most operating systems it is freely possible to manually adjust the date, time and time zone settings. Continuing with the example of Windows 10, Microsoft publishes guidance on how to do this at https://support.microsoft.com/en-us/windows/how-to-set-your-time-and-time-zone-dfaa7122-479f-5b98-2a7b-fa0b6e01b261, a screenshot from which is below:



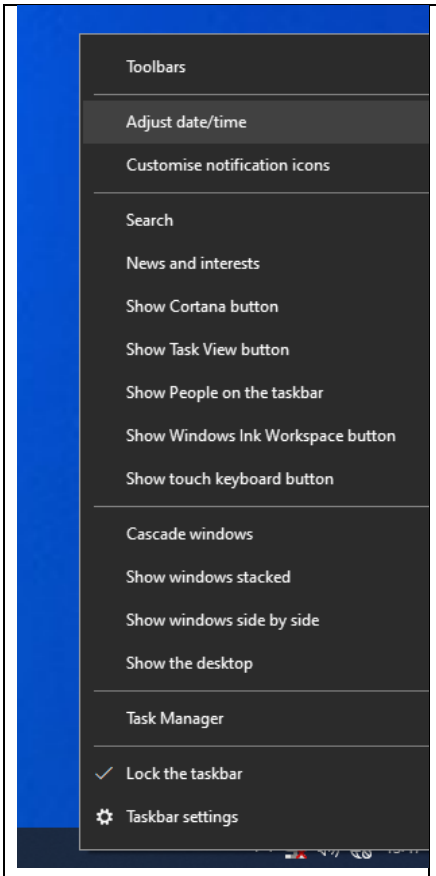
209. This clock setting is used by the operating system, and by software running on the operating system, as the local clock against which the time is checked during various system events. This includes checking the time for committing to user documents, metadata information, or entries being added to system audit or event logs.

210. Should a clock be set incorrectly, this would result in incorrect dates and times being committed to user documents or metadata information recorded, which would not match the actual date and time that the relevant event occurred.

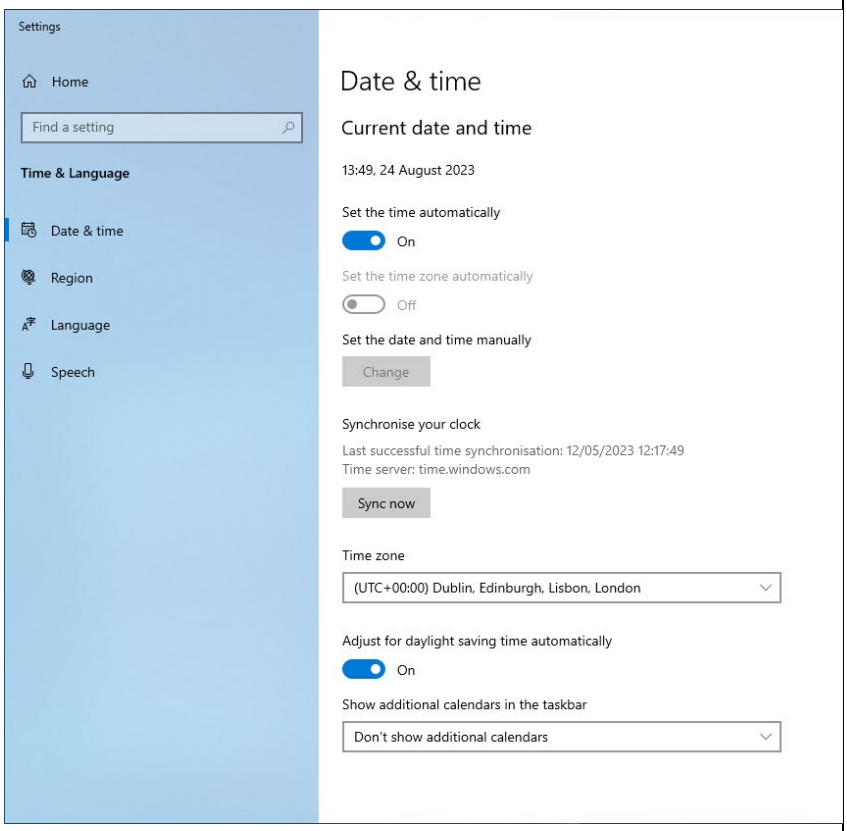
The technique of clock manipulation

211. It is therefore possible to backdate an electronic document by manipulating the clock setting on the computer, thereby creating the document in circumstances where the user-selected time will be injected into the metadata of a document. This allows the metadata to be controlled by the user and the resulting document to be manipulated so as to appear to originate earlier than its true creation.

212. I demonstrate this as follows, continuing with the example of Windows 10 and using the date "2008", one can right click on the time in the bottom right corner of the screen and selection "adjust time". This brings up a settings window for the clock, which includes options such as automatic synchronisation of the clock, or the ability to manually set the clock to a time of the user's choosing. The steps are shown in screenshots as follows:

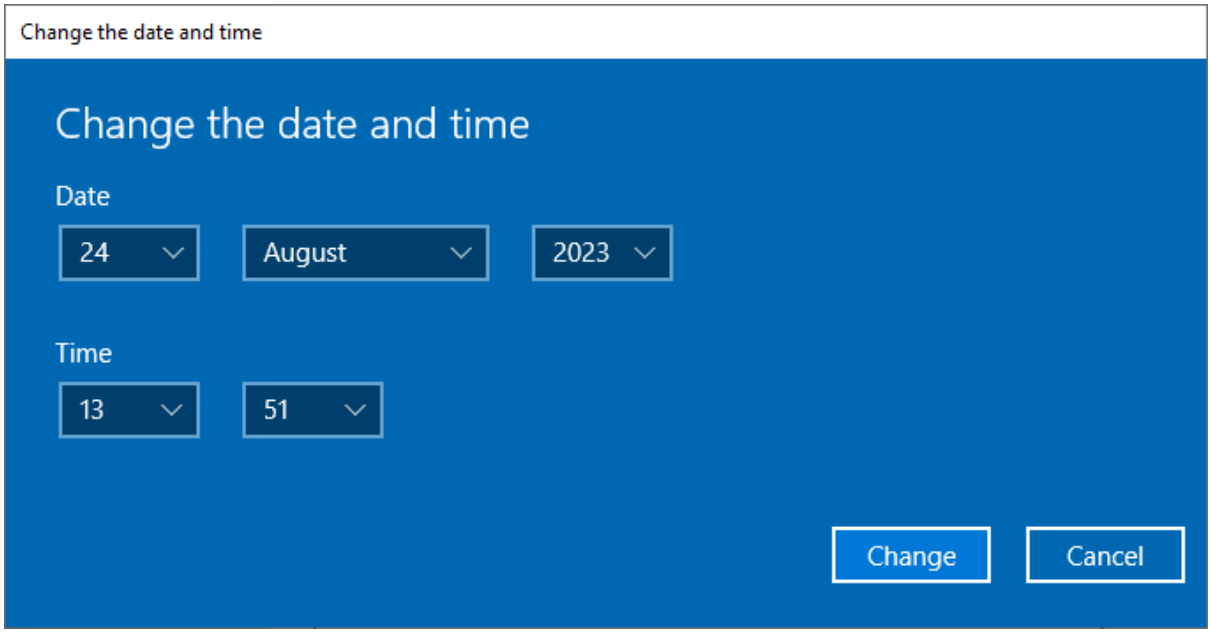


Right-click menu in bottom right hand corner of Windows taskbar

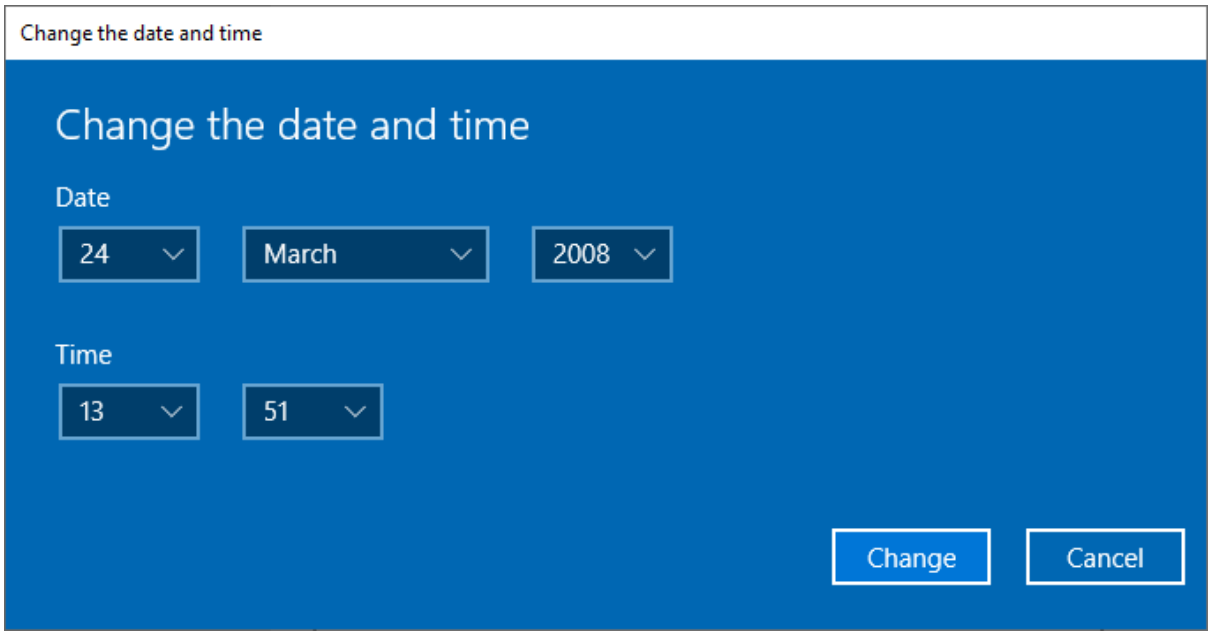


Date and time settings dialog

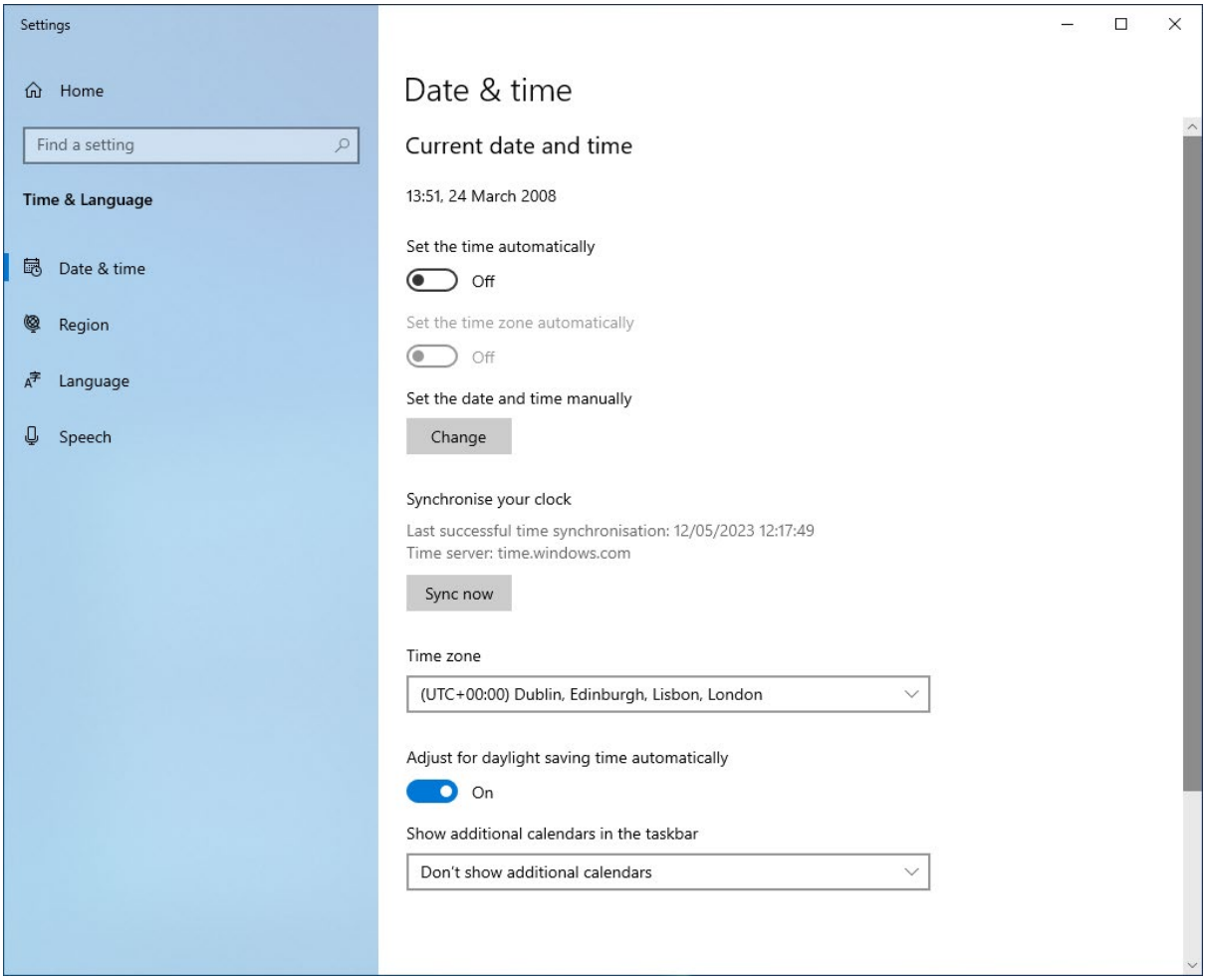
By disabling the automatic settings and clicking on Change, the user is presented with the following window:



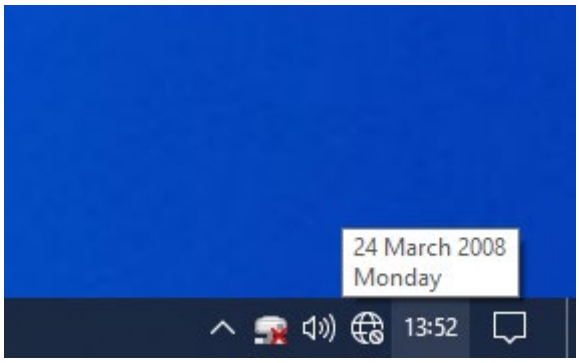
213. In this example, I have set the clock back to March 2008 as per the below:



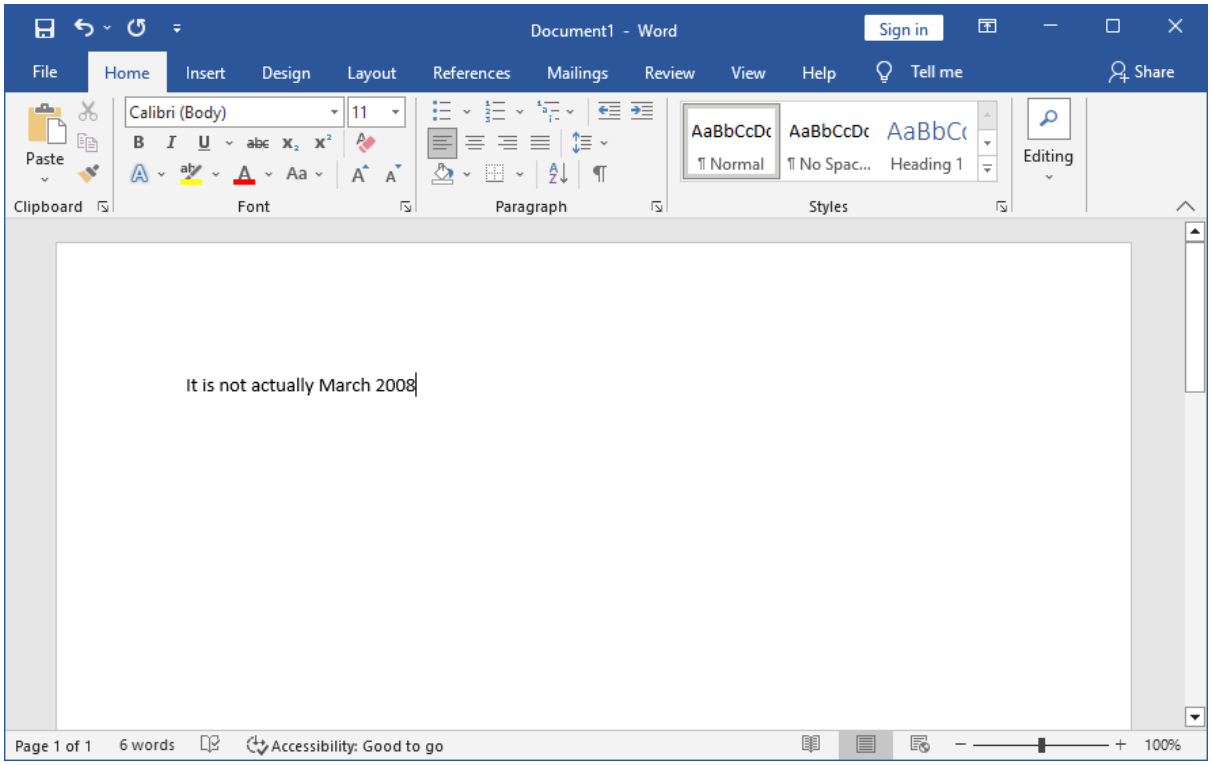
The change has been committed to the date & time settings window:



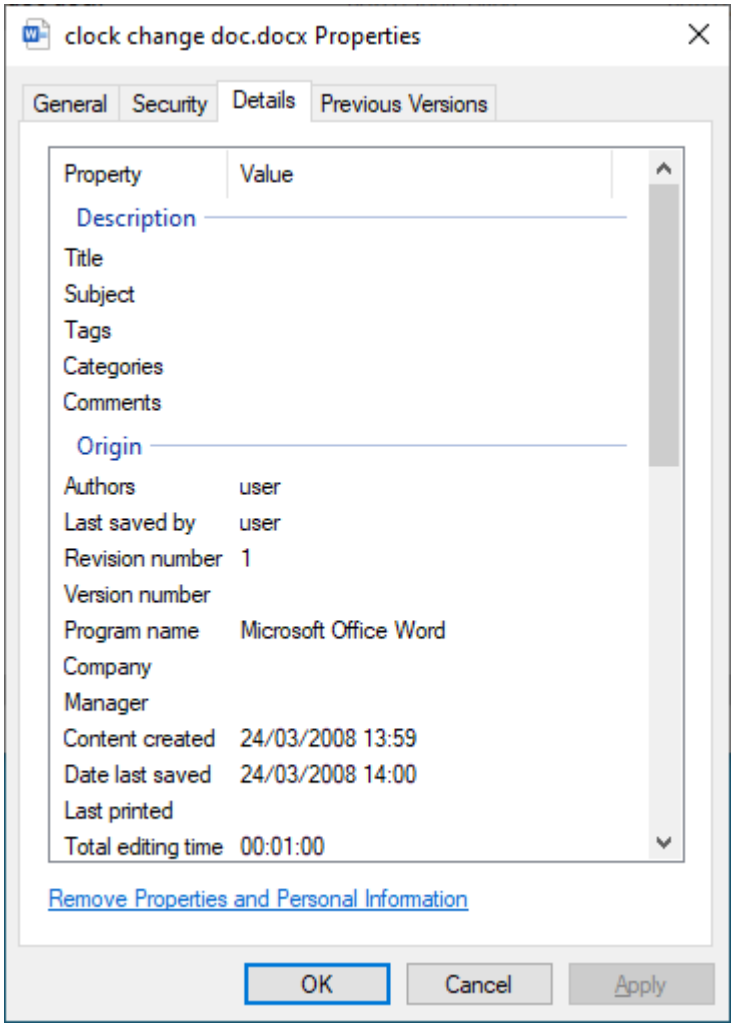
214. After closing this window and hovering the mouse pointer over the time in the taskbar it displays the full date, confirming that the change has taken effect:



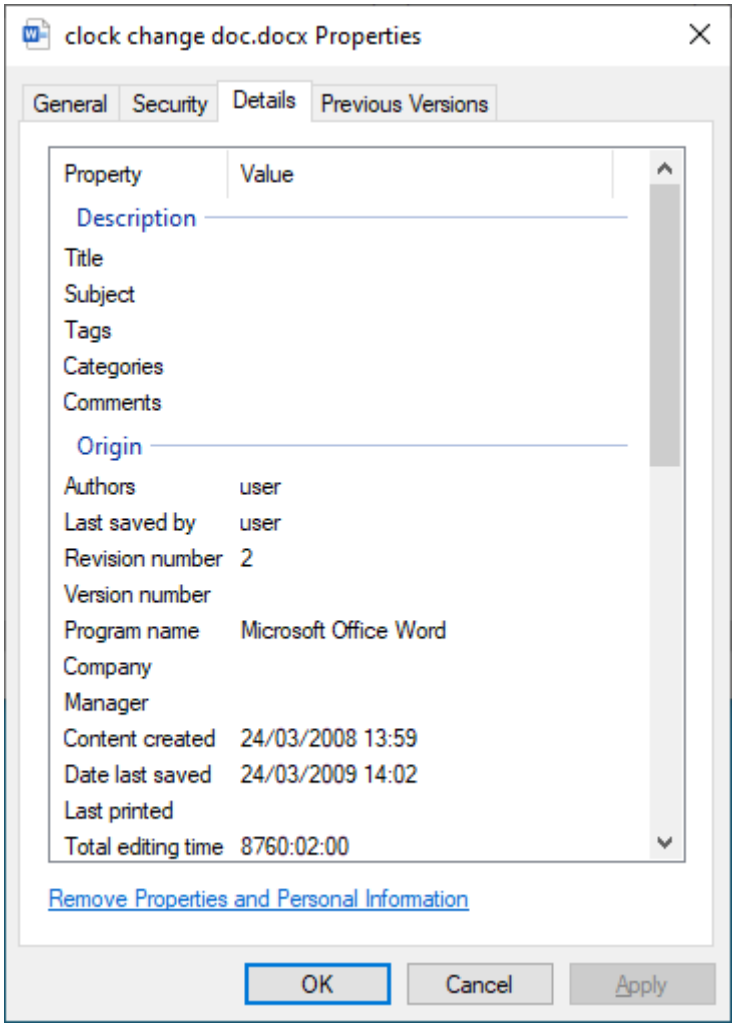
215. That setting will then be used by applications and by the operating system to assign metadata timestamps. To illustrate this, with the time set to March 2008, I launched MS Word and started to author a new document:



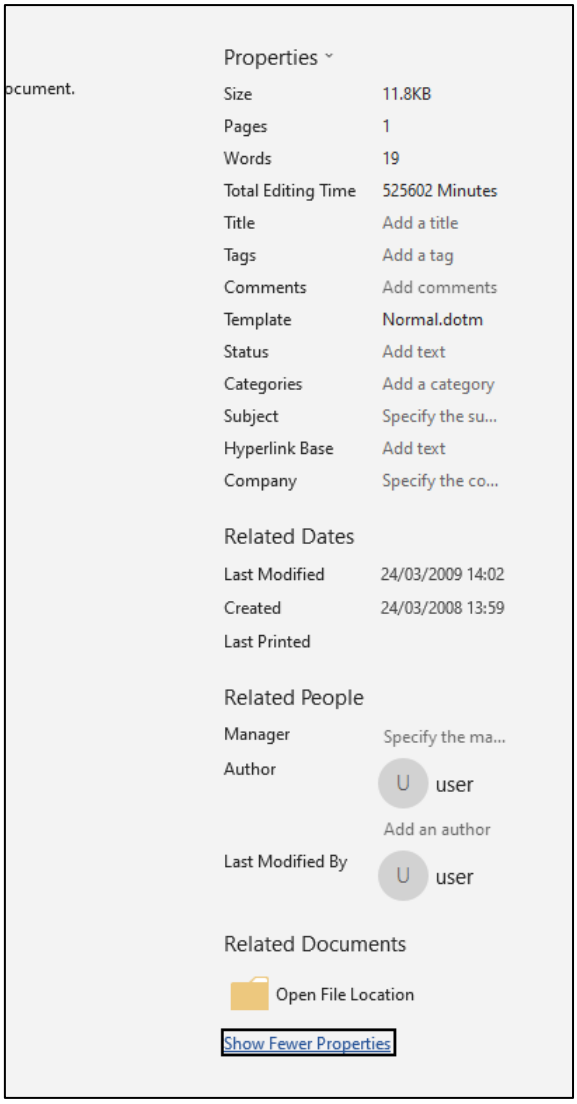
216. I then saved the document with the file name "clock change doc.docx" resulting in the following Created and Last Saved properties which are 1 minute apart, with a revision number of "1" indicating that there had been one save, and a total editing time of 1 minute (as I had spent less than a minute editing the content, and this time is given with minute-level accuracy).



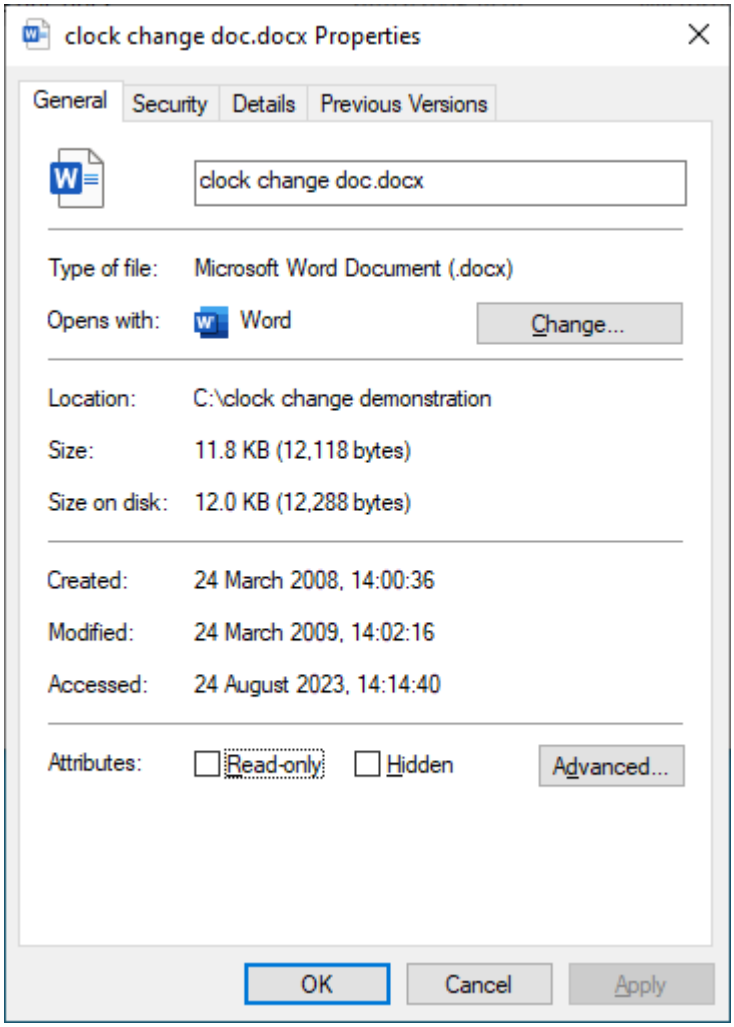
217. Repeating the settings steps shown above, I then proceeded to set the clock forward by exactly one year, to a date in 2009 and performed a save operation on the same document again. This led to the Date last saved, Revision Number, and total editing time fields being updated as shown below:



218. The same metadata information can be viewed natively within MS Word itself (the Total editing time being given in minutes, whereas it is expressed in hours and minutes in the screenshot above):



219. The screenshots above show the internal metadata, but the external file metadata is also created in the same way. Having taken those screenshots, I then set my computer clock back to the present day and accessed the document, before taking the following screenshot of the operating system file properties dialog for this file showing the equivalent operating system metadata:



220. As can be observed in the various screenshots above:

a. **Edit time.** The Total editing time is 8760 hours and 2 minutes, of which:

i. The 8760 hours were accrued artificially simply by advancing the computer clock. In this case I selected one year: 24 hours x 365 days = 8760

ii. The excess 2 minutes related to the time I actually took to perform the process, consisting of about 1 minute to create the document (and take the screenshots shown above), and about 1 minute to advance the clock and save it again.

iii. In this way, I was able in the space of just 2 minutes to create a document that was artificially dated and appeared to have been edited for many hours,

and contained no internal record of the clock manipulation steps that I undertook.

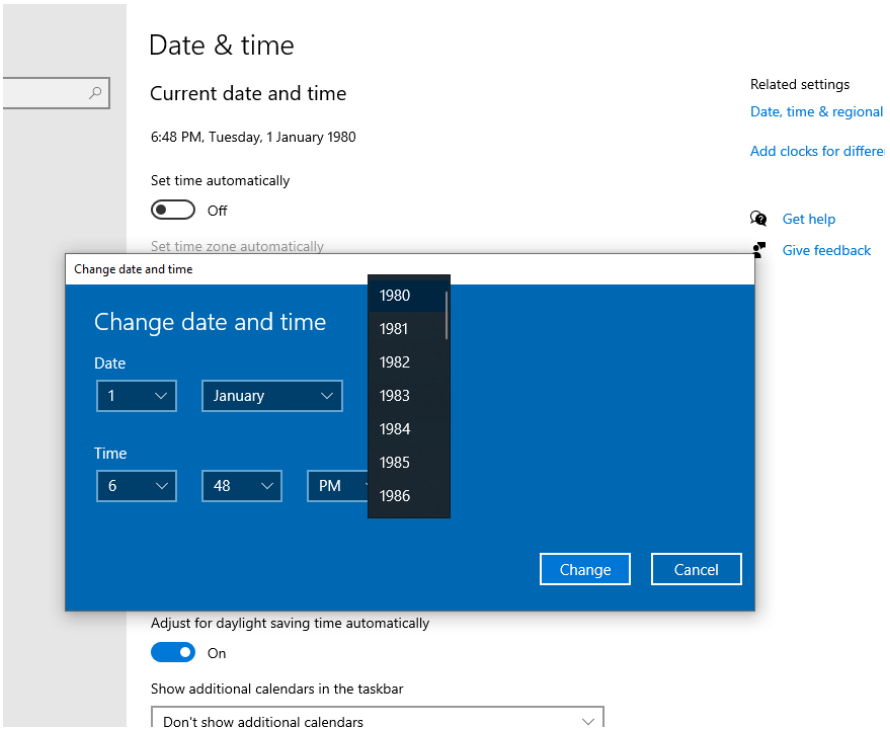
iv. I selected 1 year for demonstration purposes, but all aspects of the time are user-configurable. It would have been possible for me to advance the clock by a few minutes, or an hour, or days at my choice depending on the metadata I wanted to cause to be created.

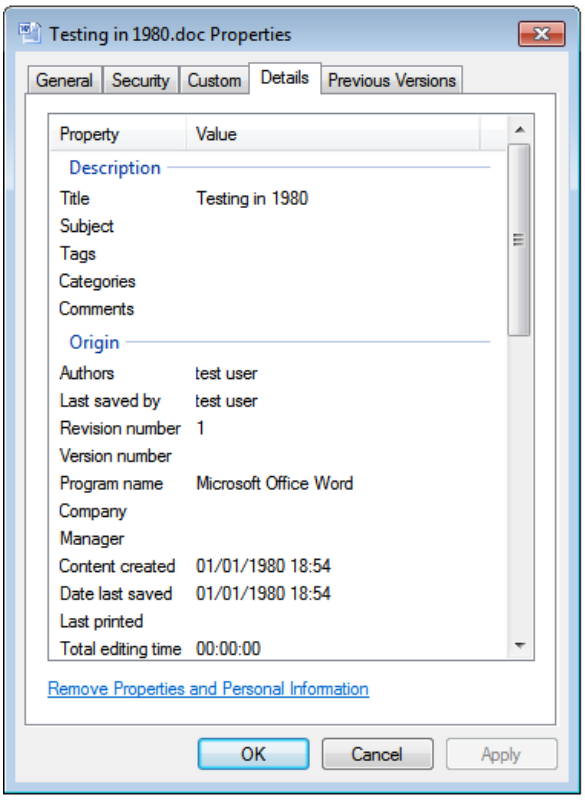
v. If the clock is altered backwards while an MS Word document is open, it will reverse or wind back the time. This will reduce the edit time, and can even result in a negative value edit time. A negative value edit time is not displayed correctly in MS Word, which lists it as an inordinately long edit time of many years.

b. **Revisions**. It can also be seen that the number of Revisions is just 2. This is consistent with the overall complexity of the document (which is just a few words for demonstration purposes and would not need many saves to retain the content as it was typed), but it is not particularly consistent with normal editing behaviour for a document that was actually edited for over a year. This demonstrates how it can be useful to review metadata fields not in isolation, but in combination with each other and the face value content of the document, to form a view as to whether they exhibit a record of normal rational behaviour or are anomalous or irregular.

c. **Last accessed:** It can be seen that at the point of my first save, the "last accessed" time stamp was set to 2008. After setting my clock back to the real time and accessing the file again, the Last Accessed timestamp was updated, but other fields were not. (Though I repeat here that the Last Accessed timestamp does not always update when a file is accessed, and it can depend on the circumstances as I have explained above).

It is possible to set the clock back further than could be practically required. In Windows 10, and the same stands for Windows 7, the furthest back the clock can be adjusted to within the Operating System is 1980. Below are two screenshots showing this setting being applied to a Windows 10 computer, and the properties of an MS Word 2003 document created when the clock was set to 1980.

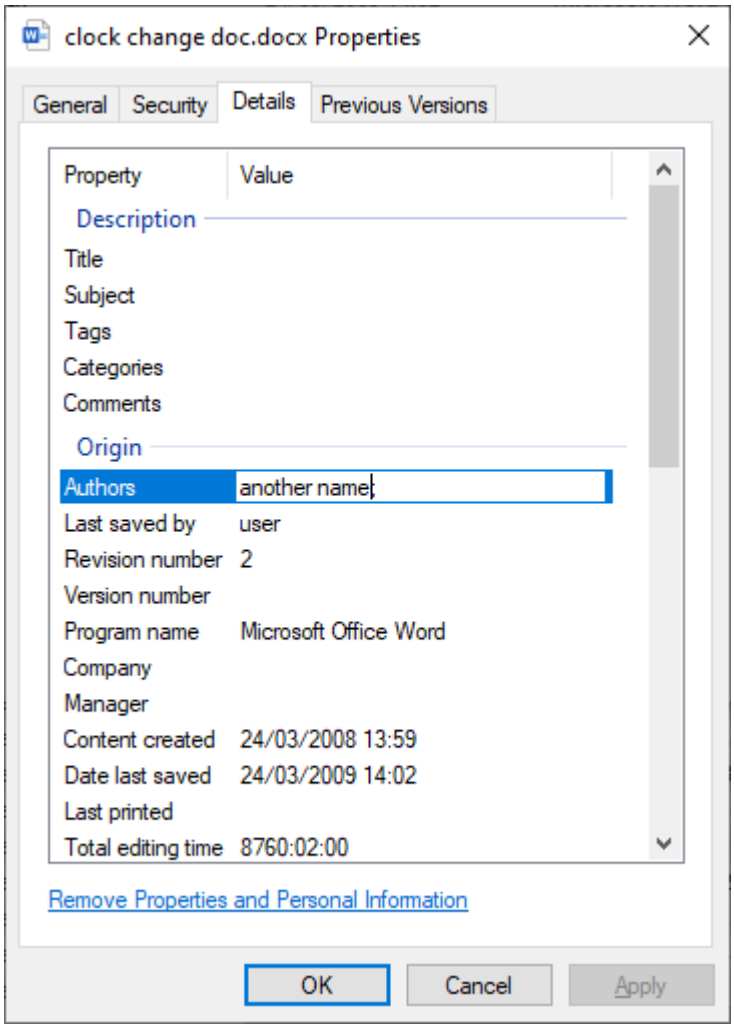




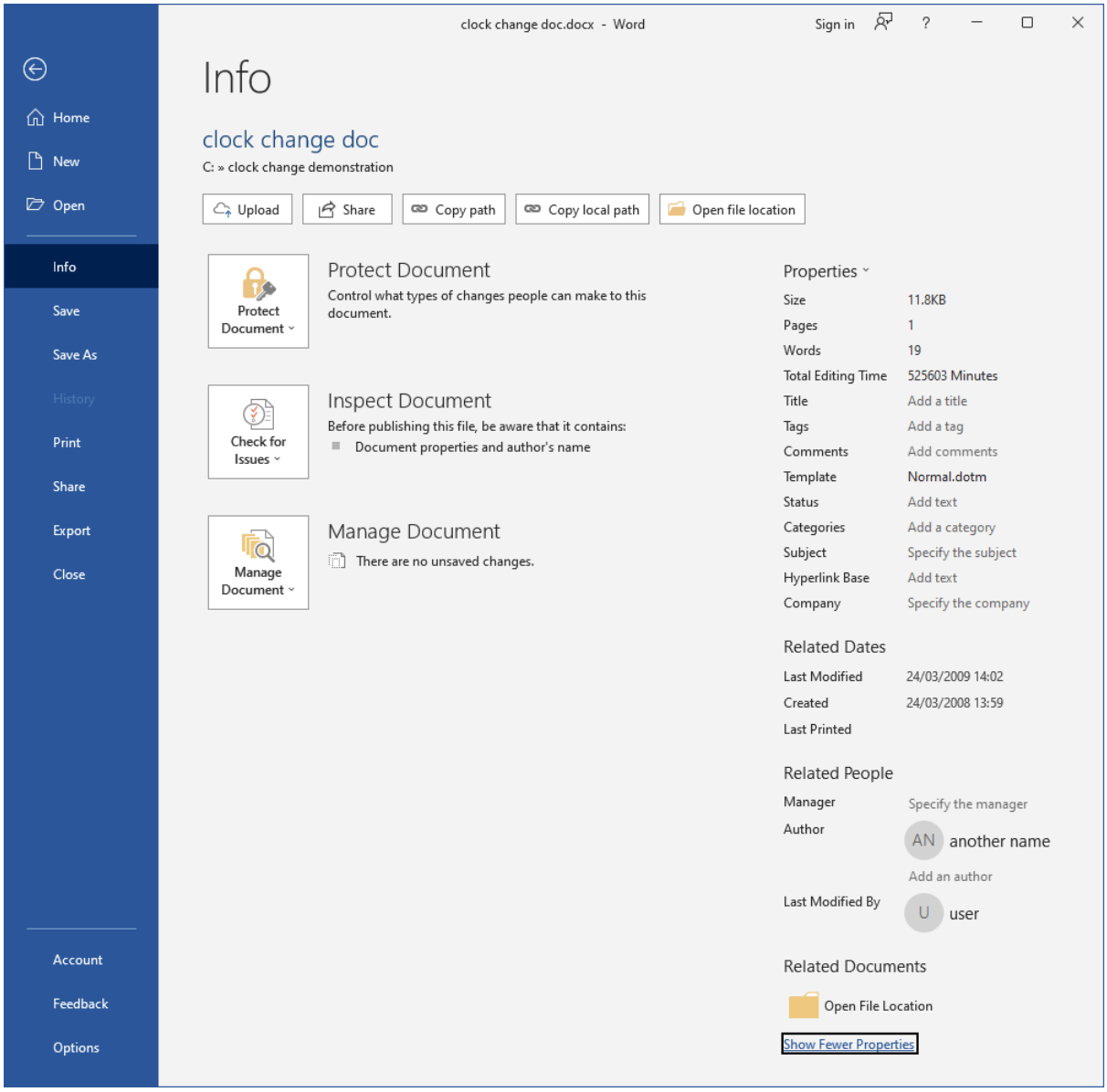
Manipulation of recorded author data

221. At this point, it is convenient to show also how the author data in a file can be updated at the same time, through direct editing.

222. As can be observed above, the username associated with my account, and which was used to record the Author and Last-saved-by metadata is simply "**User**". However by opening the properties dialog for the file it is possible for me to change the recorded original author by simple editing and typing in new information:

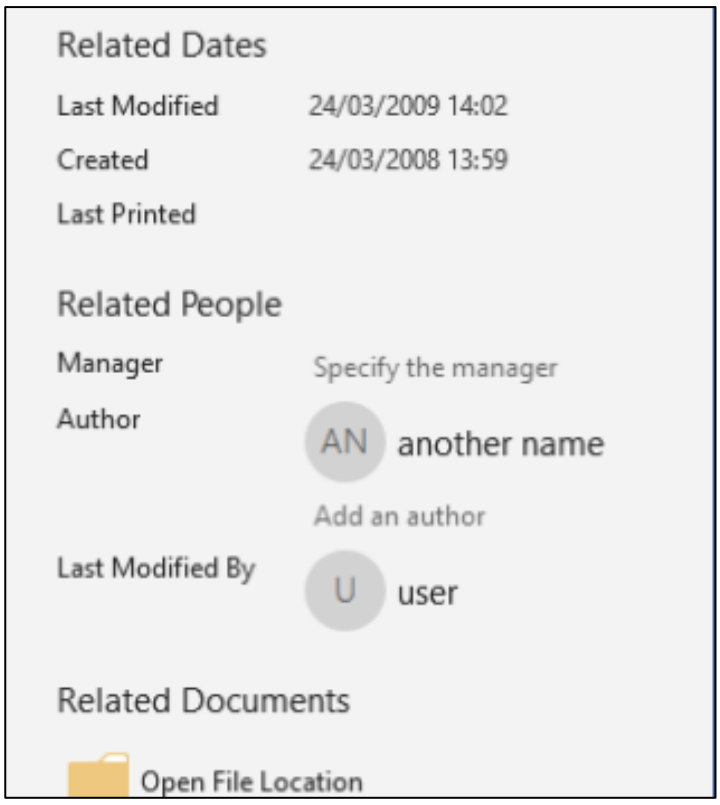


223. After saving that edit and pressing OK, opening the document within Microsoft Word records the change in the readily available internal metadata property view, even though no such user really exists:



Above – Shown in context

Below – a zoomed in screenshot of the same view:



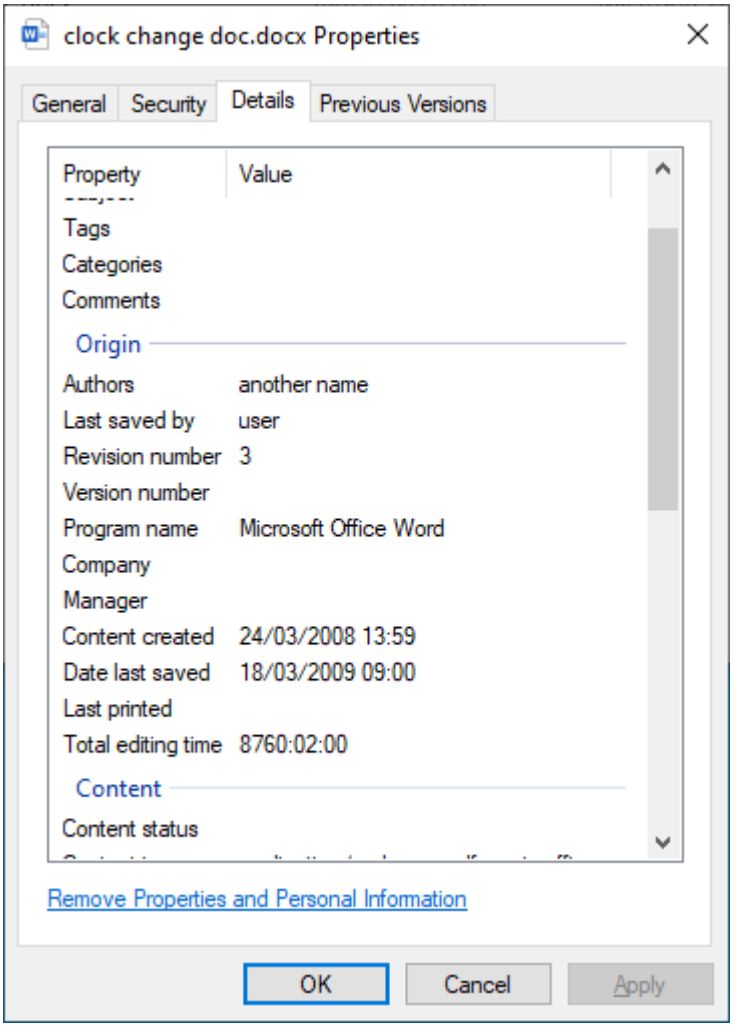
Further changes after saving, and further effects on metadata

224. Having made the changes above (including to the Last Accessed time which was set to the present day), it is helpful to show that the clock setting can still be used to make further changes to the metadata as follows.

225. I set my computer clock again but to a different date this time. Before, I had selected 24 March 2008 for my chosen 'Created' date and a year later on 24 March 2009 for my chosen "Last Modified" date. Had I been provided this document for forensic analysis, I would have remarked on the fact that the Edit Time was a very round "1 year", which I would have regarded as anomalous. It is possible to change that in the following way.

226. This time, I kept the same document open and selected a date a little earlier than 24 March 2009 and set the clock back to 9am on 18 March 2009. I then re-saved the document, which had the following effects:

    a. The Last Modified date was changed to my selected time on "18 March 2009".



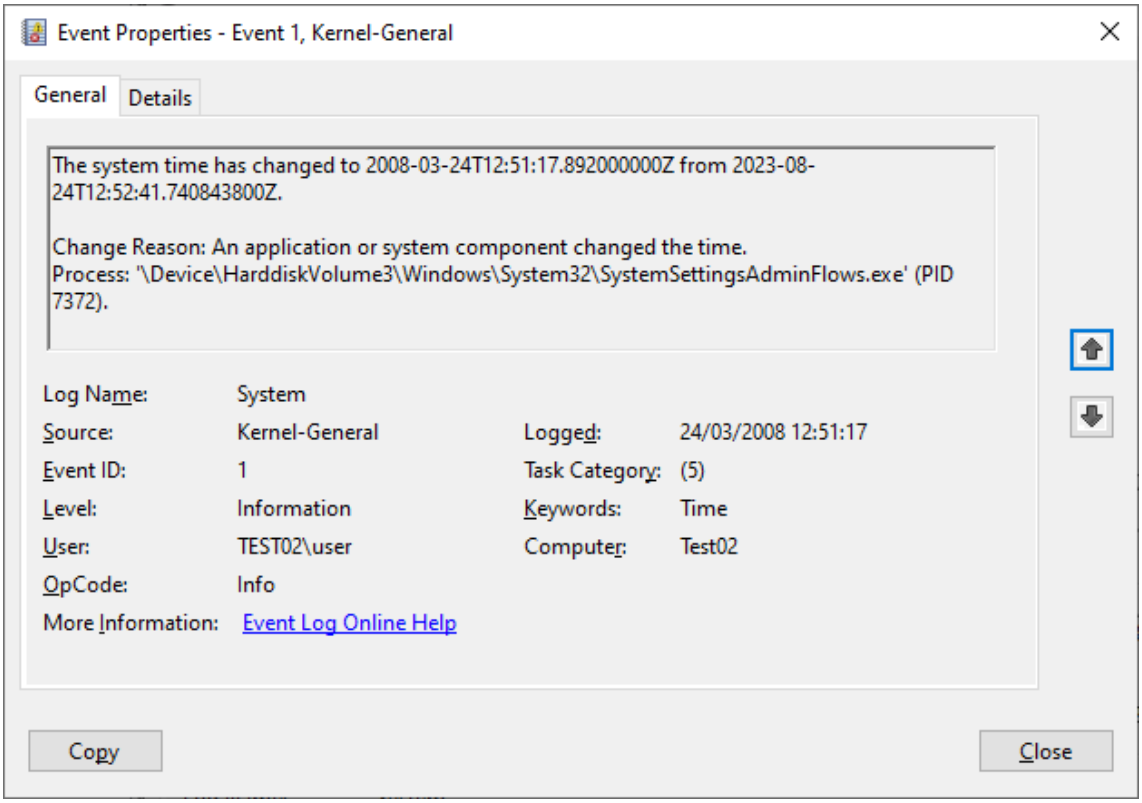
b. The OS file property "Last Accessed" was also changed to my selected time as well, overwriting the previous date which had been recorded in 2023 (as shown above).

c. As can be seen above, the internal Total editing time property however continued with the same counter. This resulted in a Total editing time remaining at 8760.02 at the point of saving, and which would continue to increase if I continued to edit the document thereafter. This is plainly incongruous with the other metadata because it is longer than the difference between the Created and Last Modified timestamps (an anomaly that I have observed within the dataset).

System logs

227. The documents, and the file metadata, will not typically contain any record of clock manipulation. However, when artefacts indicative of clock manipulation have been discovered, access to the computer equipment on which the documents were authored (or the forensic images of those computers) can help to establish whether clock manipulation has taken place.

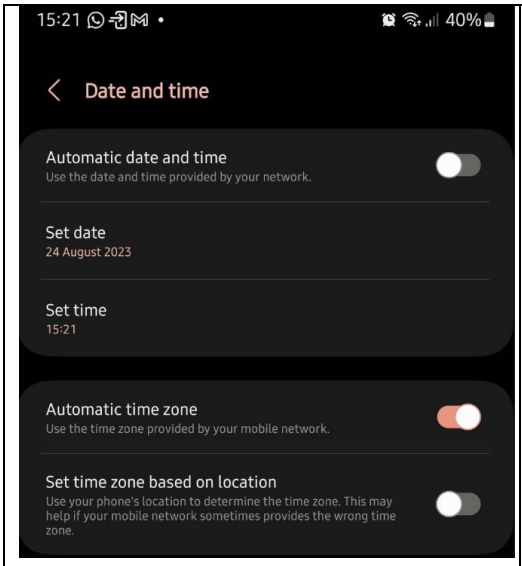
The following is a snapshot of the Windows Event Properties viewer, illustrating how a record resulting from the analysis above displays.



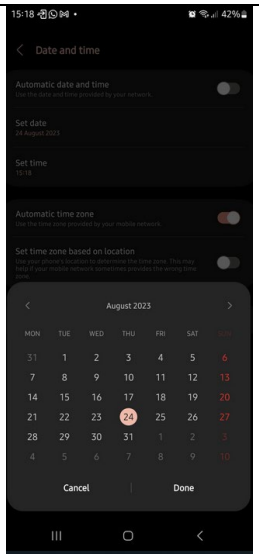
228. There are other logs which might record such data; this is just an example. Such logs can also be wiped or deleted and are not always accurate, but in this case I have not been provided with any access to the relevant images which would allow me to investigate further.

Other operating systems

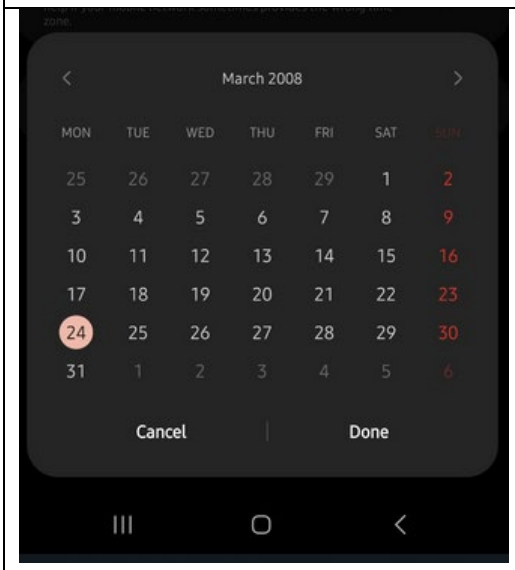
229. A similar process is available on Apple Mac computers and mobile devices such as iPads, iPhones, or Android devices. Setting the system time is a standard operation available on most computing systems, all of which work in broadly the same way. As an illustration of how it might look on a different computing system, the following screenshots have been taken by Bird & Bird on a Samsung Galaxy mobile phone running the Android operating system:



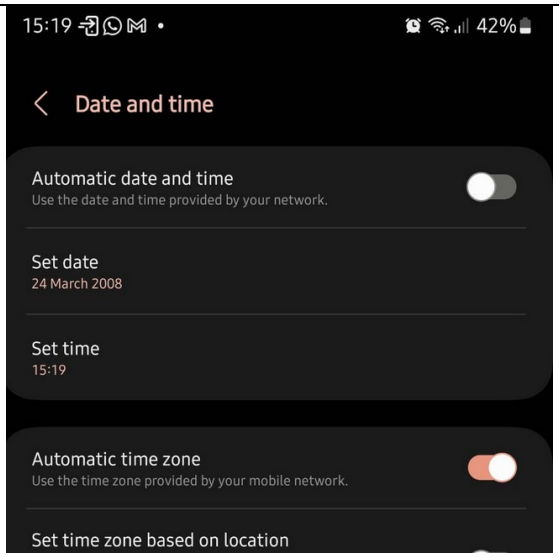
Android date and time settings showing current date



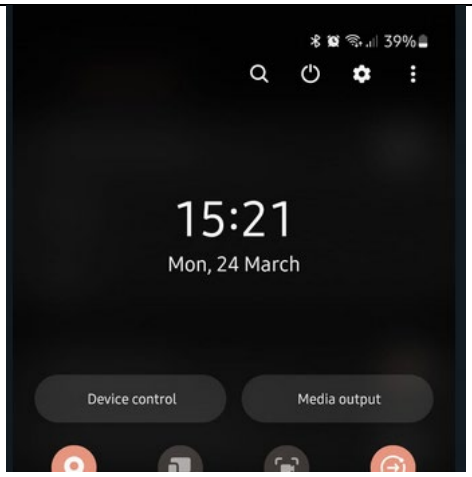
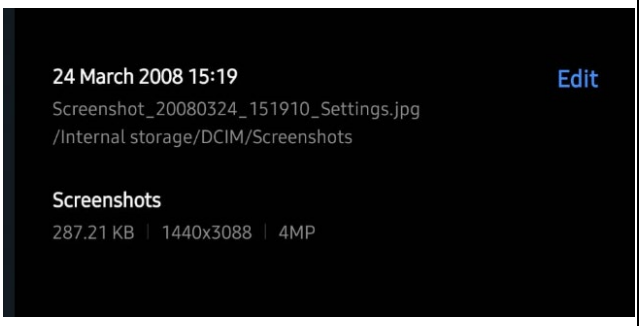
Whole-screen view showing how the date can be selected



Setting the date to 24 March 2008



Showing the date setting has taken effect in the system settings readout

	
The date and time as then displayed on the phone's main pull-down menu	*A view of the metadata for one of the screenshot files that was created – showing that it has been assigned metadata relating to a creation date of 24 March 2008 at 15:19 and that the filename has been automatically assigned with a timestamp as "Screenshot_20080324_151910".*

LIMITATIONS

230. I record here various limitations to my analysis.

231. First, as I have explained above, the disclosure dataset has omitted a large number of native format files. In other cases, the files provided are scans which do not retain any useful metadata corresponding to their creation.

232. I have also addressed above that where the authenticity of a document or documents is in question, it is prudent to analyse not just the documents themselves, but the environments in which they were authored and stored thereafter. The absence of access to the various additional sources of information identified above means that there are a number of irregularities or anomalies that I have observed as being consistent with manipulation, but which I have not had sufficient access provided to come to a firm conclusion about. In many cases, my analysis has identified multiple independent irregularities with such documents, which reduces the impact of that limitation. However, that is not the case for all of them, and I was not able to conduct the depth of analysis that I ordinarily would have done had fuller access been provided, such as

examination of comparator documents, file system information, application information, and system logs.

233. Importantly, the lack of access to forensic images or computing equipment may well mean that documents have had to be taken more in the abstract and may appear to be authentic even if they may not actually be. Apart from identifying that as a possible risk and therefore suggesting that documents which I have not addressed should not be assumed to be authentic without deeper investigation, without access to these sources of information it is not possible for me to overcome this problem.

234. As I have explained above, the external metadata provided with the disclosure dataset was provided in the form of a load file and appears to have been altered before being disclosed, in particular by the discarding of time zone information and timestamp precision. The approach taken appears to me to been inconsistent and muddled.

235. As I have also explained in several parts of the Report, there are some unexpected and therefore anomalous timestamps embedded within many of the email messages disclosed. Many of these timestamps are recent (2019 or 2020) and contradict the recorded dates of the messages, some of which date from 2014 or earlier. It is necessary to properly explore these timestamps from the original source material (i.e. the native email files). It might be the case that the inclusion of these timestamps relates to the export process from the e-disclosure database.

236. In cases where documents have been altered by their handling, it is often not possible to reliably distinguish between a document that has been altered by contamination as opposed to a document that has been altered imperfectly in the course of manipulation. In such cases, my analysis is inconclusive.

237. Another limitation of my analysis is that I have not meaningfully been able to analyse some types of document, as set out at the beginning of this Main Report.

238. There have been circumstances where I have set aside some threads of analysis in order to prioritise the analysis of other documents which I understood to be closer to the core issues. Similarly, I have prioritised the analysis of some documents over others within the time available where the required analysis is more hindered by the limitations of the disclosure dataset and indications of possible contamination by the handling of the documents.

CONCLUSIONS

239. I have formed the opinion that Dr Wright's disclosure dataset contains many documents that appear not to be original or authentic to their purported creation time. Many of these are Reliance Documents, which I understand that Dr Wright primarily relies upon as evidence to support his case. Others appear to me to be relevant to the issues in the case, as I understand them, but are not Reliance Documents.

240. In my view, a large proportion of the Reliance Documents that I have examined, as well as other documents which are apparently contextually relevant to the dispute, cannot be relied upon as being authentic to their purported dates and contain content and/or metadata which has been edited and manipulated.

241. I have borne in mind that it is not my role in the case to come to an opinion on why a document has been altered and what the purpose was of doing so, only to examine each document as it is presented to me and to come to a conclusion, if possible, on whether it is authentic based on a consideration of the information available to me and my experience of forensic review.

242. For example although the idea of "forgery" is a familiar and helpful concept for illustrating examples, I do not tend to describe specific documents as "forged" but only to comment on whether, based on the information available, they appear to be authentic or not.

243. In many cases, however, I have been able to come to conclusion not only about the fact that a document is unreliable, but have also been able to establish information about the process (or likely process) by which an alteration occurred:

    a. The minority of problems I have identified appear to me to have been the result of poor disclosure handling, by the use of processes which have contaminated some documents and their metadata. Wherever a problem has been observed which could be explained by (for example) poor handling and could be explained by making assumptions about how the handling took place, I have tended not to draw any conclusions based on those observations alone, making the assumption that they occurred as a result of some accidental handling process. However, in cases where irregularities are harder to explain this may lead me to doubt the authenticity of the documents presented. As an example of this, almost all email files in the dataset have been disclosed in .MSG format after a process of double conversion, which hinders analysis. In those cases, detailed forensic review of the computing equipment, forensic

computer images or other native sources of information might enable me to put those doubts to rest, but I cannot do so on the basis of the information provided.

b. The significant majority of problems I have found are indicative of actual manipulation of the content or metadata of documents, in a manner that I do not consider could be the result of an accidental contamination process.

c. Some documents do appear to be authentic (for example, I have taken some documents from the disclosure dataset as authentic control documents for the purpose of **Appendix PM3**). {H/20}

d. Some documents I have assessed as best I can, but my analysis is inconclusive due to degradation of quality (such as conversion to image files digitally or by hard copy scanning), or an inherent lack of information which means that forensic analysis has not been meaningfully possible (such as with plain text source code files).

244. Of the documents which I have found to be likely to be inauthentic:

a. Many contain very obvious indications of manipulation or tampering. This includes the following examples:

i. There are examples of email messages where the Transmission header exhibits a timestamp that contradicts the recorded transmission of the messages, indicating that they are not original to the date on their face;

ii. There are email items where emails that are otherwise nearly identical exhibit two very different dates, suggestive of one being created from the other by manipulation;

iii. There are MS Word documents that contain redundant revision text that include references that post-date their purported authorship, for example to facts that would not be possible to know at the purported time of authorship;

iv. There are MS Word documents that exhibit implausible or impossible Edit Times;

v. There are several documents that record their creation by reference to software which did not exist at the purported time of creation, or include fonts that did not exist at the relevant time;

vi. There are PDF files that have been edited to alter the content, the names and addresses of email From/To fields, and their purported dates and times;

vii. In one case, there is an accounting database apparently relating to Bitcoin which purports to be from 2009-2010 but which I have determined was entirely authored in March 2020;

viii. In another case, there is a document which has been first printed and scanned, but which contains digital timestamps that I have established are not authentic to the purported creation date and which have been digitally altered before printing.

b. Other documents contain indications of manipulation which are less obvious, but which can be uncovered on a detailed analysis.

c. I have also identified a number of running themes within the dataset, whereby groups of documents exhibit common characteristics that are typical of or associated with manipulation, tampering or the creation of new documents that purport to be historic (backdating). In some cases, the documents bearing those characteristics are very clearly inauthentic for multiple reasons. In other cases, the analysis is less immediately obvious, but I have observed a number of irregularities that while not determinative on their own, lead me to conclude that they cannot be taken as authentic without both a proper explanation of the problems they exhibit, and access to the computer equipment on which they were authored to allow those explanations to be tested.

245. Although my review has been as thorough as possible in the time available to me (and it has occupied almost my entire time since I was provided with the disclosure dataset), it is my view that I have not been able to thoroughly investigate all of the documents in the dataset provided. There are several limitations to the manner of production that have hindered my ability to efficiently conduct a comprehensive forensic examination of the disclosure dataset as a whole. I consider that if full access to the original forensic images had been provided, I would have been able to review the documents in the dataset more thoroughly and more quickly. As a result of

this, it is my view that other documents in the disclosure dataset cannot be assumed to be authentic without the provision of further information and access.

246. The remainder of my conclusions is set out in each of my Appendices.

DECLARATION

1. I understand that my duty is to help the Court to achieve the overriding objective by giving independent assistance by way of objective, unbiased opinion on matters within my expertise, both in preparing reports and giving oral evidence. I understand that this duty overrides any obligation to the party by whom I am engaged or the person who has paid or is liable to pay me. I confirm that I have complied with and will continue to comply with that duty.

2. I confirm that I have not entered into any arrangement where the amount or payment of my fees is in any way dependent on the outcome of the case.

3. I know of no conflict of interest of any kind, other than any which I have disclosed in my report. I do not consider that any interest affects my suitability as an expert witness on any issues on which I have given evidence.

4. I will advise the party by whom I am instructed if, between the date of my report and the trial, there is any change in circumstances which affects this.

5. I have shown the sources of all information I have used.

6. I have exercised reasonable care and skill in order to be accurate and complete in preparing this report.

7. I have endeavoured to include in my report those matters, of which I have knowledge or of which I have been made aware, that might adversely affect the validity of my opinion. I have clearly stated any qualifications to my opinion.

8. I have not, without forming an independent view, included or excluded anything which has been suggested to me by others including my instructing lawyers.

9. I will notify those instructing me immediately and confirm in writing if for any reason my existing report requires any correction or qualification or my opinion changes.

10. I understand that:
    a. my report will form the evidence to be given under oath or affirmation;

    b. the court may at any stage direct a discussion to take place between experts and has done in this case;

    c. the court may direct that, following a discussion between the experts, a statement should be prepared showing those issues which are agreed and those issues which are not agreed;

    d. I may be required to attend Court to be cross-examined on my report; and

    e. I am likely to be the subject of public adverse criticism by the judge if the Court concludes that I have not taken reasonable care in trying to meet the standards set out above.

11. I have read Part 35 of the Civil Procedure Rules and I have complied with its requirements. I am aware of the requirements of Practice Direction 35 and the Guidance for the Instruction of Experts in Civil Claims 2014.

12. I confirm that I have acted in accordance with the Code of Practice for Experts.

13. I confirm that I have made clear which facts and matters referred to in this report are within my own knowledge and which are not. Those that are within my own knowledge I confirm to be true. The opinions I have expressed represent my true and complete professional opinions on the matters to which they refer.

Signed: DocuSigned by: Patrick Madden 5943D537458F4C0... Date: 1 September 2023

Annex 1: List of Reliance Documents

ID_000050	ID_000530	ID_003052	ID_003548	ID_004015
ID_000051	ID_000531	ID_003053	ID_003549	ID_004018
ID_000071	ID_000549	ID_003163	ID_003565	ID_004077
ID_000095	ID_000550	ID_003330	ID_003566	ID_004078
ID_000127	ID_000551	ID_003464	ID_003567	ID_004079
ID_000128	ID_000553	ID_003465	ID_003568	ID_004090
ID_000183	ID_000554	ID_003506	ID_003569	
ID_000184	ID_000556	ID_003507	ID_003570	
ID_000194	ID_000557	ID_003508	ID_003571	
ID_000195	ID_000568	ID_003509	ID_003572	
ID_000199	ID_000569	ID_003510	ID_003573	
ID_000216	ID_000856	ID_003511	ID_003574	
ID_000217	ID_001379	ID_003512	ID_003576	
ID_000227	ID_001916	ID_003513	ID_003702	
ID_000254	ID_002261	ID_003514	ID_003840	
ID_000258	ID_002262	ID_003515	ID_003847	
ID_000260	ID_002330	ID_003516	ID_003860	
ID_000367	ID_002331	ID_003517	ID_003879	
ID_000371	ID_002710	ID_003518	ID_003904	
ID_000385	ID_002742	ID_003519	ID_003986	
ID_000386	ID_002972	ID_003534	ID_003987	
ID_000387	ID_002973	ID_003535	ID_003992	
ID_000388	ID_002974	ID_003536	ID_003993	
ID_000395	ID_002975	ID_003537	ID_003994	
ID_000396	ID_002976	ID_003538	ID_003995	
ID_000450	ID_002977	ID_003539	ID_003996	
ID_000462	ID_002978	ID_003540	ID_003998	
ID_000491	ID_002979	ID_003541	ID_004000	
ID_000493	ID_002980	ID_003542	ID_004009	
ID_000504	ID_002981	ID_003543	ID_004010	
ID_000525	ID_002982	ID_003544	ID_004011	
ID_000527	ID_002983	ID_003545	ID_004012	
ID_000528	ID_002984	ID_003546	ID_004013	
ID_000529	ID_003051	ID_003547	ID_004014	