

Witness Name: Christopher Michael Jackson

Statement No.: WITN10810100

Dated: 19 December 2023

## POST OFFICE HORIZON IT INQUIRY

---

### FIRST WITNESS STATEMENT OF CHRISTOPHER MICHAEL JACKSON

---

I, Christopher Michael Jackson, will say as follows<sup>1</sup>:

#### A. SCOPE AND STRUCTURE OF THIS WITNESS STATEMENT

1 I am a partner (member) in Burges Salmon LLP. I have been Post Office Limited's ("**Post Office**") recognised legal representative ("**RLR**") for the Post Office Horizon IT Inquiry (the "**Inquiry**") since 1 September 2023.

---

<sup>1</sup> The content of this witness statement generally reflects the position on 1 December 2023 when I filed my draft statement as required, which the Inquiry has reviewed. Intensive work has continued since 1 December 2023. In the final version of this statement, I have referenced 2 letters sent to the Inquiry dated 15 December 2023 as they provide updates on relevant points and can be briefly cross-referenced. Otherwise, I will provide relevant updates to the Inquiry in correspondence and oral evidence as appropriate rather than amend content that the Inquiry has reviewed.

2 This witness statement is made to assist the Inquiry with the matters set out in the Rule 9 requests addressed to me and dated:

(a) 31 October 2023 (the "**October Request**"); and

(b) 17 November 2023 (the "**November Request**")

made further to the Inquiry Chair's directions dated 15 September 2023 (the "**Directions**"). I have below also referred to the October and November Requests together as "**the Requests**".

3 In overview:

(a) the October Request requires clarification and explanation of points of engagement relating to the mechanics and practicalities of disclosure that were set out in letters from my team to the Inquiry dated 13 [WITN10810101] and 16 October 2023 [WITN10810102]; and

(b) the November Request requires explanations of: (i) the events that led to the Mimecast/Exchange issue and what is being done to address the resulting problems and (ii) the structural (EDRM) disclosure review that was summarised in my letter of 1 September 2023 to the Solicitor to the Inquiry [POL00126339] as I became RLR.

I have copied or paraphrased (where clearer to do so) the relevant questions into the headings and sub-headings of the sections of this statement that answer each question.

- 4 The explanations and clarifications required by the Requests are best provided in the context of the overall position. Further, the position and thinking on behalf of Post Office has developed since the mid-October letters that are the subject of the October Request. For example, BSFf have been working closely with KPMG LLP ("**KPMG**") on how the issues of repeating copies of documents can be best addressed to assist the Inquiry in its review of material whilst still making available to the Inquiry near duplicates and documents that make clear the different contexts in which copies of documents appear and reappear, which are required by the Inquiry.
  
- 5 I anticipate that it will be more helpful first to set out the wider position before drilling down into the specifics. I am also conscious of the detail required to respond fully and properly to the questions and issues covered by the Requests<sup>2</sup>. I have therefore set out responses in the following sections and sequence below:

---

<sup>2</sup> Many of the communications between BSFf and Post Office (and other communications in connection with its Inquiry participation) are subject to legal advice privilege. I have however aimed to give full, and I hope helpful, answers and explanations on each of the points in a way that means that privilege does not get in the way. However, I do not have authority to waive any legal professional privilege and nothing in my evidence is intended to do that.

- (a) **Section B** (from page 9): an overview of my professional background and that of the BSFf team; the scope and timing of our assistance to Post Office in the Inquiry and how that links to the co-ordination and co-operation with other professional advisers.
- (b) **Section C** (from page 14): a short overview of factors relating to disclosure (generally and Post Office specific) that are directly relevant to the difficulties and problems that have occurred and to the options to resolve them.
- (c) **Section D** (from page 23): the Mimecast/Exchange issue, the sequence of events relating to it, and the proposed solution (responding to paragraph 1 of the November Request).
- (d) **Section E** (from page 69): The structural (EDRM) review (responding to paragraph 2 of the November Request): the reasons for it; the work involved and remaining.
- (e) **Section F** (from at page 79): responding to paragraph 1 of the October Request (relating to BSFf's letter to the Inquiry dated 16 October 2023 [WITN10810102]).
- (f) **Section G** (from page 90): responding to paragraphs 2-4 of the October Request (relating to BSFf's letters to the Inquiry dated 13 [WITN10810101] and 16 October [WITN10810102]).

6 Where my knowledge and belief set out in this witness statement has been materially informed by another person or by documents that I have

reviewed, I acknowledge that person or those documents. Colleagues from BSFf have assisted me in preparing this witness statement. I have in this statement explained my understanding of technical or system issues. Except where expressly stated, I do not have technical knowledge of particular disclosure technologies so I would need to defer to others if it would be useful also to drill down into any specific issue involving the detail of any of the disclosure technologies.

7 Other external advisers also advise Post Office, including Herbert Smith Freehills LLP (“**HSF**”), Peters & Peters Solicitors LLP (“**P&P**”) and KPMG and they have provided factual and, in the case of KPMG, technical input to me on disclosure issues.

8 The current situation is not one that anyone would wish to see continue. Post Office has asked me to convey its apologies for the current situation and to assure the Inquiry and other Core Participants that it is a Post Office priority to get to a position where hearings (and planning and preparation for hearings) can take place from a stable basis with the risks of further emerging data source issues minimised and managed so far as is practicable.

9 I made the following observations and commitments in my letter to the Inquiry of 1 September 2023 [POL00126339]:

*“We are mindful of the seriousness of the issues and events being reviewed by the Inquiry and the acute human and other impacts that those*

*have had upon the Postmasters and others affected. Both in terms of our approach and our instructions from POL, that awareness informs and underpins all aspects of our work for POL during the rest of the Inquiry and to the inputs that the Inquiry will understandably demand of POL. That of course applies to the points set out below.*

*Neither I nor other members of the BSff team had any prior involvement with any work for the Post Office nor the matters that gave rise to the Inquiry's Terms of Reference. As a team we therefore recognise that we do not yet fully understand everything that has gone before, nor all of the complexities. We will however continue to work intensively to get across those issues and to engage frankly and constructively with the Inquiry and with those representing the Postmaster Core Participants ("CP's") and other CPs.*

*The issues being considered at the September hearing have, necessarily, involved significant amounts of detailed explanation in witness statements, disclosure statements and correspondence. The purpose of this letter is to seek to stand back from that detail and to provide a frank overview of the position based on our current understanding and our assessment since our appointment. The work to build that understanding is ongoing; we are seeking to take a structured and systematic approach to it.*

*As various of the witnesses for the September hearing have confirmed from their own perspective, my understanding and direct observation is also that POL's instruction and wish is to provide all relevant evidence that the*

*Inquiry wishes to see, so that the full factual position can be examined and become known. That is the attitude and instruction from the POL team with whom we are working, the great majority of whom have also come fresh to the issues that are being examined by the Inquiry.*

*I have been instructed by POL (and it would in any event be my intended approach professionally) to flag to the Inquiry if ever there were to be an attempt to withhold evidence that should be disclosed in relation to the Terms of Reference and the events leading up to the Inquiry being set up. I sense however that that is unlikely to arise; the issues faced are really those of scale, complexity and practicability.*

### ***Proposed Engagement***

*My aim and request is that there can be continued (formal and minuted as necessary) engagement with the Inquiry's senior team on these critical issues so that the Inquiry is updated on the work POL is undertaking. We hope such an approach will best support the Chair to continue to plan for the vital remaining stages of the Inquiry. Whilst we will provide updates in correspondence, with issues of this complexity we consider that the ability to have a discussion on points of concern may be beneficial for the Inquiry and for POL in assisting it.*

*That is of course a matter for the Inquiry to consider but I reiterate that I, and colleagues, are happy in that context to meet with you and your colleagues regularly and as you would find helpful. I will also, as and if*

*necessary, attend as RLR any future disclosure hearings to provide formal updates.”*

- 10 I re-affirm those comments and commitments. All the BSFf team’s efforts for Post Office will, throughout, remain focused on giving the best professional support practically achievable to assist the important work of the Inquiry.
- 11 In the answers to the Requests, I have aimed to include proposed solutions. I am conscious that emerging problems with, and frank updates to the Inquiry on, Post Office’s disclosure have been deeply and understandably frustrating to the Inquiry, to Postmasters and their families (including, in particular, those who have been attending on matters of great importance to them only then to see hearings cancelled at short notice) and to those witnesses who have been affected. I understand fully the reasons for those reactions and for the profound distrust in many quarters, which is the starting point for any exchanges on disclosure given the underlying earlier events relating to Horizon that the Inquiry is charged to investigate.
- 12 However, I confirm that all my experience acting for Post Office since May 2023 indicates to me that all the professional advisers working for Post Office on the Inquiry (external and internal to Post Office<sup>3</sup>) are behaving

---

<sup>3</sup> I have summarised below the current numbers within the Post Office Inquiry team as well as those for BSFf. As with the BSFf team, in practice most of the Post Office Inquiry team now in place started work during the course of 2023 and have had no, or little, involvement with the facts, actions and approaches that have given rise to the Inquiry’s Terms of Reference.



properly and professionally, working intensively and with significant resource, to provide all requested evidence to the Inquiry. Were it ever to be suggested otherwise that would be a matter of profound professional concern.

## **B. OVERVIEW OF PROFESSIONAL BACKGROUND AND SCOPE OF BSFF WORK**

### **Qualifications and professional background**

- 13 I am a solicitor and solicitor-advocate (civil). I have been in practice since 1988. I trained at Macfarlanes and qualified as a solicitor in 1990. I moved to Burges Salmon in late 1991 and became a partner (then under the Partnership Act 1980) in 1997. I have been a member (under the Limited Liability Partnership Act 2000) of Burges Salmon LLP since 2004.
- 14 Since 1990 I have worked on, and since 1997 have been a partner leading teams in, complex and/or large-scale matters for public and private sector organisations, including at various times, commercial litigation disputes, public inquiries and inquests, major procurement challenges, judicial reviews, criminal prosecutions and matters relating to strategic safety issues and economic and safety regulation.
- 15 Whilst the subject matter and sectors of those cases and projects has varied significantly, the main underlying common thread has been complex organisational or project failings or problems and the risk factors that led to them. For example, I (or other partners in our immediate team) have been

involved for an organisation involved in all major UK rail accidents, including the resulting investigations and inquiries, since privatisation. Procurement challenge matters have, since 2008, involved work in a range of sectors including IT systems, education, nuclear, defence and transport for public authorities and bidding entities. I have also been involved with other public inquiries and inquests.

- 16 Those categories of work in public inquiries or litigation have often involved complex, large-scale disclosure exercises for central government, public corporations or other entities similar to Post Office or for private organisations. However, the scale, challenges, complexities and problems faced here in relation to the Inquiry are of a greatly different order of magnitude and difficulty even to those in other very large-scale situations and projects. I note that Gregg Rowan made similar observations at paragraphs 36-41 of his statement [WITN09950100] for the 5 September disclosure hearing.
- 17 The Chair confirmed my designation as RLR on 30 August 2023 and I replaced Mr Rowan in that role with effect from 1 September 2023. HSF continues to assist Post Office in relation to certain issues on the Inquiry and related matters, including (with the material involvement and assistance of P&P on criminal matters – see Mr Rowan’s witness statement [WITN09950100] for further details of the firms’ respective roles) the operational conduct of Inquiry Phase 4.

18 I have given further details below on the respective roles of the various firms and the co-ordination between us.

**BSFf**

19 Neither I nor, to the best of my knowledge or understanding, BS or Ff has had any professional role or involvement assisting Post Office generally or in relation to the Horizon IT system prior to being appointed 6 months ago in May 2023. Neither BS nor Ff is or has been on any Post Office panel for legal or other work.

20 Post Office has engaged BS from the Crown Commercial Service (CCS) framework RM6179. It ran a competitive, regulated procurement process commencing in February 2023 for services to support it in Phase 5 of the Inquiry onwards, including preparatory work. BS was formally appointed for that scope from 22 May 2023. Mobilisation and work had started shortly before that date.

21 Ff is BS's approved CCS Key Sub-Contractor relating specifically to public inquiry and complex inquest work. However, in practice BSFf works closely as a combined team, to deliver Inquiry-related services to Post Office. Effectively operationally – as opposed to legally/contractually – it is an integrated joint venture intended to provide greater depth of resource, experience and combined skills than either of the firms could provide

individually<sup>4</sup>. Although Rule 6 of the Inquiries Rules 2006 requires the identification of one RLR, the BSFf team is jointly led by me and Oliver Carlyon, an Ff partner.

- 22 The nature of the disclosure exercise required by the Inquiry necessitates a very significantly resourced legal team. The combined BSFf team working wholly or predominantly on the Horizon Inquiry over recent months is now over 170 professionals (including document reviewers and project managers but not including business support colleagues such as those in Finance and IT team). That team is very large relative to any with which I have previously been involved, or am personally aware of, in other (even very significant) inquiries or litigation. The BSFf team has continued to work in parallel, and collaboration with, HSF, P&P and the Post Office team such that the total number of professionals now working on these issues over recent months has exceeded 350.

### **Scope of involvement and responsibility**

- 23 BSFf mobilised to assist Post Office during May 2023. There had been a senior team short introduction meeting with Post Office, HSF and Counsel on 29 March 2023 and then transition briefing meetings with Post Office

---

<sup>4</sup> BS has been involved in public inquiries and major inquests for clients as core participants. Ff has separately had a long - established practice advising public inquiries, as well as core participants in other inquiries. The collaboration was established in 2022 to combine the resources and approaches of the two teams following the establishment of CCS legal panel RM 6179. BSFf currently acts as legal adviser to several UK public inquiries. That work is however (with limited overlaps) mainly carried out by colleagues not working on the Post Office Horizon IT Inquiry.

and HSF following appointment during June, July and August. Mobilisation and transition were considerable undertakings given the size and complexity of the Inquiry and the fact that, by that stage, it had been ongoing for approximately 3 years and the Inquiry's Terms of Reference relate to a period of over 20 years, covering events from the late 1990s to the recent past.

- 24 BSFf assists Post Office on structural matters that might affect Post Office's support to the Inquiry (I provide more detail on that work below) and we assisted Post Office for the 5 September 2023 disclosure hearing. August was an intensive period for that reason, combined with the multiple areas of work for Inquiry Phases 5-7.
- 25 Issues affecting different phases require BSFf, HSF and P&P to collaborate together and with the Post Office Inquiry team and other teams within Post Office. Where such issues have arisen to date, the collaboration has been, and continues to be, regular, with a number of meetings each week, and is constructive.
- 26 Outside direct involvement with the Inquiry, HSF and P&P also assist Post Office in relation to matters with similar and related facts and issues, such as Post Office's Horizon Shortfall Scheme and Overturned Convictions Scheme (in the case of HSF) and Criminal Cases Review Commission and Appeal cases (in which P&P acts). BSFf does not assist Post Office on those areas of work.

27 From late July, BSFf has been working on detailed disclosure requirements relating to Phase 5 under statutory notices served by the Inquiry, detailed forward planning and preparation for Inquiry Phase 5 and work on Inquiry Phases 6 and 7 issues. We have noted the very great intensity and pressure of the work involved across all Inquiry Phases.

### **C. DISCLOSURE – OVERVIEW AND POST OFFICE**

28 I summarised in paragraph 11 of my letter to the Inquiry of 1 September 2023 [POL00126339] my understanding of the causes of the scale and complexity of Post Office disclosure:

*“My understanding from what we have seen since May 2023 is that this is down to a combination of factors including (but not exclusively):*

*(a) POL’s own long and complicated organisational history and internal structures over decades (and longer) including a demerger during the last 20+ years during which the Horizon problems and events have occurred.*

*(b) Multiple sites and the absence until recently of any ‘data universe’ map of hard copy and electronic repositories (locations and systems) of potentially relevant documents leading to emerging sources from both ‘known unknowns’ but also ‘unknown unknowns’.*

*(c) Multiple document systems (current and historic) and interactions between different systems.*

*(d) A complicated mix of hard copy, digital and e-media sources from various different eras and without any central record. Some sources are local, others central, or are a hybrid of both.*

*(e) The evolution (through the collation and adding of different source repositories from different providers and at different times with different methodologies) of the Relativity database operated by KPMG for POL. This is also complicated by system constraints on all disclosure databases including Relativity. Functionality and usability declines materially once databases get above a certain size. I am not a technical e-disclosure expert but my understanding is that the 60million documents currently held are approximately 30 Terabytes of data in total and that a Relativity review workspace database starts to have serious functionality problems at or around 10Tb.*

*(f) The scale of data involved (as others have confirmed, now over 60 million documents with more inevitably to be found as the data mapping continues and specific requests for Phases 5-7 are formulated and targeted).*

*(g) As a result of different inputs from different sources and providers, variability in data quality and therefore also functionalities (for example email threading or use of CAL – computer assisted learning – or TAR – Technology Assisted Review – that would ordinarily be available and are commonly used in Relativity disclosure projects being either not available or only partially available.*

*(h) The need to respond swiftly to incoming evidence requests as the Inquiry evolved, potentially led to a focus on responding to individual requests, whilst balancing the factors brought into play in all large disclosure exercises of scope vs time vs avoidance of irrelevant material etc.*

*(i) Practical difficulties in the use of search terms on issues which – necessarily – are not always easily defined – for example processes, bugs/errors/defects and other terms used in a wide variety of contexts – some highly relevant to the Inquiry and others not so.”*

29 Subsequent work has strengthened that view. The scope of the Inquiry is necessarily wide in time and range of issues. Historic data governance problems, many of which were embedded within Post Office’s data landscape over many years, have risen to the surface under the scrutiny of the Inquiry and Post Office’s internal and external Inquiry teams.

30 These also link to wider dynamics in complex disclosure exercises. I am conscious that the factors summarised in paragraph 31 below are well-known to the Inquiry and to Core Participants. However, I have reprised them briefly for context because of the perceptions that have arisen in the context of the recent problems in Post Office’s disclosure and the consequent regrettable disruption to hearings and to the individuals involved in those hearings. The factors also feed into the proposals set out in section E and F in terms of what will be required to get to a position of



greater confidence, to the timings involved and to the levels of residual risk. They are also relevant to the October Request (Sections G and H below).

31 Unfortunately, no large and/or complex modern disclosure exercise can in practice be configured to produce every document within an organisation's custody and control that responds to the applicable terms of reference. Rather they can only be designed and run to produce the best achievable evidential results available by reference to the constraints of time, resource, knowledge, technology and complexity in the particular situation<sup>5</sup>. Based on my experience and discussions I have had over many years with professionals involved in disclosure exercises, the main reasons for that include<sup>6</sup>:

- (a) Before the mid to late 1990s, an organisation's records often mainly comprised hard copy documents stored in identifiable, physical locations. Digital technology resulted in massive proliferation of data and repositories and very significant increases in the number of documents, communications and other data created and retained.
- (b) Systems change organically and rapidly as technology evolves, becomes out of date or redundant and is replaced, often without any

---

<sup>5</sup> The Inquiry's Disclosure Protocol fairly reflects this reality in confirming that searches should be "*reasonable in all of the circumstances*" and "*comprehensive, thorough and rigorous*".

<sup>6</sup> See Post Office's letter to the Inquiry dated 10 September 2021 [WITN10810103] and HSF's letter to the Inquiry dated 15 October 2021 [WITN10810104].

central records or overarching system design. The ability to locate responsive documents or repositories is also often inhibited by loss of corporate memory/knowledge over time as people leave and by restructurings (as happened in 2012 with the Royal Mail Group and Post Office separation).

- (c) Additionally, document volumes are too vast for every document to be reviewed manually. I understand from KPMG that, on some analyses, over 80 million documents are now held on its Relativity platform for the Inquiry – on a conservative estimate there are at least 70 million documents – and that is only a portion of the documents in Post Office's data universe. As an indicative calculation based on a relatively high review rate of 40 documents per hour, a very large team of 100 reviewers each working full time (8 hours per day, 200 days per year) would, in perfect circumstances take nearly 11 years to complete a first level manual review of 70 million documents. The review rate would also be slower if, for example, the issues to be coded needed to be complex.
- (d) Parameters – for example – search date ranges for specific searches, custodians to be searched, search terms or other techniques to be used, repositories required to be investigated are therefore commonly (in litigation) discussed closely between the parties and the subject of direction and/or guidance from the court and (in inquiries or complex inquests) the subject of operational

discussions and meetings involving the relevant core participant(s) and inquiry team(s).

- (e) Post Office's data universe is more diffuse and complex than that of many other organisations that I have seen professionally but in common with others comprises many "live" electronic data sources, "dead" electronic data sources (some of this data is likely to have been retained but not actively considered for some time and some of this data is likely to have been deleted or lost as part of normal cycles of change and data/document disposal<sup>7</sup>), eMedia<sup>8</sup> (such as CDs or USB drives) and hard copy documents that may be in archives, offices or elsewhere.
  
- (f) Communication now often takes place across multiple platforms. A meeting that might once have led to the preparation of formal minutes, might now be recorded in an attendance note, personal notes (which could be in many different formats) and "side-bar" conversations by email or in a collaboration platform. Within the more than 20-year period covered by the Inquiry's terms of

---

<sup>7</sup> This should not happen when litigation or an Inquiry is in contemplation – measures such as litigation holds should be put into place.

<sup>8</sup> Electronic media (or eMedia) are devices containing data recorded via electrically based processes such as hard drives, random access memory (RAM), read-only memory (ROM), disks (such floppy disks or CDs), flash memory, memory devices (including USB devices), phones, mobile computing devices, networking devices, office equipment, and many other types. See: [Electronic Media - Glossary | CSRC \(nist.gov\)](#)

reference there have been material cultural shifts in how individuals work and the tools they work with.

- (g) The ability provided by technology to generate material by interaction with multiple parties creates huge amounts of full duplication and near duplication. For example, a single document might be emailed to 20 people, amended by several but not all of them and then reattached in different contexts and forwarded to different groups of people. Replication over tens, hundreds or more individuals over a long period in different contexts produces a labyrinthine intermingling of documents and communications. The same documents may also be saved to multiple data repositories (e.g., emails to Mimecast, Exchange or local devices or documents to SharePoint and OneDrive, each of which would be an exact or near duplicate of each other). This is the family document and duplication (or near-duplication) problem. A reviewer or review team will see chains or families that are often many pages or many tens of pages long that look very similar or identical but may or may not be identical or the context of which has subtly (but potentially substantively) changed. Reviewers can manually distinguish between them only with intense effort and focus, which is impracticable where timescales (relative to the volume of documents for review – a mix of relevant and irrelevant documents) are short, particularly as material necessarily falls to different reviewers in those circumstances, and/or the applicable terms of

reference do not direct that focus. We have been working with KPMG to find ways to reduce the impact for the Inquiry in its work of the resulting conundrum and paragraphs 76 and 121 (and the associated appendices) below summarise actions recently taken and proposed as solutions.

- (h) The near duplicates and large families issues are aspects of document review being a manual process that involves the application of human judgement to code documents, for example, as to whether they are responsive to the applicable terms of reference, whether they are subject to legal professional privilege and whether redactions should be applied and, if so, where they should be applied. Between different reviewers there will be divergent, reasonable value assessments. Those valid divergences increase where the number and combinations of the issues being coded are greater. All disclosure exercises will involve the review of documents that are clearly and obviously within scope, but most will also involve instances where multiple reviewers (or even the same reviewer at different times, in part because that reviewer will acquire greater experience with the dataset and issues) could look at the same document and reasonably make different coding decisions. The potential for genuine human error is also unavoidable. Both divergent approaches and human error should be – and have been – reduced by system design, quality checks and proper instruction and supervision but they cannot be eliminated at any stage of a

review exercise. Technological methods, such as email threading, TAR and CAL, can assist but, to date, have not been reliable because of the variability of data quality as noted at paragraph 11(g) of my letter to the Inquiry dated 1 September 2023 [POL00126339]. I have summarised at paragraphs 89 below the work that is ongoing to try to improve that situation.

- (i) Where required search parameters (for example in a Section 21 Notice or Rule 9 Request) are broad, complex in combination and/or concept based it is more difficult for reviewers to assess whether documents respond and how they need to be coded<sup>9</sup>.

32 Whilst the risks cannot be eliminated, BSFf have worked hard with Post Office and KPMG to mitigate risks. For example:

- (a) recruiting appropriately skilled individuals into the teams at all levels;
- (b) close engagement with Post Office subject matter experts and members of its Inquiry team;
- (c) extensive onboarding training and reviewer guidance (which is updated on an ongoing basis in response to feedback,

---

<sup>9</sup> For example, the s21 (03) Notice dated 21 July 2023 is highly complex to scope, review and code because of the combinations and nature of the issues and relationship of those issues to each other and to individuals over a long period. The s21 (08) Notice dated 8 October 2023 is an example of a request dealing with a specific area and period that is less complex to scope, review and code. This is discussed in more detail in paragraphs 99 to 103 below.

correspondence with the Inquiry and events, such as hearings, within the Inquiry);

- (d) multiple tiers of review;
- (e) establishing a tier one and tier two reviewer buddy system and defined escalation routes;
- (f) frequent (normally daily) thematic and issue discussions amongst the disclosure team as a whole with additional such calls for specific workstreams;
- (g) proactive monitoring of any signs of concern (e.g., lower coding accuracy, declining level of reviewer engagement or review rate) that leads to direct, tailored feedback;
- (h) close ongoing engagement at operational and senior level with KPMG as Post Office's e-disclosure provider; and
- (i) quality control checks based on samples of documents and targeted searches.

My understanding based on confirmations from HSF (see, for example, Mr Rowan's 23 August witness statement [WITN09950100]) and P&P is that they have taken similar steps in their review exercises for the Inquiry.

**D. EXCHANGE/OFFICE 365 ISSUE AND PROPOSED SOLUTION  
[NOVEMBER REQUEST PARAGRAPH 1]**

33 Paragraph 1 of the November Request stated:

“Microsoft Exchange 365

*Please set out detail of the issue that has arisen in respect of Microsoft Exchange 365. This should include the following:*

- a. When the issue was discovered, by whom and in what circumstances.*
- b. POL's understanding of the cause of the issue and where you consider responsibility for the issue lies.*
- c. How it is that the issue was not identified at the time that searches were originally carried out over Mimecast and the checks that POL carried out to ensure compliance with its obligations in that regard.*
- d. The way in which the issue is being resolved and the date on which such an exercise is likely to be completed.*
- e. The steps that are being taken to remove documents that are duplicative or duplicative in material respects. If and insofar as steps are not being taken, please explain why.”*

A footnote to sub-paragraph c. stated: *“Please note that the Inquiry is not expecting a detailed analysis of compliance with each and every Rule 9 request or Section 21 Notice. This question is directly aimed at the process by which POL satisfied itself that its original use of Mimecast was sufficiently comprehensive.”*

- 34 Post Office has yet to conclude investigations into the issue that has arisen in relation to data held on Microsoft Exchange Online (part of Microsoft 365



services) (“**Exchange**”) and its impact. However, I set out below to the best of my knowledge:

- (a) an explanation of the substance of the issue based on my current understanding of Post Office’s email systems today and as they have evolved since the 1990s as explained to BSFf by Post Office’s IT team;
- (b) an explanation of how the issue was discovered, by whom and in what circumstances, as well as communications with the Inquiry from 18 August 2023 on it. Where events happened prior to my/BSFf’s first awareness that this was an issue requiring potential investigation during mid-August 2023 onwards and/or involved work that was ongoing in parallel with which we were not directly involved, I base my understanding on documents provided by Post Office and its other legal advisers and discussions that colleagues in the BSFf team have had in order to investigate the sequence of events;
- (c) an explanation of how Post Office is undertaking investigation and technical analysis to process Exchange data in a manner to reduce the number of duplicates for review and ultimately produced to the Inquiry (this is my understanding based on what has been explained to BSFf by Post Office and KPMG); and

(d) an explanation of the current working plan by Post Office (with timescales where available) in respect of the different Inquiry Phases.

35 In drafting this evidence in Sections D-G and the Appendices to my statement, I have had the benefit of engaging with technical experts from Post Office and KPMG and the support of several experienced colleagues who have had further such engagement. I am not a technical expert.

**The Issue: Current email data systems for storing emails sent and received: email data repositories held by Post Office and the role of Exchange**

36 Email is the primary operational communication channel for Post Office both internally and externally and has been for much of the relevant period that is the focus of the Inquiry. Therefore, email is rightly a key category of electronic data. However, email is not a description of any specific data repository. Today, when an email is sent to or from an address on the postoffice.co.uk email domain, there are several potential repositories where that email will or may be stored:

(a) First, and most recognisably, when an email is sent from or to an email client<sup>10</sup> such as Outlook (Post Office's current email client) on a user's device this is stored in a local email data file or mailfile on

---

<sup>10</sup> An email client is the software application that is used, for example, to access, manage and send emails. See: [Mail User Agent \(MUA\) - Glossary | CSRC \(nist.gov\)](#).

the device itself and email data on that local file can be viewed from the email client even when the device is offline ("**local mailfile**").

(b) However, emails are not sent directly from or to an email client. Rather, the email client (i.e. Outlook), connects with a cloud-based mail server that sends out or receives the email. Post Office currently uses Exchange as its mail server. There is a server-level email data file or mailfile ("**Exchange mailfile**") that synchronises with and replicates the local mailfile. Permanent deletions<sup>11</sup> of email data by users at local client level will synchronise and replicate in the server mailfile after 30 days unless a relevant litigation hold has been applied (which would prevent permanent deletion from the Exchange mailfile). By design and because of the application of the litigation holds we are instructed Post Office have put in place, Exchange mailfiles would be a more complete record of emails than local mailfiles so there would be no benefit to harvesting a local mailfile if an Exchange mailfile also exists.

(c) In addition, Post Office utilises a further email gateway platform that records a copy of emails transmitted within Post Office's Exchange server and through which emails between its Exchange server and an external email domain must pass. The current platform used by Post Office is Mimecast. Amongst other email services, Mimecast

---

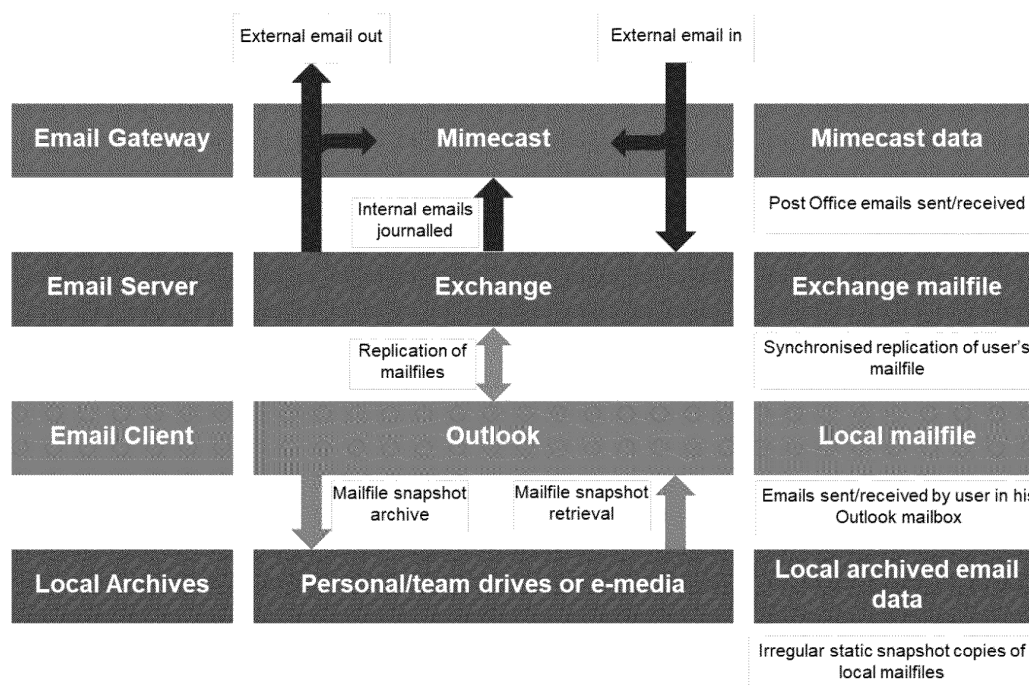
<sup>11</sup> I.e., where an email is deleted from a user's inbox and then from that user's deleted items folder.

services include (in the simplest terms) a repository that keeps a separate, immutable copy of:

- (i) all external emails transmitted between the postoffice.co.uk email domain and any other email domain; and
  - (ii) all internal emails sent between postoffice.co.uk email addresses transmitted within the Exchange server itself but are then uploaded to Mimecast.
- (d) This function (known as “journalling”) by Mimecast creates an archive of email data as it flows into, out of and within Post Office that is separately held on the Mimecast platform. Importantly, Mimecast only journals live email traffic once Mimecast has been activated – it does not journal email data that pre-dates its activation and operation. Such legacy email data would have to be specifically exported from existing sources and imported into Mimecast for ingestion to be included in the Mimecast archive. Together, I refer to the email data that Mimecast captures whilst active and any legacy email that it has ingested in this statement as “**Mimecast data**”.
- (e) Finally, it remains possible for a system administrator or some users (with relevant permissions) to make a copy of their local mailfile at any particular point as a static snapshot which could be separately stored elsewhere either on a local device drive, a network drive or

cloud-based storage such as SharePoint or OneDrive or indeed on physical electronic storage media such as a USB stick, CD or other physical storage media. However, today I understand that this should normally be for temporary or exceptional purposes (e.g., IT fault troubleshooting). In this statement, I refer to static email data of this kind as **“local archived email data”**.

37 The current email system and its related email data repository as I understand it can be illustrated as follows in this (simplified) diagram:



38 Although it may appear from the above that local mailfiles, Exchange mailfiles and Mimecast data are the same, this is not the case and they are not designed to be. Conceptually:

- (a) Email data on local mailfiles and Exchange mailfiles will reflect what is in a user's Outlook mailbox (including items held in their Deleted Items folder and other folders). Absent the imposition of system level litigation holds applied at server level, items that are permanently deleted from a user's mailbox will not be retained on their local mailfile or the Exchange mailfile. I am instructed by Post Office that litigation holds were introduced by Post Office at various points for various purposes and were put in place for certain parts of the business in 2014, 2016 and 2020 in contemplation of various litigation at the relevant times and ultimately Post Office-wide in respect of Exchange mailfiles from March 2021. However, up until these points, emails permanently deleted by users will not appear within their corresponding local mailfiles or Exchange mailfiles (after 30 days in the case of the latter). By comparison, emails that are journalled on Mimecast are immutable and retained until deleted by the system. I understand from Post Office that there is no automatic deletion process set in Mimecast and so the retention period for Mimecast data is in practice indefinite (up to 100 years); and
- (b) Additionally, local mailfiles and Exchange mailfiles will hold data that is not email data at all – most notably calendar and contacts data but also notes and tasklists and system/server messages. They would also contain draft emails and other emails that had not been sent for any reason. By comparison, Mimecast only journals

transmitted email data (albeit that will include transmitted calendar appointments).

39 Consequently, for emails sent or received after the point at which a journalling gateway platform such as Mimecast has been activated, the most complete record of email data should be that data repository that is held on Mimecast. For legacy email data pre-dating the activation of Mimecast which has been ingested into Mimecast, the completeness or otherwise of that aspect of Mimecast data will only be as good as the data record exported to it and as processed for ingestion.

40 I understand from Post Office that Mimecast was activated in or around late 2015. Allowing for transition time, there should therefore be a high degree of confidence that any and all emails sent or received from early 2016 onwards are held on Mimecast. However, out of an abundance of caution, Post Office is undertaking checks and I will update the Inquiry further if those investigations indicate any systemic issues with Mimecast journalling of emails transmitted from 2016.

**The Issue: Legacy pre-2016 email data systems for storing emails sent and received: email data repositories held by Post Office and role of Exchange**

41 Prior to the introduction of Mimecast, however, the relevant history of Post Office's email systems and email data repositories is complicated given the long period of time covered by the Inquiry's terms of reference (see in particular the background contained in Post Office's First Interim

Disclosure Statement [POL00114170ds]). During that time, there have also been generational changes and regular updates to IT systems, applications, devices, software and suppliers. In addition, there has been the separation of Post Office from Royal Mail Group from 2012 onwards with resultant impacts on separation of IT architecture and data. The quantity, scale and more than 20-year timespan of these changes mean that individual instances of data loss should be expected, although they are clearly unhelpful to the task of getting a complete evidence trail.

42 Given the period, restructurings, complexity of IT systems and staff turnover, loss of institutional knowledge has been a key factor in the ability of Post Office to reconstruct its understanding. The current understanding of email systems over the entirety of the relevant period that the Inquiry is investigating has been based on internal investigations at Post Office by consulting subject matter experts and searching available (limited) records but unfortunately cannot be complete. I set out in Appendix 1 Post Office's understanding of the summary position on pre-2016 email data as explained to BSFf by Post Office accompanied by a diagram prepared by Post Office to illustrate its current understanding of how its email systems have evolved.

43 The reconstructed detail of Appendix 1 demonstrates the complexity of Post Office's legacy and current email systems and data repositories but the practical working conclusions below can be made. Based on current understanding of the position with technical investigations yet to complete:



- (a) Post Office user emails sent and received post-1 January 2016 should be captured on Mimecast. Post Office is undertaking technical checks but there should logically be a high level of confidence that Mimecast data captures post-2016 emails as fully as possible. Any discrepancies ought to be exceptional. To be clear, Post Office does also hold email data for this period as mailfiles (local and Exchange) but they would be expected to be fully duplicative of Mimecast data for this period for the reasons set out above. Indeed, Exchange data for this period should hold fewer emails than are held in Mimecast.
- (b) Post Office user emails sent and received between 2012 to 2016 exist on Mimecast. However, there cannot be said to be the same high level of confidence that all emails from this period currently held by Post Office across all its data repositories will be in Mimecast. This is because pre-2016 emails ingested by Mimecast were supplied by Royal Mail Group in late 2015 following the formal separation of the organisations as part of the complex technical project of separating IT systems and data. It was provided as a subset of data from a separate email gateway journalling system, ProofPoint, which was in use by both organisations before Mimecast. Post Office does not have records on, or full visibility as to how, ProofPoint was operated by Royal Mail Group and how the ProofPoint legacy data disks were produced for Post Office by ProofPoint on Royal Mail Group's behalf. The intention of the

ProofPoint transfer, however, was, at its highest level, that any emails in the ProofPoint system where the sender or recipient was a postoffice.co.uk email address would be exported and ingested by Mimecast. Post Office understands that a copy set of these disks has been located within Post Office's archives (although it would require further forensic analysis to confirm if necessary). However, as stated above, it is understood by Post Office that the entirety of the ProofPoint legacy data on them has been ingested into Mimecast so the disks themselves would be a duplicative source.

(c) Mimecast:

- (i) would (generally) not contain emails pre-dating 2012 as ProofPoint and the email journaling at gateway level that came with it were first introduced by Royal Mail Group in 2011/12 but would include such emails if part of an email chain or as an attachment associated with a post-2012 email.
- (ii) would not contain any 2012-2016 emails that may have been deleted from the ProofPoint archive prior to production of the ProofPoint legacy data disks. Post Office does not have knowledge of how Royal Mail Group operated ProofPoint but it has no current reason to believe that Royal Mail Group applied deletion periods that would materially affect the record during this period); and

(iii) would only contain email data associated with postoffice.co.uk email addresses. Certain Post Office staff may, depending on their function, historically have had emails associated with other email domains such as royalmail.com and those would not have been ingested into Mimecast unless one of the other parties to such emails also had a postoffice.co.uk email address (in which case, they would be captured from the other party).

44 During this 2012-2016 period therefore:

- (a) There would possibly be an additional amount of email data that may exist on Post Office mailfiles (local or Exchange) that would not be on Mimecast. However, the extent of any difference for emails sent and received by Post Office users during this period between Exchange mailfiles and Mimecast depends on the content and quality of the ProofPoint legacy data ingested into Mimecast in respect of that individual. A technically complicated de-duplication exercise between Exchange and Mimecast would need to be designed and undertaken as logically there will be a very large amount of duplication between those two datasets during this period and the current position on that analysis is set out below.
- (b) There may also be email data in local archived email data sources that would not be on Mimecast and/or Exchange. However, as described above, these would not routinely have been created by

users during this period and would not been encouraged or permitted during this period so far as storing them on local drives or physical storage devices and media goes with the introduction and adoption of SharePoint, OneDrive and similar cloud-based storage and/or network drives. SharePoint, OneDrive and similar cloud-based storage as well as network drives are known data repositories and searches for mailfiles can be and have been searched for to provide email data where appropriate. So far as local archived email data stored on physical devices and media go, if and when found, Post Office has assessed the possibility that they may contain non-duplicative responsive material. Such physical media devices and storage include, for example, certain individual USB sticks and laptop folders located for individual custodians likely to be of relevance to the Inquiry and the back-up tapes which HSF has investigated and reported regularly to the Inquiry on previously all of which were assessed and, as necessary, harvested and reviewed.

- 45 Post Office user emails sent and received pre-2012 are not generally found on Mimecast at all (except for limited users involved in piloting ProofPoint before its implementation at Royal Mail Group) and save where part of an email chain or as an attachment associated with a post-2012 email (see the letter from Post Office to the Inquiry dated 10 September 2021 [WITN10810103]). Post Office understands therefore that there is particular interest in email data from other sources during this period in

respect of relevant custodians active at that time. Such email data may be held by Post Office:

- (a) on mailfiles (local or Exchange) to the extent that they have been retained and not permanently deleted or lost in data migrations during upgrades or replacements or other IT issues. Given the time that will have elapsed since 2012 until relevant litigation holds were first applied at server level (variously from 2014 onwards), logically it is not expected that there would be particularly material amounts of email data dating before 2012 that still remain in user Exchange data, however, that is still being investigated by Post Office; and/or
- (b) on local archived email data sources that may still exist and be found by or provided to Post Office from time to time. Again, where those old archives have been migrated over the years to current SharePoint, OneDrive and similar cloud-based storage storage as well as network drives, these are known data repositories and searches for mailfiles can be and have been searched for to provide email data where appropriate. And again, so far as local archived email data stored on physical devices and media go, if and when found, Post Office has assessed the possibility that they may contain non-duplicative responsive material with a focus on pre-2012 email data.

### **The Issue: Summary and Responsibility for the Issue**

46 Taking the explanations above in respect of email repositories, it currently appears that where Post Office has not harvested Exchange mailfile data that it holds then:

- (a) There should (logically as I understand the position) be a minimal risk that Post Office has not harvested fully relevant email data it holds for relevant individuals which were sent and received post-2016.
- (b) There is a limited but not immaterial risk that Post Office has not harvested fully the email data it holds for relevant individuals which were sent and received post-2012 up to 2016. This risk will vary by individual (including their length of service and the extent to which they have kept emails in their Outlook mailbox) and depending on whether any data from this period has already been harvested from identified local archived email data sources. There is likely to be extensive duplication between any Mimecast data, Exchange mailfiles and any local archived email data during this period. Careful de-duplication (against all Mimecast data collected) will be (and has already been) required to understand the extent of non-duplicative material and reduce the amount of duplicative data for review and that is ultimately produced to the Inquiry. De-duplication will be complex (and I understand already has been in relation to Inquiry Phase 4).

- (c) There is a risk that Post Office has not harvested fully the email data it holds for relevant individuals for emails sent and received pre-2012. This risk will vary by individual (including their length of service and the extent to which they have kept emails in their Outlook mailbox) and depending on whether any data from this period has already been harvested from identified local archived email data sources. De-duplication against existing email data sources will also be required but where local archived email data sources have not already been identified and harvested for individuals, in principle harvesting any Exchange data from this period would not be duplicative. However, as I have said above, it is not anticipated that there would be material amounts of email data still held on Exchange from prior to 2012, although that is still subject to further investigations by Post Office.

47 In relation to responsibility for the issue:

- (a) In terms of the responsibility (duty) to address the issue, this rests with Post Office. In terms of delivery, that, operationally, will need to be by advisers by phase.
- (b) In terms of the cause and reason for the issue arising I have set out above the underlying technical reasons for the difference between Mimecast and Exchange and below the related factual sequence of understanding as it developed over the period since 2017 (as I currently understand them). These are events with which I was not

involved and the sequence has been put together from the documents. Save as specified in this statement, I have not spoken to any of the individuals involved. It would not therefore be right for me to comment further beyond the facts as I understand them.

**When was the issue discovered, by whom and in what circumstances (and what has previously been communicated to the Inquiry)?**

48 I have covered in this section the period from 10 July 2023 to the discussion of the issue with the Inquiry on 3 November 2023. I have also set out the position before that date in responding to the request as to how it was that the issue as not identified at the time that previous searches were carried out.

49 There were 2 specific instances of which I am aware where a potentially relevant email was identified as existing, but which could not be found in Mimecast data collected and that triggered recent proactive investigation of the issue:

- (a) A series of documents had been provided by Post Office on 19 May 2023 as part of a FOIA request in May 2023, but the full suite had not been provided to the Inquiry in response to any prior Rule 9 request. Those documents were then produced on 30 May 2023 to the Inquiry. These documents are referred to in the witness statements of Ben Foat dated 21 June 2023 (prepared for the disclosure hearing on 4 July 2023) [POL00118164ds] and Gregg Rowan dated 23 August 2023 (prepared for the disclosure hearing



on 5 September 2023; see paragraphs 53 and 54) [WITN09950100]. The FOIA request response produced documents that were not found in any Mimecast data. I understand from HSF that the documents were, however, identified by Andrew Wise who, while searching through his Outlook email client, located an email (with attachments) that could not be found in Mimecast. That led to an initial query by HSF of Post Office on 10 July 2023 as to how Andrew Wise had located the email in question which, over time, developed into a query as to whether there was a repository separate to Mimecast.

- (b) Separately, BSFf identified on 14 August 2023, during document review in response to the s21(03) Notice (dated 21 July 2023), an email chain comprising 1) an originating email with attachment and 2) a response to the originating email re-attaching the attachment (produced to the Inquiry as [POL-BSFF-0136285] and [POL-BSFF-0136286]). The email chain was from a collection of email data from Mimecast searched as part of Post Office's response to the s21(03) notice. The BSFf reviewer sought to identify the originating email but was unable to in the Mimecast data extracted at the time. BSFf escalated this to KPMG on the same day.

50 Enquiries and investigations took place during July and August, initially between HSF and Post Office to establish whether these were exceptional

anomalies, or whether there was a potentially greater issue. KPMG became involved also from mid-August.

51 HSF wrote to the Inquiry in its letter dated 18 August 2023 [POL00124516], informing it of the Andrew Wise email issue and investigations into whether there were potential further repositories:

*“The document referred to in paragraph 44(b) of Ben Foat's Second Witness Statement (an email dated 31 August 2011 sent by Andrew Wise) was not contained in the CCRC workspace:*

*- As explained in previous correspondence with the Inquiry (including our letters dated 12 August 2022 and 30 June 2023), as part of the Royal Mail Group/POL separation, there was a wholesale change to the email servers that POL used. Most of the archived email data pre-dating 2012 was retained by RMG and now no longer exists. The earliest email available to be harvested from Andrew Wise's mailbox is dated December 2011 (i.e., after the date of the email referred to in paragraph 44(b)).*

*- POL are continuing to investigate where the document was saved and whether further repositories may need to be harvested.*

*We note that the document referred to in paragraph 44(a) is dated 23 May 2011 and is an earlier email in the same chain as the email referred to in paragraph 44(b). For the avoidance of doubt, this email was not collected*

*from Andrew Wise's mailbox. It was recovered from the deleted items of another custodian (Robert Daily)."*

52 Enquiries of and technical investigation by Post Office's IT team continued during the remainder of August and during September in parallel (from BSFf's perspective) with the intensive work taking place at that time, in particular, in relation to the 5 September disclosure hearing, remediation in relation to the three specific disclosure issues to which that hearing related and the work on the response to Section 21 Notice (03).

53 On 6 October 2023, BSFf wrote to the Inquiry [WITN10810105] in connection with s21 (03), including an update that:

*"A review of possible additional data sources, which includes the collection and review of instant messages and review of some hard copy documents to consider if they are responsive to the Notice and/or Terms of Reference. For example, as previously notified to the Inquiry in correspondence dated 18 August 2023, BSFf together with POL, HSF and KPMG are investigating a Microsoft Exchange repository that may contain emails covering the period before POL started using Mimecast (2016). From initial investigations, it appears that these emails primarily span 2011 to 2016 although there are some outliers at either end. This repository is called a "mailfile" and it may contain emails that are not held within Mimecast. Whether or not the emails are 'new' requires complex technical knowledge and work and there is not yet a clear answer as to the extent of duplication between the Microsoft Exchange repository and Mimecast. That work is*

*being progressed by POL as quickly as possible and POL will provide an update to the Inquiry as soon as it is in a position to do so."*

- 54 On 16 October 2023, BSFf wrote to the Inquiry [WITN10810102] in relation to a range of disclosure issues (this letter is a subject of the October Request) and requested a meeting with the Inquiry to discuss several issues including disclosure issues to best support the Inquiry. BSFf proposed in an email the following agenda in relation to disclosure issues (amongst other matters):

*"Please find attached letter for your kind attention. As the Inquiry will be aware, POL has requested a meeting with the Inquiry to discuss a number of areas where it would be helpful to meet in person. The purpose of the meeting is to:*

- 1. To assist with the Inquiry's visibility over work being conducted by POL;*
- 2. To understand the Inquiry's direction so that POL can plan its work to assist the Inquiry;*
- 3. To sight the Inquiry on specific disclosure challenges faced by POL and discussion about ways to best align with the Inquiry's timescales/critical path.*

*We would propose the agenda for the meeting to be as follows and would be grateful for the Inquiry's comment and input on the same:*

- 1. Approach to disclosure:*

- a. *Harvesting and searches of additional repositories identified in response to s.21 and structural review;*
- b. *Mimecast vs Exchange server;*
- c. *Third party material;*
- d. *Inquiry's expectations of the document review process (search terms, level of reviewers etc);*
- e. *Prioritisation of disclosure by POL (subject to information provided by the Inquiry);*
- f. *Cut off date to apply to disclosure searches (noting CLI footnote 3)."*

55 On 20 October, HSF wrote to the Inquiry [WITN10810106] with a further update on disclosure issues and at paragraphs 54 to 59 explained that:

*"Email repositories*

*54. Further to our letters dated 18 August, 4 September and 6 October 2023, with support from KPMG, POL has continued to investigate the extent to which certain emails that are not available on Mimecast might be held on other repositories and may need to be harvested.*

*55. We understand that this exercise has been time-consuming, complicated and is ongoing. Whilst there continue to be significant uncertainties, POL's preliminary understanding continues to be that, in addition to POL's Mimecast archiving system, some custodian email data*

*is also held in Exchange (or Office) 365, and that (where available) custodian "mailfiles" can be harvested from Exchange 365.*

*56. The work undertaken so far by KPMG indicates that there are instances where emails are not on the Mimecast archive and are available in Exchange (and, inversely, that some emails available in the Mimecast archive are not available in Exchange). As yet, POL has not been able to identify why this occurs. POL is still in the process of investigating the issue and its implications and will write to the Inquiry with a substantive update when more is known.*

*57. In the meantime, POL has extracted Exchange 365 mailfiles where available for certain custodians, including (on an urgent basis) certain Phase 4 witnesses who are due to give evidence in the coming weeks and potential Phase 5 witnesses. KPMG have been seeking to interrogate data relating to some of these custodians and have tested deduplication workflows in order to try to understand the extent of duplication / new material.*

*58. Whilst this work is ongoing, we understand from KPMG that initial sampling indicates that there is significant overlap between the data from the two sources, but also that the mailfiles on Exchange 365 do contain additional documents. Furthermore, md5# deduplication has had limited effect in respect of this dataset. KPMG are still testing alternative solutions, including custom deduplication, but based on the work conducted so far, it appears likely that isolating potentially new and relevant documents from*

*mailfiles on Exchange 365 will be a complicated and possibly manual process which (for technical reasons) might nevertheless leave a volume of duplicative material for review.*

*59. POL will keep the Inquiry updated on a regular basis as the investigation of this data progresses.”*

56 Further to BSFf's letter to the Inquiry dated 16 October 2023 [WITN10810102], a meeting was subsequently scheduled with the Inquiry for 3 November 2023 and the Inquiry reverted with the agenda on 1 November 2023 [WITN10810107]. The Exchange issue was not specifically included but anticipating that it would be discussed under disclosure issues, I wrote to set out our (and my) understanding of the issue as it then stood on 2 November [POL00165906].

**Why was the issue not identified at the time searches were originally carried out and what checks were carried out by Post Office?**

57 I have summarised in this section my understanding of the position from the exchanges that I have seen primarily relating to 2 periods:

- (a) Relating to the period in 2017 when Post Office was involved in the GLO proceedings.
- (b) Relating to the Inquiry from February 2021 onwards.

58 As already noted above, Post Office will have been advised by a series of external legal advisers throughout these periods. The Inquiry will recall

that Post Office has given a limited waiver in respect of certain privileged documents up to February 2020, which would not cover much of the privileged material between Post Office and its legal advisers on this aspect. However, I have sought below to set out the sequence without the need to refer to privileged material.

- 59 The earliest relevant statement relating to Post Office's consideration of its own data repositories and resulting position (including on emails) is set out in its Electronic Disclosure Questionnaire (GLO EDQ) dated 6 December 2017 for the GLO [POL00000657]. It was prepared and signed on Post Office's behalf by Womble Bond Dickinson LLP with input from Post Office. Within that GLO EDQ , it was stated variously in response to Question 3 of Part 1 that:

*“Until c.2012, Post Office employees used Lotus Notes. Microsoft Exchange was introduced on the separation of Royal Mail and Post Office and when introduced, Post Office's employees' emails which were stored in Lotus Notes were transferred into Microsoft Exchange.”*

*“A backup copy of the Lotus Notes database was taken as part of the migration exercise and it may be possible, though not straight forward, to identify and export data from this backup. Post Office does not believe it would be necessary to access this copy due to the transfer of data into Microsoft Exchange.*



*When Microsoft Exchange was introduced Post Office also introduced email archiving – initially by Proofpoint and from February / March 2016 onwards by Mimecast.*

*The emails stored in Proofpoint were transferred into Mimecast. These archives store all emails sent to or from a Post Office employee and emails cannot be removed from the archive (unless special permissions are granted to do so). It is understood that this archive will hold emails dating back to 2012 including for Post Office employees who no longer work for the business.*

- (a) In response to Question 3 of Part 1 of the GLO EDQ also went on to note that Skype Instant Messages “*would be held in each Custodian’s “Conversation Folder” in Microsoft Exchange. There is no archiving.*”

In response to Question 13 of Part 1 of the GLO EDQ added that: “*When an employee ceases to be an employee of Post Office, their laptops are re-distributed within the business. Their emails would remain stored in Mimecast (as explained further in Question 3) and documents stored in SharePoint” Appendix B further noted against “Lotus Notes and Microsoft Exchange” that:*

*“Microsoft Exchange is Post Office’s principal email software used by all employees. Microsoft Exchange was introduced by Post Office in c.2012.*

*Prior to Microsoft Exchange, Post Office employees within Royal Mail used Lotus Notes.*

*Please see Question 3 for an explanation of archiving.*

*It is anticipated that an average user can be expected to send and receive 31,000 emails a year. Extracting the accounts of all the Key Custodians for one year would therefore capture around 2,511,000 emails (plus attachments)."*

60 Based on Post Office's current understanding of its email systems and repositories (as summarised in this statement), unfortunately, it appears that these descriptions in the GLO EDQ in hindsight were not accurate or were over-simplified:

- (a) Although Exchange is mentioned in the context of emails and instant messages, it is not identified as a separate server-level source of email data. References to Exchange in the Appendix B to the GLO EDQ appear to equate it to "*email software used by all employees*" which would describe the Outlook email client;
- (b) The statement that Post Office used Lotus Notes until 2012 is understood to be incorrect, current understanding is that Post Office stopped using the Lotus Notes email client and Lotus Domino servers and started using the Microsoft Outlook email client and Microsoft BPOS-D servers over the period from 2008 to 2010. For completeness the statement is also incorrect as before Lotus Notes, it is currently understood that Post Office used versions of MSMail;

- (c) Consequently, the suggestion that Post Office introduced ProofPoint email archiving at gateway level in 2012 *at the same time* that it moved to Microsoft Exchange and Outlook also does not match the currently understood timeline;
- (d) The assertion that all Lotus Notes data would have transferred to Exchange in 2012 is also understood now to be incorrect since, as well as the period of migration to Exchange occurring between 2008-2010, Post Office's current understanding is that not all old email archives would have been migrated. Only those files associated with active users at the time would have been migrated to Exchange at the time (if at all). It follows that the indication that Lotus Notes archived data would be duplicative of Exchange is therefore also not (always/fully) correct; and
- (e) Finally, although implied at most, any reading of the GLO EDQ, as suggesting that Mimecast (and before it ProofPoint) is a complete repository of:
  - (i) Outlook emails either in whole or in part from 2012 onwards;  
or
  - (ii) Lotus Notes emails imported into Exchange or any other emails pre-dating 2012.

would not be correct based on current understanding.

- 61 The specific reasons for these issues in the GLO EDQ are not clear or known to me at this time and it would require much more investigation (likely going beyond documentary review) to pinpoint the specific cause or causes of the issues with the GLO EDQ identified above. However, I reiterate again the complexity around the technical and legacy issues in this area, lack of institutional memory over the lengthy timeframe and that Post Office has had to (re)build its knowledge and understanding in this area. It appears however from the circumstances at that time that that has resulted in a number of areas of lack of precision in use of terminology, understanding and possibly communication between different disciplines (in particular Legal and IT).
- 62 Following the GLO EDQ and having harvested Mimecast (incorporating ProofPoint data), I understand from Post Office that gaps in pre-2012 email correspondence were identified in custodian emails at various points and local archived email data was searched for and where it was located it was added to Relativity workspaces. Consequently, awareness increased between Post Office and its advisers that email data repositories pre-2012 could be contrasted with email data repositories post-2012. Post Office took steps to identify and provide such material from local archived email data where relevant and, in particular, it found certain of the snapshot repositories on SharePoint and OneDrive as well as local storage on laptop devices. I understand from Post Office that Exchange data was not identified as a separate data source for harvesting and was not harvested for the purposes of the GLO.

63 On 19 August 2020, P&P produced the Disclosure Management Document (plus Annex) [POL0042261] [POL00039560] in the context of the criminal convictions appeals and Post Conviction Disclosure Exercise (PCDE) (the PCDE DMD) with inputs from Post Office. The data gathered for the GLO formed a part of the proposed disclosure for that process and to that extent at least there was a degree of reliance on the underlying methodologies adopted previously in respect of that exercise. I understand however from P&P that the PCDE DMD was also informed by P&P's own enquiries of Post Office's IT team specifically regarding email data (particularly pre-2012) as this had not been explored in any detail in the GLO. An Addendum and an Annex were produced by P&P in the PCDE [POL00142414] [WITN10810108] dated 13 January 2021 (references in the documents to '13 January 2020' are typographical errors). The 13 January Addendum referred to email review for the PCDE but specifically in the context of Mimecast data, however, the 13 January Annex noted against email repositories for Post Office:

(a) In respect of Post Office emails post-2012:

*"Post-2012 email data (Mimecast)*

*NB. Although described in the DMD spreadsheet as "Post 2012 e-mail data (Mimecast)", in fact the Mimecast data dates from December 2011 onwards."*

(b) In respect of Post Office emails pre-2012:

*“(i) During the RMG and POL separation, there was a change in the email servers and software used by all employees (from Lotus Notes to Microsoft Exchange). Only the email data of existing POL employees (i.e. those employed at the time of separation and who continued to be employed by POL thereafter) was transferred across to POL’s new servers.*

*Electronic Filing Cabinet (EFC), which contains pre-2012 email data and the Lotus Notes back up data that had been provided to WBD as part of the GLO (not including legal/security), was uploaded to the GLO dataroom and has been digitally searched by P&P for the case-specific and GDR. Analysis of the results of the EFC searches reveals very little relevant material related to the Legal/Security teams. All relevant material has been extracted and reviewed.*

*In relation to e-mail data of existing employees for whom pre-2012 data seems to exist, P&P’s initial review has identified 7 priority custodians and 15 non-priority custodians. The data for the priority custodians has been located (December) and extracted for search & review (currently ongoing).”*

- 64 Further, on 19 December 2022, P&P produced a Second Addendum [WITN10810109] updating on additional repositories located up to that period including data sources such as the devices and storage tapes from Chesterfield, which have been reviewed and formed part of previous updates from HSF to the Inquiry.

65 The PCDE DMD, Addenda and Annexes reflect Post Office's developed understanding at the time (as it stood) and since the GLO EDQ that Mimecast in fact contained emails after 2012 but not before 2012. The explanation captured in the 13 January Annex [WITN10810108] in particular in respect of pre-2012 emails does unfortunately (in hindsight) however continue to reflect some of the looser use of terminology adopted in the GLO EDQ. In respect of the date that Post Office stopped using Lotus Notes, it also continues to state incorrectly that it was 2012. Exchange was, as previously, not itself identified to be a separate available data repository for email data.

66 Subsequently, the Inquiry is aware of the contents of the four Interim Disclosure Statements [POL00114170ds] [POL00114173ds] [POL00114176ds] [POL00114177ds]. The First Interim Disclosure Statement dated 27 May 2022 [POL00114170ds] is of particular relevance as it describes many of the challenges experienced by Post Office in its disclosure that I have also touched on in this statement. As with each of the Disclosure Statements, in the usual way, and necessarily, it was based on the signatory's understanding of the position as reported to them. In respect of email repositories, that statement explains the current understanding at the time that:

*"19. Prior to 2012, I understand that POL's provider of email servers and software was Lotus Notes. Following the Separation, POL began to use Microsoft Exchange instead of Lotus Notes. At the same time, POL began*

*to use an email archiving system called Proofpoint. Since the beginning of 2016 POL has used Mimecast as its email archiving system. The emails that had previously been stored in Proofpoint were transferred into Mimecast.*

*20, I understand that there are a number of limitations to the email data that POL possesses, including:*

*a. Only those who were identified as being current POL employees at the time of the Separation (i.e. those employed by POL and who continued to be employed by POL thereafter) were transferred across to POL. Accordingly, POL does not hold copies of email data in respect of those employees who left the business prior to or at the time of the Separation.*

*b. At the time of the Proofpoint/Mimecast migration, only emails sent to or from a postoffice.co.uk email account were migrated, despite POL employees having access to and being able to use royalmail.com email accounts. The consequence of this is that POL did not receive emails solely between royalmail.com email accounts, even if those emails involved POL employees. Furthermore, the migration from Proofpoint to Mimecast will not have captured any deleted email data.”*

67 I note that the understanding of the position recounted in the First Interim Disclosure Statement [POL00114170ds] is a further evolution of Post Office’s understanding of Mimecast and its limitations. Again, this reflects how Post Office was continuing to build its understanding of these systems



throughout. However, once again, in hindsight it is unfortunate that the chronology for Post Office moving from Lotus Notes to Exchange and coinciding with email archiving with ProofPoint no longer accords with the understanding of Post Office as I have set out in this statement. The matter of pre-2012 emails is not specifically addressed in the First Interim Disclosure Statement save to note that legacy “E-filing Cabinets” as part of Lotus Notes had formed part of GLO repository searches. However, I note that data repositories where local archived email data (such as old Lotus Notes .nsf files) are known now to be found were referenced in that statement as known repositories of data such as SharePoint and other team drives, file servers, the NAS Drive and laptops. However, Exchange data is not itself identified as a separate data repository for emails.

68 For completeness:

- (a) Whilst the Second Interim Disclosure Statement dated 18 October 2022 [POL00114173ds] deals primarily with what it calls “*hard copy documents*” it can be seen (e.g., from paragraph 17) that this includes references to physical data storage devices and eMedia such as CDs and tapes which are a potential repository for local archived email data;
- (b) Whilst the Third Interim Disclosure Statement dated 30 November 2022 [POL00114176ds] deals primarily and further with “*hard copy documents*” again this includes eMedia such as floppy disks (e.g.,

paragraphs 44 and 66). Section G specifically addressed a question from the Inquiry in relation to emails post-2000:

*“Question No. 2(d) of Inquiry's 10 November Letter*

*The Inquiry assumes that, like many businesses, POL may have relied more heavily on paper-based communication in the period 1995-2000 (letters, faxes, etc), with an increased reliance on electronic communication (emails, etc) thereafter. Would you please address this assumption and, if it applies in relation to POL and its predecessors, explain if this issue was considered by POL as part of the broader ‘approach adopted to ensure reasonable steps taken to search potentially relevant hard copy locations (Q1/Q2 2022 to present)’ (Second Interim Disclosure Statement, section F)? If the issue was considered, please explain how it was incorporated into the approach.*

*For the reasons explained above, it is not the case that POL actively made this assumption. POL is not in a position to confirm definitively that it would have relied more heavily on paper-based communications in the period 1995 – 2000, with an increased reliance on electronic communications thereafter. I understand from a current POL employee who worked in the security team in the mid-1990s that there was a considerable amount of paper-based communication but also that electronic communications (i.e. email) were in use around that time and floppy disks were also used to*

*transfer material. For context, I understand that HSF has conducted searches for email data held in POL's Relativity databases in the period from 1995 to 2000 (including across the entire GLO and Inquiry databases, as well as the mailbox data of 124 custodians harvested for the purposes of responding to requests received from the Inquiry) and has only identified 63 native emails from this period.*

*As noted above, POL undertook (and continues to undertake) searches for material responsive to the Inquiry's requests with regard to its electronic databases (which, as noted at paragraph 31 above, already contained material which had been harvested from hard copy document repositories for the purposes of the GLO and the PCDE) and, where it was considered unlikely that responsive documents may be contained on those electronic databases, its hard copy document repositories”*

- (c) Whilst the Fourth Interim Disclosure Statement dated 12 January 2023 [POL00114177ds] deals primarily and further with “*hard copy documents*” again this includes eMedia such as CDs (e.g., rows 2 and 3 of the attached table).

69 Taking the Interim Disclosure Statements overall, they present further detail and insight into Post Office's data universe and understanding, in particular, around Mimecast. However, with hindsight, it remained the case then and up until very recently that it reflected Post Office's understanding at the time that Lotus Notes was used up to 2012 coinciding with the

introduction of ProofPoint and email archiving. The statements also do not identify Exchange data as a separate available repository of email data (in conjunction with other known and identified repositories such as SharePoint, OneDrive, network drives and physical storage devices and media). Whilst work undertaken for the GLO, PCDE and prior disclosure exercises will have been taken into account and built upon, we understand that Post Office did also conduct extensive review activity with its advisers to support the development of the Interim Disclosure Statements.

**How is the issue is being resolved and when is it likely to be completed?**

70 There are a variety of factors concerning Exchange of which we are aware that affect how it is interrogated:

- (a) Exchange is not a complete record of all emails sent/received.
- (b) Exchange holds significant volumes of data. This reflects working practices with electronic data. For example, the Exchange data items for custodians named in respect of 1 request made for Inquiry Phase 5 (including some who are still employed by Post Office) are, in many cases, in the low millions each (equating to 1-2TB of data each). Further, analysis of that data by KPMG needs to be performed on KPMG's systems. This is because Exchange has limited analysis functionality, whereas KPMG has access to systems with more precise analysis functionality so KPMG can manage that analysis (KPMG do not have direct access to Post

Office's systems for information security reasons). Consequently, the time taken for analysis reflects machine time required for identification and migration of potentially large volumes of data from Post Office's systems to KPMG and for analysis on KPMG's systems (assuming there are no issues with the data transfer).

71 HSF and P&P are instructed with respect to Phase 4 (see paragraph 17 above for further details). My understanding of their plans is derived from discussions with individuals from those firms and is set out in the following paragraphs. I understand that their approaches are to an extent driven by the fact of upcoming witness hearings and the need to take urgent efforts to check whether additional documents need to be disclosed to the Inquiry (and witnesses) for each such hearing.

72 For HSF (the information in this paragraph has been provided to me by HSF):

(a) HSF is assisting Post Office with disclosure of emails harvested from Exchange and I understand HSF have undertaken the following review exercises on an expedited basis because of anticipated hearing dates in November and December 2023:

(i) In relation to Catherine Oglesby, approximately 31,887 documents were collected from Exchange via party / participant-based searches using a combination of known email addresses and wildcard terms. Following application of

keywords, approximately 7,469 documents were reviewed (on a full family basis) and on 10 November 2023 32 documents were produced to the Inquiry.

- (ii) In relation to Gareth Jenkins, approximately 8,744 documents were collected from Exchange via party / participant-based searches using a combination of known email addresses and wildcard terms. All 8,744 documents were reviewed in full and, on 10 November 2023, 3,045 documents (comprised of 2,134 parent emails and 911 attachments) were produced to the Inquiry together. Noting the breadth of Question 2 of Request No. 30 (which seeks all emails between POL employees and Mr Jenkins), on 17 November 2023, an index was provided to the Inquiry which identified documents that might be of greater interest than others in the production (subject, of course, to the Inquiry's own views).
- (iii) P&P has had primary carriage of reviews of Exchange documents relating to Paul Whitaker. However, HSF assisted with the review of approximately 3,245 documents located via search terms for the Castleton case study – on 16 November 2023, 2 documents were produced to the Inquiry.
- (iv) In relation to Elaine Cottam, attempts were made to locate emails via a combination of address book searches and party

/ participant-based searches using a combination of known email addresses and wildcard terms – no potentially relevant emails were found.

- (b) In addition, steps have been taken to harvest and deduplicate Exchange documents relating to Andrew Winn and Andy Dunks, who Post Office understands may be scheduled to give evidence in January 2024.
- (c) Whilst the workstreams detailed above are those which were put in place to deal with (at the time) forthcoming witnesses, HSF have been giving thought to how Post Office might deal with the issues arising from the discovery of the data contained on Exchange which affects the civil elements of phase 4 of the Inquiry more broadly. Post Office will update the Inquiry further.

73 For P&P, (the information in this paragraph has been provided to me by P&P):

- (a) P&P has been and is assisting Post Office to search for, review and produce material from Exchange relating to witnesses giving evidence in respect of the criminal case studies (“CCS”) Module of Inquiry Phase 4. POL, P&P and KPMG have been working to search, review and produce material on a witness-by-witness basis, bearing in mind the Inquiry’s hearings schedule, to ensure that as

many of the scheduled Inquiry Phase 4 hearings as possible can go ahead as planned.

- (b) In summary:
  - (i) What documents P&P reviewed and whether P&P used search terms was dependent upon the number of documents returned. Post Office has updated and will continue to update the Inquiry in correspondence.
  - (ii) Data in respect of the witnesses who have remained scheduled to give evidence up to and including Friday 8 December 2023 has been produced to the Inquiry.
  - (iii) Data relating to the witnesses scheduled to give evidence in the week beginning 11 December 2023 has been identified. It will then be reviewed and produced as soon as possible before their respective hearings.
  - (iv) P&P is proceeding on the basis that the Inquiry will call Graham Ward (in addition to other witnesses) in January 2024 and Post Office and KPMG will harvest data for review accordingly.
- (c) P&P's initial view is that any retrospective remediation work with respect to the criminal case aspects of Phase 4 should broadly reflect the approach taken in remediation exercises previously



undertaken by P&P relating to (a) policies and procedures and (b) training, experience and qualifications.

(d) The approach is under consideration and anticipated to identify what (if any) remediation work is required for the following: CCS witnesses; Rule 9 Requests 6 and 14; and the workstreams relating to policies, procedures, training, experience and qualifications. This will be informed by Post Office's understanding of whether/to what extent the Inquiry wishes to further examine these issues, the number of potentially relevant custodians and anticipated timing and likelihood of finding relevant documents. Post Office will update the Inquiry separately also in correspondence in this regard. The approach will also be informed by the Post Office's ongoing PCDE. As part of the PCDE, Exchange data in respect of all potential future appellants will be interrogated and any material that is identified as being responsive to any criminal related Rule 9 or Section 21 Notice, CLI 49 or the Inquiry's terms of reference and Completed List of Issues, will be produced to the Inquiry.

(e) Post Office will update the Inquiry separately also in correspondence in this regard.

74 We (BSFf) have considered – from our (still building and therefore far from complete) knowledge – potential impacts on Inquiry Phases 2 and 3:

- (a) Phase 2 concerns “*Horizon IT System: procurement, design, pilot, roll out and modifications*”. This relates to events before 2012. As explained above, Exchange is understood to hold data not available in Mimecast or other Post Office systems potentially from before 2012 (introduction of Proofpoint) and 2016 (introduction of Mimecast), depending on the email custodian. The working conclusion is that further investigative steps about potential impact on Inquiry Phase 2 of the Exchange issue would logically not be productive, taking into account, amongst other things, the low likelihood that individuals would have been communicating by email in the period up to 2016 about procurement and roll out processes that completed more than a decade earlier and retained those emails in their Exchange emails. Post Office will conduct further work to validate that working conclusion but I expect that (at most) there will be immaterial numbers of non-duplicative documents on Exchange that are likely to have material probative value to Inquiry Phase 2. Post Office would however welcome engagement with the Inquiry on this working conclusion based on the analysis in this statement of the Exchange issue and the ongoing validation work I refer to in this paragraph.
- (b) Inquiry Phase 3 concerns “*Operation: training, assistance, resolution of disputes, knowledge and rectification of errors in the system*”. Relevant issues and documents may appear before or during the 2012-2016 period. BSF’s current analysis is that

(building on the explanation above regarding Inquiry Phase 2) the approach should be to consider for each witness who gave evidence for Inquiry Phase 3, the dates of their employment at Post Office to identify whether/to what extent they may have sent or received emails up to 2016. For those identified, a proposed prioritisation for search and review of data for those individuals (for example by likely relevance) would then be shared with the Inquiry for comment.

75 Regarding Inquiry Phase 5, BSFf have searched email data with respect to related Inquiry requests. 1 Inquiry request to date (s21 (03)) has identified named custodians and consequently required the specific collection, search and review of Post Office specific custodian email data. As explained in correspondence to the Inquiry, for the custodians identified in that request I understand that:

- (a) Post Office has undertaken Address Book searches for each custodian named. For all custodians for whom we have been told by Post Office that Exchange data is available, pre-2016 data has been migrated to KPMG, searched using the applicable search terms and reviewed. Documents responsive to the relevant request were produced to the Inquiry on 30 November 2023.
- (b) POL has undertaken, and is undertaking, participant-based searches for each of those named custodians. For those whose responsive data was of a manageable size, the data was migrated to KPMG, searched using the applicable search terms and pre-2016

data is currently being reviewed. For those custodians whose responsive data was so voluminous that it was not feasible to transfer to KPMG, Post Office is seeking to re-run the searches with a cut-off date of 2016 to reduce the amount of data to transfer. Once that has been transferred, it will be searched and reviewed. We will update the Inquiry as soon as possible regarding anticipated timeframes for that review and production.

- (c) For avoidance of any doubt, I understand that each of the above (searches and review based on Address Book and participant-based searches) have applied global de-duplication using what I understand is the standard forensic processing MD5 Hash approach only.
- (d) Post Office is considering further (but is not currently using wildcard searches of Exchange data in respect of the request referred to above (or Inquiry Phase 5 requests to date because of the issues below). We understand that Post Office is continuing to investigate the nature, scope and output of wildcard search functionality, (with KPMG) what overlap it has with Address Book and participant based searches (and with data held in Mimecast), and that HSF and P&P have identified that the approach returns false positives (in the sense that emails to/from/cc/bcc are not always of the custodians searched for).

- (e) In relation to additional custodians in relation to Inquiry Phase 5, the primary immediate further work would relate (under s21 (03)) to other custodians who might have referred to any of the specific individuals in relation to specific issues. We will write separately to the Inquiry on that issue as it is not readily possible to set out the potential thinking without setting out the nature and content of s21 (03).
  
- (f) In addition, BSFf are considering whether previous Rule 9 requests made of Post Office and directed to HSF may be relevant to Phase 5 and, where that is the case, whether and to what extent the search strategy involved searching for emails the completeness of which may be affected by the Exchange/Mimecast issue. AS with each of the issues above, work on the best approach, options on which direction from the Inquiry will need to be sought, and resulting timescales is ongoing. Post Office anticipates writing to the Inquiry to keep it updated ahead of the disclosure hearing on 12 January.

**How is Post Office investigating technical analysis to process Exchange data to reduce the number of duplicates?**

76 Details of this process and analysis are in Appendix 2.

**E. STRUCTURAL (EDRM) REVIEW [NOVEMBER REQUEST PARAGRAPH 2]**

77 Paragraph 1 of the November Request stated:

*“Please set out the detail of the structural review. This should include the following:*

*a. In simple terms, what the review involves and the phases/issues which are likely to be affected.*

*b. When the work on the review commenced, including the reasons for it not being undertaken at an earlier stage.*

*c. What work remains and the date on which such an exercise is likely to be completed.”*

#### **What the review involves and the phases likely to be affected**

78 I summarised the reasons for and the approach to Post Office’s structural review in my letter to the Inquiry dated 1 September 2023 (my first letter as RLR) [POL00126339] I have set out below the relevant paragraphs (I have not included the footnotes from the letter):

*“...In the light of these factors [see paragraph 11 of that letter], and the opportunity afforded to us to assess the position during the course of a transitional hand-over period, the structural review to which Diane Wills refers in her statement, is being taken forward by revisiting the EDRM (Electronic Discovery Reference Model) stages. The Inquiry will be aware EDRM is the generally recognised global methodology for complex disclosure exercises. It involves looking separately at each of the key stages of identification; preservation; collection; processing; review; analysis and production. In practice that involves a system review of all*

*sources of data and systems (electronic and hard copy), how they are being captured and processed. It will also involve looking at the viability (or not) and time involved (if viable) of restructuring the Relativity databases. That structural review is underway.*

*We are conscious that there has, for example been very intensive work ongoing on hard copy data repositories and that the Inquiry has received updates from HSF on this on 22 August and on 31 August. The same confirmatory exercise is being carried out in relation to digital repositories and also the interactions between different systems.*

*This is to check, to the best level achievable, all of the relevant elements that make up POL's disclosure in the light of the factors summarised at paragraph 11 above: sources of data; types of data; those that have been successfully captured and those that remain to be captured for potential relevance to the remaining Phases of the Inquiry, how it is currently held and accessed in Relativity and whether this can be improved. Each of the implementation processes and actions (all of the stages in the chain of what is being done by whom) will be looked at to seek to reduce risks and make any achievable improvements.*

*This is being done mindful of the reality that the focus of attention and review to date evidentially has been on Phases 2-4. We do not have anything like the same level of knowledge and detail on those phases and related work as do HSF and therefore defer to them on that issue. However, from our understanding and involvement since our instruction our sense is*

*that detailed and thorough searches and data collation have occurred in relation to those Phases. The focus in the review is therefore on Phases 5-7.*

***The relevance of the review to POL's support for the work of the Inquiry***

*The work on the review will of course continue in parallel with our work in responding to the live requests from the Inquiry and we do not anticipate it impacting negatively on that. However, the issues set out at paragraph 11 above add an additional layer of complexity to that work. In terms of timing we anticipate that the review itself will take a number of weeks. If structural changes to the Relativity database are viable and bring material benefits, the scale of data and resulting processing time is likely to take 12-15 weeks. However, that structural work (if actioned) would be done in parallel with continuing review work in the existing system and should not affect that continuing work.*

*The review work is required to be able to enable POL to comply with current requests of the Inquiry in relation to Phase 5 and future such requests. In particular, POL wishes to establish that all ascertainable data sources have been identified and collected to the full level reasonably achievable so that the review pool contains the source data potentially relevant to the specific request/requirement. The review work on the existing pool will continue whilst that is done in parallel.*



*The aim of the work will also be – as with any complex disclosure exercise – to inform the necessary interactions and balancing between different factors including resource, efficiency, and depth of review achievable relative to different timescales. Those factors obviously involve unavoidable choices in any review exercise – for example on depth achievable vs time available. The aim will remain that the support from POL to the Inquiry can be effective and efficient. However, the reality is that it will not be possible to mitigate all of the factors set out at paragraph 11 above. Many are historic matters inherent in a disclosure exercise of this nature...”*

79 That remains an accurate summary. In overview the structural review involves:

- (a) the testing and validation of past assumptions for the Post Office’s disclosure exercise;
- (b) the assurance of disclosure-related work conducted to date against objective standards; and
- (c) where a need is identified, the completion of reasonably achievable remediation work.

80 The Electronic Discovery Reference Model (EDRM)<sup>12</sup> breaks a disclosure exercise down into the stages of identification, preservation, collection,

---

<sup>12</sup> I understand the EDRM was developed almost 20 years ago and has been updated and improved on an ongoing basis ever since by lawyers and other professionals active in eDisclosure across common law jurisdictions. I understand EDR M

processing, review, analysis, production and presentation. While the framework moves from the left-hand side (i.e., identifying potential sources of information and determining their scope, breadth and depth, location, availability and known limitations) to the right-hand side (presenting material before the appropriate forum (e.g., a court or public inquiry), it is not strictly a linear process; each stage interacts with the others, so approaches are updated as knowledge increases of the specifics.

81 If stages on the left-hand side have not been conducted effectively, problems are particularly likely to arise and they are likely to compound through the EDRM stages. Putting it simplistically the quality of review is obviously heavily dependent not only on how the review is done but the quality and completeness of the review pool on which it is based. However, for the reasons that I summarise at paragraph 31 above that is – in any complex exercise – easier to state than to implement. The focus of the current structural review is on the left-hand side of the EDRM (i.e., identification, preservation, collection and processing).

82 The review is not directed at a single (or multiple) specific Inquiry Phase or issue. The main activities are:

- (a) An exercise to identify the full extent of Post Office's electronic data universe to validate assessments of all data sources as to whether

---

to be the global standard disclosure framework (although it is expressed to be for electronic disclosure, the framework works – and is used – also for hard copy documents).

they are reasonably likely to contain data that might be relevant to the Inquiry's terms of reference and whether that any such data has been preserved and, as appropriate, collected and processed for review.

- (b) A consolidation of Relativity workspaces to reduce time and operational complexity when responding to requests from the Inquiry and to ensure that newly processed data has more coherent metadata, which will enable more effective use of other Relativity functionality (such as email threading and textual near-duplication).
- (c) An exercise to validate custodian data mapping to facilitate an assessment of whether further identification work is required and whether any additional preservation or collection actions need to be taken (e.g., in relation to potentially materially relevant data in the possession of third-party professional adviser or individual custodians).

**When did the work commence and why was it not undertaken at an earlier stage?**

83 The specific structural review to which I referred in my letter to the Inquiry dated 1 September 2023 [POL00126339] being conducted by BSFf started following our instruction in May. We had anticipated the use of the EDRM structures for disclosure work in our tender submission and discussions on

it with Post Office therefore started during June and July during mobilisation.

84 It would not however be a fair inference to view the structural review as an indication that investigative work to identify, preserve and collect sources had not previously been conducted for the purposes of Post Office's disclosure exercise. My understanding is that extensive elements of investigation work have been done by Post Office and by various advisers over several years (including during the GLO proceedings). The Inquiry receives, for example, detailed updates from HSF on the work that it is carrying out.

85 I understand and, since the instruction of BSFf, have seen that detailed efforts were undertaken to review factors relating to disclosure. However, a comprehensive understanding of Post Office's data universe has not yet been achieved and it is, and has been, developing as institutional knowledge is reconstructed.

86 Complexity has also been due in part to the fact that it builds on several previous waves of overlapping disclosure processes in the past including for the GLO, Criminal Appeal proceedings, malicious prosecution proceedings as well as for the Inquiry. My understanding is that this has meant that Post Office's main data repositories have been identified and harvested from time to time, for multiple purposes, from multiple sources and into multiple Relativity workspaces. Post Office had, prior to our instruction, instructed and actively participated in more wholesale structural

disclosure activity intended to develop its Inquiry-related disclosure, recognising that its data repositories and its knowledge of those repositories had built up over a series of layers and years. Much of that work has involved taking fresh reviews of Post Office's data universe and identification of electronic and hard copy data.

87 For example, Post Office had already conducted (commencing in around December 2022) a hardcopy audit procedure (which BSFf has not advised on or been involved with) of 228 Post Office site locations. I understand that an enhanced self-certification process was supplemented by an on-site search conducted by a team from Innovo Law, comprising 2 solicitors with public inquiry experience and a former police officer who has previously led investigative work on public inquiries.

88 As part of the work, Post Office from around May 2022 onwards has been building up internal knowledge of its current and legacy IT architecture. These processes are ongoing. Post Office started in June 2023 the exercise to identify the full extent of Post Office's electronic data universe. That work was initially progressed by the same team from Innovo Law and a group of Post Office subject matter experts coordinated by Post Office's Chief Data Architect. It is continuing now as a collaborative project with advisory inputs from senior BSFf lawyers, feedback from BSFf's front line disclosure teams and a wider range of Post Office SMEs, who combine seniority, relevant technical expertise and residual long-term institutional knowledge. That first stage work of identifying Post Office's potentially

relevant data universe is expected to complete within the next few weeks and feed into the structural review.

- 89 I also initially raised with KPMG in July 2023 the possibility of consolidating or rationalising the multitude of Relativity workspaces to facilitate better de-duplication, email threading and analytics at workspace level and to facilitate faster electronic searches and review. Post Office have taken that project forward and we are liaising with KPMG to develop and scope a methodology to improve delivery of disclosure to the Inquiry. Care is being taken not to impact work in responding to Inquiry requirements and also to try to avoid adding rather than reducing complexity. Post Office has already approved the concept in principle providing that cost-effective options are technically feasible within a timeframe that allows Post Office to best assist the Inquiry.

**What Work Remains and on what date is it likely to be completed**

- 90 I have summarised the main ongoing activities of the structural review at paragraph 82 above. Appendix 3 contains further detail about those activities, the current position on each and the known future work. I anticipate that unfortunately some elements of the work would require some further months to complete and some of the timescales are not predictable with certainty. However, I have included Post Office's current best estimates of duration. For example, the examination of the 5 servers

and 12 back-up tapes located at Chesterfield<sup>13</sup> is not easy to predict due to technical complexities. The timescales are driven primarily by technical work required and the delivery of that before any resulting legal work can be scoped and carried out.

91 New potential data sources are being identified that require further investigation. There is a resulting balance of the need to inform the Inquiry on new issues on disclosure which may impact its work without undue delay. However, equally the Inquiry will not wish to be troubled on items that turn out to be not relevant. For example, the recent Post Office FileShare correspondence from BSFf to the Inquiry dated 10 and 17 November 2023 [WITN10810110] [WITN10810111]. If helpful to do so we will continue to outline potential repositories but not trouble the Inquiry with detail until after investigations have progressed meaningfully.

#### **F. RESPONSE TO PARAGRAPH 1 OF THE OCTOBER REQUEST**

92 I have set out responses below on points arising from BSFf's letters to the Inquiry dated 16 [WITN10810102] and 13 October 2023 [WITN10810101]. For the reasons mentioned at paragraphs 31(d) and 54 above, the aim of the letter of 16 October and the covering email and draft agenda [WITN10810112] was to summarise Post Office's understanding and current approaches for engagement with the Inquiry and to seek a meeting

---

<sup>13</sup> 6 servers are being examined but 1 has been confirmed, I understand, not to contain stored data. 13 back-up tapes are being examined, but 1 has been confirmed, I understand, to be a cleaning tape that would not contain stored data.

so that any points of concern for the Inquiry on the approaches and related timing impacts could be discussed, addressed and any practicable changes incorporated.

- 93 BSFf's letter to the Inquiry dated 13 October 2023 [WITN10810101] replied to the Inquiry's letter to BSFf dated 9 October 2023 [WITN10810113] in relation to BSFf's letter to the Inquiry dated 11 September 2023 [WITN10810114]. BSFf's letter of 13 October aimed to provide the clarifications sought and sought discussion on any points or concerns:

*"We hope that this letter assists to clarify matters raised in the Inquiry's letter of 9 September 2023. We would welcome a call with the Inquiry following receipt of this letter to discuss the points and any ongoing concerns, particularly around the Inquiry's concerns on duplicates. We can make ourselves available at any time next week on Wednesday 18th or Thursday 19th October 2023."*

- 94 A meeting was subsequently arranged for 3 November 2023. The October Request was issued on 31 October 2023. The points set out below have not to date been the subject of discussion with the Inquiry. I have therefore summarised the position below.

- 95 The context and practicalities of disclosure exercises generally summarised at paragraphs 31 to 32 above was the background for paragraphs 6 and 7 of BSFf's letter to the Inquiry dated 16 October 2023 [WITN10810102]. The letter was seeking to be direct about different



dynamics involved between the remediation exercise for the Three Issues (as defined by the Inquiry) as against ongoing disclosure.

96 The remediation exercise for the Three Issues was of course to be completed with the very full levels of rigour that resulted from the discovery of the problems with item level de-duplication, specific search terms and family documents.

97 I also confirm in relation to disclosure work going forward that:

(a) KPMG has been instructed and is proceeding on the basis that global/family level de-duplication should be used. Item level de-duplication will not be applied without specific agreement. KPMG has confirmed that no review involving a request by the Inquiry since the concerns with the Three Issues arose has involved the use of item level de-duplication in a manner that would exclude documents from review<sup>14</sup>.

---

<sup>14</sup> Item level de-duplication has not been used to exclude documents from review. HSF and P&P have, I understand from HSF, on occasion, instructed KPMG to use item level duplicate analysis to identify documents within a draft production set that: (i) are exactly duplicative of documents that had been produced to the Inquiry previously; and (ii) which are either standalone documents or documents attached to parent documents that had been reviewed and assessed as providing neither additional relevant content nor context. HSF has stated, by way of example, this was explained in HSF's letters to the Inquiry dated 14 [WITN10810115] and 25 August 2023 [WITN10810116].

- (b) KPMG has confirmed that no item level de-duplication has been applied to any review exercise or production carried out by the BSFf team.
- (c) Search terms are being examined, tested and refined rigorously.
- (d) Post Office and BSFf (including the senior BSFf individuals responsible for operational and strategy decisions regarding Post Office's Inquiry disclosure exercise) considered all of the correspondence, evidence and Directions relating to the Three Issues and used it to inform and develop our approach for each EDRM stage of Post Office's Inquiry disclosure exercise.

98 Search strategies are necessarily specific to each request. To date there have been two main notices that have led to review and productions that BSFf has dealt with (leaving aside those recently served and currently in progress): these are s21 (03) and s21 (08). For reasons of Inquiry confidentiality I will not go into the detail of these in this witness statement because of its likely circulation. However, the difference between them is useful to illustrate the different types of approach that are necessarily applied to different situations.

99 S21 (03) involved multiple individuals and multiple issues extending over a long period. The Inquiry specifically requested that the issues were interpreted broadly and that Post Office took an abundance of caution approach (both of which were done). It also involved, in capturing those

issues many phrases or search terms common in everyday language, many of which were also in use in operational contexts within Post Office over the more than 20-year period covered by the Inquiry's terms of reference.

- 100 In contrast, s21 (08) involved the identification, collation and provision of a specific and defined cohort of material to the Inquiry. For example, where it was possible to identify one example of a document requested then information in that document led to searches being conducted for documents with similar or the same wording and formatting. The learning from further documents identified then helped to identify further lines of enquiry. The enquiries became more focused and granular until all reasonable searches and lines of enquiry had been exhausted.
- 101 Different – although both rigorous – search strategies in relation to families (both designed to address and head off the potential issues which had arisen as part of the Three Issues) were adopted therefore in each case.
- 102 In dealing with the review for s21 (08), because the review pool was much narrower and because this was not an exercise involving multiple emails with large families and very large amounts of duplication, every related document was reviewed. Where any document was relevant, the surrounding related documents were produced. In terms of scale, this exercise led to 199 documents produced. This review exercise, although narrower in scope, still required a large team working full time (or nearly full time) for a material period.

103 In contrast, because of the nature of it, s21 (03) had a potential review pool in the low millions if broad search parameters were used and over 500,000 documents in the review pool when the parameters were refined and iterated. In terms of approaches to family documents, for this specific notice, BSFf therefore had to devise a methodology that sought to avoid the problems that had arisen within the Three Issues (minimising to the extent practicable the risks of not picking up relevant documents within families) whilst being able to respond with relevant evidence within manageable timescales and reducing the – already material – risk of adding to duplication or marginal relevance issues in the production set for the Inquiry. I have responded below on de-duplication including recent work with KPMG on that to try to assist.

### **Review of Family Documents**

104 To address the family documents and duplication factors, the instruction given to Tier 1 reviewers on review of family documents on s21 (03), and by default for all disclosure workstreams unless a particular workstream requires a different approach is that below. This is instructed in the detailed formal guidance and also training given to reviewers:

#### ***“Reviewing family documents***

*18. The Inquiry has stressed the importance that potentially relevant documents are reviewed in their family context, rather than in isolation. The documents that you will review in batches will not be in full families, so you*

*will need to click the Family Group icon at the bottom right corner of the screen when viewing a document, to view the "child documents".*

19. *Where you code a document in your batch as relevant to the section 21 you then MUST code all documents in the family group of that document. You may need to review the family documents in order to determine the relevance of the parent document.*

20. *If you code a document as Not Relevant, do you not need to view and code the family documents UNLESS there is anything in that document to suggest the child documents are relevant, for example the title of an attachment suggests possible relevancy ("GLO talking points" etc.).*

21. *It is important to remember when you code a document as Not Relevant, you not only exclude that document from further review but ALL child documents, unless they are Search Term Responsive (and will therefore be reviewed at Tier 1). Therefore it is important you are certain of relevancy before coding a document as Not Relevant. Always err on the side of caution and code as Relevant when you are uncertain.*

22. *The Inquiry has stated that a parent email to a relevant document will most likely be relevant. Please err on the side of caution and code all parent emails to relevant documents as relevant, unless they meet the specific criteria for exclusion at paragraph 4 of the Notice (see page 14 for guidance). If you are unsure if a document meets the criteria for exclusion, mark it as relevant."*

- 105 The guidance to reviewers, including the above concerning approach to family documents, was produced and then updated in light of the correspondence, evidence and Directions concerning the Three Issues.
- 106 As an example, where a reviewer marks parent document (A) as irrelevant, its family documents (B) and (C) would not be reviewed at that stage. However, if (B) and/or (C) are themselves responsive to search terms applied to identify documents potentially relevant to the Notice/Request then they will be included in a batch of documents and will be reviewed separately.
- 107 An alternative approach would be one that looked to review every family member of every document that had initial 'search term hit' even if that document and its parent had been reviewed and found to be non-relevant. Conservatively, because of the nature of families in digital communications, that might well increase the number of documents that had to be reviewed by a factor of three or four times or more. In a small review that has little impact (so was the approach adopted in relation to s21 (08)). However, in an exercise involving several hundred thousand documents, the impact on timings and responsiveness would be very material.
- 108 It was understandable that such an approach needed to be taken during the remediation of the Three Issues, following the issues with item level de-duplication in relation to the remediation of prior requests. However, with complex requests involving large review pools, targeted risk mitigation search strategies in place and global/family level de-duplication having

been applied to the review pool, applying it would create material difficulties. If the result was also a relevance rate of around 1.5% to 2.0% (extrapolating from the relevance rate found during the remediation exercise), the level of additional time/delay in production and cost would be very significant relative to the level of gain. The risk of adding documents of marginal or no relevance and, therefore, adding to the problem of duplication for the inquiry in productions might well also be increased.

### **Reconsideration of Previous Searches**

109 In relation to reconsideration of previous searches:

- (a) The Three Issues remediation exercise has, I understand from discussions with HSF, P&P and Post Office, sought to redress the impact of the Three Issues in relation to affected Rule 9 requests (as has been reported to the Inquiry in regular correspondence sent between July and November 2023).
- (b) BSFf and Post Office will of course be alive to and will consider new information that arises, which may impact upon POL's response to a previous request. For Inquiry Phases 5-7 Post Office and BSFf are, and will continue, to look at the best practicable search strategy for each request and will put in place quality control and risk mitigation measures as appropriate. These will be, and are, all kept under review. Such an approach will ensure that Post Office meets each of the Inquiry's requests as far as reasonable in all of the

circumstances noting the constraints that I have highlighted in Section C above and the measures that are being taken to try to address some of those constraints.

- (c) In addition, where documents of interest likely to be of relevance to the Inquiry's terms of reference and therefore to the Inquiry are identified by Tier 1 or Tier 2 reviewers, even if they are not immediately responsive to the particular request, these are escalated and produced to the Inquiry. This happened recently, for example, in relation to a body of additional material identified during the s21 (08) review.

#### **Use of Search Terms vs. Enquiry Based Searches**

110 In terms of the use of search terms alongside other search techniques, this very much depends upon the specific requirement in particular section 21 notice or rule 9 request. When a request is relatively narrow and can be targeted for example at a particular period, relatively limited set of custodians or category of documents, it may well be practical and necessary to review within those parameters all documents identified. Where the request requires production of broader categories of documents covering multiple topics then search terms have to be used to identify a review pool of documents that is realistically capable of review.

111 For the reasons set out in my letter to the Inquiry dated 1 September 2023 [POL00126339], use of technology/AI techniques alongside other search



approaches has not to date been viable. For such techniques to be possible the data and the underlying metadata must be consistent or have at least a large degree of consistency, and the eDisclosure provider must have sufficient understanding of history and management of the whole dataset. That is not the case with the data in the existing Relativity database. This is largely because of the variable quality of the data accumulated over time and the diffuse nature of the total Relativity database, which has been built up incrementally and from different sources, in respect of some of which KPMG has had limited (if any) visibility over collection, processing and management. The consolidation proposal (discussed above in Section E), once implemented, involves improving the quality of the data as well as rationalising the various databases. Therefore, it may be possible in the future to use some of these techniques in parallel with other search methodologies. We would only however wish to do that after discussion of the approaches with the Inquiry. Also, I am mindful that the timescales and the critical path for the Inquiry hearings might in practice mean that this potential additional capability will only come in at a relatively late stage, and therefore may have to be targeted at specific issues (in discussion with the Inquiry) in a way that best assists.

### **Potential Engagement with the Inquiry**

112 In all these areas above, I am conscious that they involve professional value judgements as to the best approaches. We will continue to set out in response to each request received how we have approached the particular

review and the reasons we have approached the particular review in that way. I reiterate however that Post Office would welcome discussion if this would assist the Inquiry following the point of receipt of a particular request (or even in draft confidentially in advance in scoping it) to best adapt the way in which it can respond practically to the Inquiry's recommendations.

**G. RESPONSE TO PARAGRAPHS 2-4 OF THE OCTOBER REQUEST**

113 We fully appreciate the real problems that duplicates and near duplicates present for the Inquiry (and, for completeness, for Post Office and its advisers) in conducting the review.

114 At the core of the problem is a need to reconcile two competing priorities:

- (a) the requirement not to exclude documents that:
  - (i) appear ostensibly the same but may have one or more differences, some of which may prove to be important; or
  - (ii) are in fact exact duplicates but appear in different contexts and that context itself may be significant to an understanding of, for example, whether an individual was not aware of a particular circumstance or set of facts.

and

- (b) the difficulties of volume and repetition (with resulting duplication of effort and time) caused by production of identical or very similar repeating documents or sets of documents.

- 115 That conundrum is compounded in large exercises which cover long periods, multiple issues and multiple individuals who will be interacting and communicating in different ways and in different combinations. The issue is therefore not an inability to identify duplicates at a specific time. Rather the issues in BSF's letter to the Inquiry dated 13 October 2023 [WITN10810101] related to ways the priorities discussed in paragraph 114 might be addressed operationally in a way that meets the Inquiry's requirements.
- 116 This problem is compounded by material variances in the quality of the data held in the Post Office Relativity database. As I note above, the reasons include different data having come from different sources and different applications using different processes at different times and in held in different workspaces within Relativity. The variability can result from, for example, the same document or similar documents being held in different formats (images or text or other) or because different applications create different metadata. To the disclosure systems therefore they are different documents even if all the content is identical. What is meant therefore by identification of duplicates is not straightforward. In any disclosure context strictly only an identical MD5 Hash (or #) test can be immediately identified as a full duplicate. I have set out some further context below.
- 117 There are several technological methods to de-duplicate, but each has limitations and potential downsides:

- (a) MD5 Hash – This is an industry standard de-duplication process discussed in detail at the 5 September 2023 disclosure hearing. It uses a highly sensitive (and therefore precise) algorithm to de-duplicate documents, but, consequently, even tiny changes to metadata between two documents (such as might arise if they were processed onto Relativity using different software or different processing criteria) would result in it treating those documents as different. It is therefore safe to operate in the sense that it would take a highly precautionary approach and only exclude absolute duplicates. As a result, however, it will admit into review and production documents with tiny (including probably many inconsequential) differences.
- (b) Textual Near Duplicate (TND) identification – This is another industry standard process to identify duplicates, that can be highly useful, but must be used with care. While setting a 100% minimum similarity percentage parameter would group exact textual duplicates, differences in the metadata might be of importance and mean that it would not be appropriate to use this method to de-duplicate. Further, any lower setting, even 99%, could lead to unpredictable results. For example, the 1% difference between two documents could indicate a likely irrelevant divergence, such as different renderings of the same URL (for example, hyperlinked text that says “*Click here*” may render in one email as the text “*Click here*” and in another as the hyperlinked URL). It could, however,

indicate an important difference that significantly changes the substance of the text (for example, the difference between: “X must do that” and “X must not do that”). The TND process is agnostic as to the cause of the distinction. It also has compatibility issues with certain document types or where optical character recognition<sup>15</sup> is inaccurate (OCR has technical limitations, particularly with handwritten documents or documents with manuscript comments). I also understand that Relativity identifies a principal document against which others in the TND group are compared to generate the percentage difference. However, Relativity identifies the principal document as the one with the greatest amount of text on the assumption that would be the most complete document. That may not be the case, and consequently the document(s) of most interest may be assigned percentage similarities lower than 100%.

- (c) Custom processing hash – I understand this to be very similar to MD5 Hash, with corresponding advantages and disadvantages. This is not an industry-standard approach and must be used with care. While it can be a useful tool, it relies on having a strong understanding of the data set, which is not always available. Essentially, this involves designing a particular combination or set of parameters tailored to produce a particular result based upon the

---

<sup>15</sup> OCR is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine -encoded text. It is the process that makes the text in, for example, a PDF searchable.

particular fields which will be most effective when applied to the specific document population involved.

- 118 For these reasons great care is needed when considering using any of the techniques other than full MD5 Hash de-duplication to exclude documents from either a review pool or from a production set. To do so would introduce material risk of excluding something with minor but potentially significant differences. That would also not, as we understand them, meet the inquiry's current requirements for the reasons at paragraph 114(a) above.
- 119 However, TND and/or Custom hash techniques can more readily be used to prioritise work following production. With this in mind, BSFf has been working closely with KPMG recently as set out in paragraph 2.16 of BSFf's letter to the Inquiry dated 13 October 2023 [WITN10810101]. BSFf wrote to the Inquiry on 24 November [WITN10810117] regarding an additional load file in relation to s21 (03) and to set out suggestions as to how TND and other specified coding fields could be deployed by BSFf and/or the Inquiry (if the Inquiry would find it helpful) to separate out near duplicates, or other categories for lower priority review and to target high priority documents or areas.
- 120 The issue is compounded by the various scenarios in which documents may exist, relate to each other and have been produced. These scenarios impact the volume of data to be searched, that is produced, and the nuances which any de-duplication process must take into account.

- 121 I also attach at Appendix 4 a further technical schedule that we have prepared with KPMG that summarises these and further options. Those options in Part A are those that the Inquiry might consider using on the material which has been provided to it by BSFf on behalf of Post Office. Those in Part B are further options which BSFf and KPMG could run against the full dataset, using additional levels of Custom hash technique (with the consent and involvement of the Inquiry and its eDisclosure providers) to further refine and de-duplicate for the purposes of prioritisation. By using these techniques, we are hopeful that the Inquiry's document reviewers, solicitors and counsel team would be able to de-prioritise a large quantity of near duplicates in their work, whilst having the ability subsequently to circle back round, as required, to look at those documents as needed in context or if there are minor variants that prove to be potentially evidentially significant.
- 122 By way of illustration, we have identified that emails that are the same in substance and metadata may be identified by Relativity as non-duplicates due to the inclusion or content of disclaimer wording at the bottom of the email chain. For example, emails are identified where one contains a disclaimer and another, which appears to be a duplicate, does not. In addition, we have identified that in emails sent to 2 or more recipients, the disclaimer that appears in the email received by each recipient is updated to refer to them meaning they have different disclaimers and each copy of the same received email has a different MD5 Hash value.

123 We welcome the opportunity to assist further also in operational discussions with the Inquiry and/or its eDisclosure providers<sup>16</sup> if helpful. To support this, BSFf has provided an update by letter dated 15 December 2023 on its proposed approach to seek to assist the Inquiry in relation to duplicates and near-duplicates during review of documents by BSFf on behalf of POL and after production of those documents [WITN10810118]. In addition, BSFf has provided an additional loadfile containing metadata fields, which hopefully will assist the Inquiry (see the letters dated 24 November 2023 [WITN10810117] and 15 December 2023 [WITN10810119]).

124 In relation to paragraph 2 of the October Request, the factors and concerns that led to the proposal to disclose by way of list previous documents that were responsive to the search terms used, were specific to s21 (03) in summary these were:

- (a) The volumes involved and resulting delay and impact on timings of response.
- (b) The duplication of review activity – and resulting cost – between the Inquiry and Post Office/its external reviewers. The material

---

<sup>16</sup> We anticipate that the Inquiry has already been considering HSF and BSFf productions for Inquiry Phase 5, and potentially Inquiry Phases 6 and 7, cross-referring to documents produced. We are concerned to ensure that any de-duplication effort has minimal impact on any cross-references the Inquiry has already made. We are concerned to ensure that any de-duplication does not inadvertently de-duplicate the specific copies of documents on which the Inquiry's preparations rely.



comprises documents which had all been provided evidentially to the Inquiry by Post Office, as relevant to the Inquiry's terms of reference in response to previous rule 9 and section 21 obligations and therefore available to the Inquiry. As at 31 October 2023 I understand that 151,580 documents had been disclosed pursuant to 47 Rule 9 requests and 11 Section 21 Notices (other than s21 (03) and s21 (08)).

- (c) To seek to reconcile those factors with the Inquiry's request, we therefore provided a list of prior disclosed documents responsive to the search terms used for s21(03).

**Please explain whether any significant changes have been made in respect of the resourcing of POL's Disclosure exercise since the last disclosure hearing held 5 September 2023 (i.e., numbers of reviewers, hours worked etc.)**

125 The resourcing of POL's disclosure exercise has materially increased since 5 September 2023 (and in fact prior to that as BSFf mobilised and built up its own disclosure team as documents became steadily available for review and as they received individual CU clearances). This is because the disclosure teams deployed by HSF and P&P have remained intensively busy in connection with remediation (relating to the Three Issues), Inquiry Phase 4 work and the further disclosure activities reported on an ongoing basis to the Inquiry. 4 law firms (BSFf obviously being an integrated team

drawn from 2 firms) each now have large disclosure teams deployed and working intensively.

126 In terms of the relative sizes of the different teams, these are broadly similar between HSF and BSFf. Indicatively, the teams for each have been since 6 September 2023 (these are not full time equivalent – FTE – numbers<sup>17</sup>):

(a) HSF: 171 total colleagues involved across all activities (lawyers, other professionals and colleagues) of whom 68 were trainees, paralegals or other first-tier reviewers)

(b) P&P 45 total colleagues involved across all activities (lawyers, other professionals and colleagues) of whom 17 were trainees, paralegals or other first-tier reviewers).

(c) BSFf: 175 total colleagues involved across all activities (lawyers, other professionals and colleagues) of whom 80 were trainees, paralegals or other first-tier reviewers).

127 It is not possible to identify the number of lawyers on disclosure activities specifically. Whilst trainees, paralegals and other first-tier reviewers will be focused wholly or primarily on disclosure, more senior team members,

---

<sup>17</sup> Each of HSF and BSFf had similar numbers of colleagues – in the low tens – within these total numbers who were each involved during the period to a relatively low level of activity (under 20 hours)

certainly within the BSFf team, with some specialist exceptions, are mostly involved in a mix of disclosure and other work.

128 In terms of work on disclosure activities the number of disclosure-related hours worked per week (on average) during the period has been at approximately the same level for BSFf relative to HSF/P&P combined.

**Statement of Truth**

I believe the content of this statement to be true.

**GRO** .....

Christopher Michael Jackson

Date: 19 December 2023

## APPENDIX 1: POST OFFICE EMAIL SYSTEMS TO 2016

- 1 For ease of reference, where I refer to “Royal Mail Group” in this appendix I refer to Royal Mail Group Limited and/or its relevant predecessor at relevant times as applicable.
- 2 A diagram reproduced at the end of this appendix, which was provided to BSFf by Post Office, illustrates the evolution of Post Office’s email systems and data repositories. However, in terms of the key events in the evolution of Post Office’s email data repositories, BSFf’s understanding from Post Office is as follows:
  - (a) Until the early 2000s, it is understood that Royal Mail Group used early versions of Microsoft Mail or MSMail. Post Office has very little information relating to this period in relation to email data including as to quantity although email would be expected to have been used on a relatively limited basis. In any event, it does not believe that it has and it has not encountered any email archives from this period. Post Office is therefore not aware of any email data repositories from this period.
  - (b) From the early 2000s, Royal Mail Group started using IBM Lotus Notes as its email client combined with a Lotus Domino Server. It also additionally utilised a Sendmail Gateway. There was no journaling functionality at gateway level. We understand from Post Office that due to mailbox memory sizes at the time, it was not

uncommon for users to create local archived email data to keep older emails. However, due to migration programmes since and passage of time, few of these snapshot repositories would still exist and Post Office is not aware of any structured repository of Lotus Notes data archives. Some of these however still exist (either in native .nsf Lotus Notes format or that have been converted since to .pst files), can be found on Post Office's SharePoint/OneDrive network and have been previously searched for and located email data has been processed onto Relativity. Where these old archives have been stored on physical storage devices or media instead, as and when located by or provided to Post Office, the possibility that they may contain non-duplicative email data is assessed and data harvested as required (e.g., the hard drives and back-up tapes located by Post Office at Chesterfield, which have been the subject of previous updates to the Inquiry by HSF).

- (c) In and around 2008-10, Royal Mail Group reverted from IBM to Microsoft. It changed its email client from Lotus Notes to Outlook, its email server from Domino to Exchange (hosted on Microsoft Business Productivity Online Suite Dedicated or "BPOS-D") and its email gateway from Sendmail to IronPort (albeit there was still no journaling functionality at gateway level). During this migration process, efforts were made to convert Lotus Notes data repositories (.nsf files) to Microsoft data repositories (.pst files) at server mailfile level and for local archived email data. As with any such conversion

process at this scale, it is understood that there will have been individual instances of legacy data loss, but Post Office has no information as to material events of data loss.

- (d) After migration from IBM to Microsoft, it remained possible for locally archived snapshots of email data to be created by users and others and stored on drives or physical devices and media. However, from this period with the adoption of cloud-based storage and applications such as SharePoint, the practice by email users of creating local archived email data on physical devices or media became increasingly discouraged until no longer permitted in practice. When any such physical devices or media are located by or provided to Post Office, the possibility that they may contain non-duplicative email data is assessed and data harvested as required.
- (e) In and around 2011/12, Royal Mail Group changed its email gateway provider from IronPort to ProofPoint. For the first time, ProofPoint introduced email journalling at gateway level in a manner broadly equivalent to that described for Mimecast at paragraph 36(c)-(d) above in the main body of my witness statement. However, given the historical restructuring context and length of time that has passed, Post Office does not have information as to:
  - (i) exactly when ProofPoint may have been activated by Royal Mail Group in 2012 (but understands that it would not have been a uniform date for all individual users in any event) and

data captured for some users will go back to 2011 if they were pilot users; and

- (ii) retention periods and other settings that were applied to ProofPoint by Royal Mail Group during its mobilisation and operation. Although Post Office has no reason to believe that Royal Mail Group would have applied deletion settings to ProofPoint that would materially affect the journalling of data between 2011/12 and 2016.
- (f) Additionally, it is understood by Post Office (but not known conclusively) that Royal Mail Group did not export a legacy email data file from Exchange pre-dating ProofPoint into that system. In other words, it is understood that ProofPoint did not ingest pre-2012 email data from Royal Mail Group's Exchange mailfiles into its archive.
- (g) From 2012 to in or around 2016, there were a number of key changes to Post Office email systems coinciding with the period in which Post Office demerged from Royal Mail Group:
  - (i) Post Office adopted Microsoft 365 as its productivity platform (from Microsoft BPOS-D) and user Exchange mailfiles were migrated across at server level;

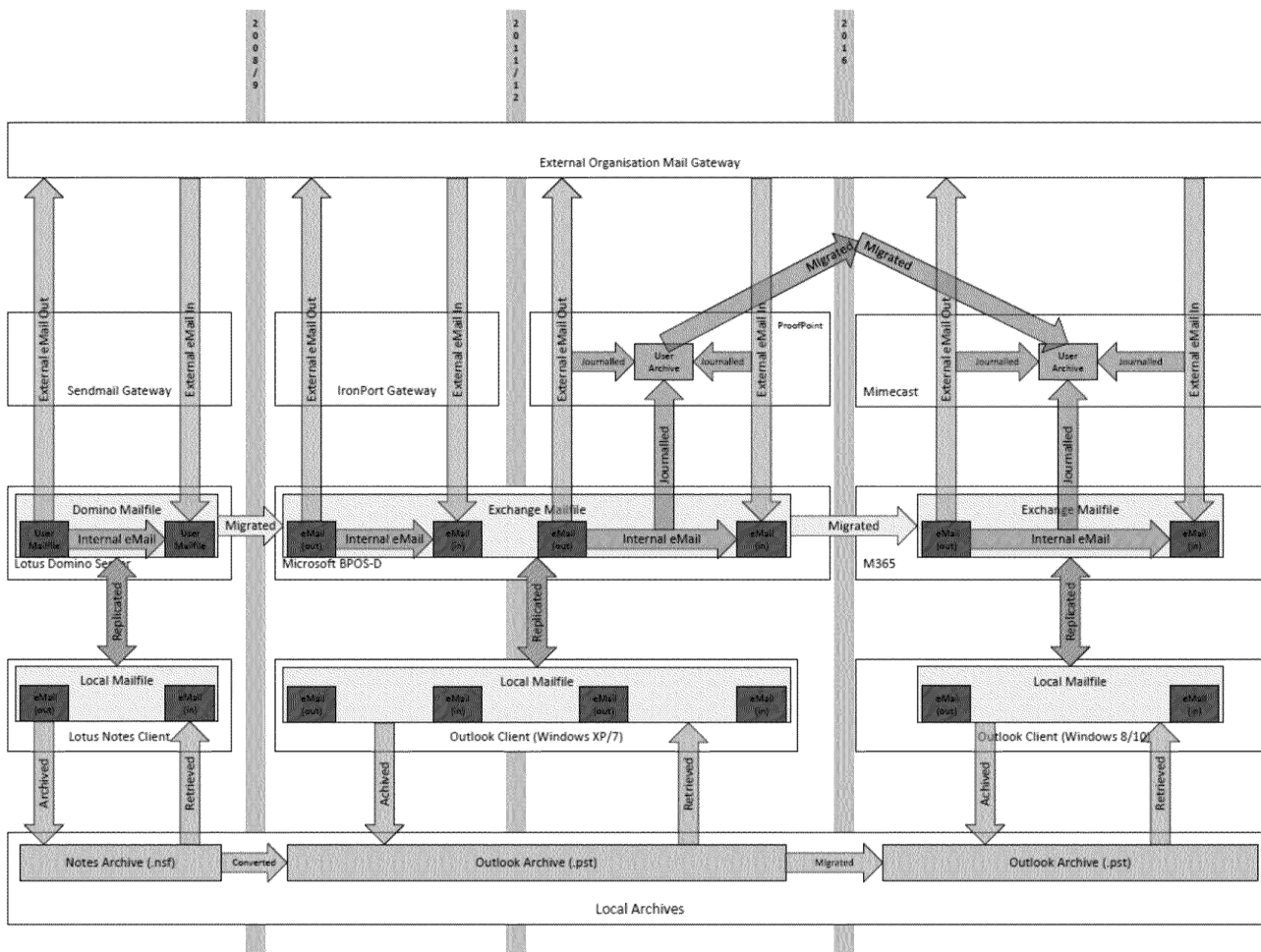
- (ii) Post Office updated its Outlook email client and local mailfiles were restored from the Exchange mailfiles that had migrated across to the new version of Exchange; and
  - (iii) Post Office adopted Mimecast as its email gateway in or around late 2015. The operation of Mimecast once activated is described above at paragraphs 36(c)-(d) of the main body of my witness statement. However, as explained above, the Mimecast data will include any legacy mailfile data exported to it, processed and ingested into Mimecast.
- (h) Post Office has confirmed that the legacy mailfile data ingested by Mimecast at its activation came from Royal Mail Group's ProofPoint email gateway system in or around late 2015 to allow for continuity of the immutable journalled email record at server level. As ProofPoint itself was only activated by Royal Mail Group in or around 2011/12, the Mimecast data would therefore broadly not be expected to contain email data pre-dating 2012. Moreover, Post Office understands from investigations that:
- (i) it was Royal Mail Group that instructed ProofPoint to create the ProofPoint legacy email dataset and it was provided to Post Office on a number of disks ("**ProofPoint legacy data**"); and



- (ii) Post Office does not have records of how Royal Mail Group specifically created the ProofPoint legacy data. However, it does understand that only email data in relation to postoffice.co.uk email addresses were included. This meant, for example, that where a user who had access to multiple email addresses including royalmail.com email addresses, these other emails would not have been exported across to or ingested into the Mimecast data. These non-postoffice.co.uk emails could however remain in their Outlook mailboxes and Exchange mailfile, However, given that formal separation between Royal Mail Group and Post Office occurred in 2012, this particular issue of multiple email addresses likely has only limited impact and only for very longstanding Post Office staff.
  
- 3 Throughout the entirety of the period above, there will in addition have been multiple upgrades and replacements of IT equipment, software, operating systems, physical devices and media, back-up and support systems and more as well as IT issues or system or application failures. Any and all of these events entail the risk of email data loss and doubtless individual instances of data loss did occur during such events. However, Post Office is not aware of any specific events of material email data loss.

DRAFT SUBJECT TO LEGAL ADVICE PRIVILEGE – NOT FOR FORWARDING BEYOND ADDRESSEES IN WHOLE OR PART OTHER THAN AS SPECIFICALLY AGREED

129 Figure – Post Office email flows and systems from early 2000s to date:



---

## APPENDIX 2: POTENTIAL DE-DUPLICATION OPTIONS

### Introduction

- 1 KPMG was instructed to assist Post Office and BSFf with an exercise to analyse whether documents identified within Exchange were duplicate, or duplicative in material respects, with documents that already existed in the Relativity database (for instance, as sourced from Mimecast).
- 2 The Exchange data for 13 custodians was included in the initial scope of this analysis. This data was extracted by POL Cyber for the date range 1 January 1995 to 1 January 2016.

### Approach

- 3 Since email data had been collected from a variety of different sources over time, KPMG's analysis was conducted across four of the main Relativity workspaces:
  - (a) BSFf Processing Workspace (3.6 TB): the main workspace being used as a data repository for BSFf's responses to Phases 5-7 of the Inquiry.
  - (b) POHIT Processing Workspace (19.4 TB): the main workspace being used as a data repository for HSF's responses to Phases 1-4 of the Inquiry.

- (c) GLO Workspace (5.7 TB): the legacy workspace used as a data repository for Womble Bond Dickinson's work related to the Group Litigation Order.
  - (d) CCRC Processing Workspace (1.1 TB): the workspace used for any new data received by KPMG for P&P's work related to the Criminal Cases Review Commission and the Inquiry.
- 4 KPMG considered four different forms of duplicate document analysis when comparing the Exchange data for the 13 custodians with data that already existed in the four Relativity workspaces listed above:
- (a) Relativity Processing Duplicate Hash. Similar to a MD5Hash value, this is a unique and forensically accurate digital fingerprint of a document created by Relativity using a SHA256 hash during processing.
  - (b) Manual Custom Hash. This is a bespoke approach which KPMG uses based on a concatenation of metadata fields: Message ID, Unified Title, and Sort Date/ Time (hours and minutes without the second value).
  - (c) Message ID. A comparative analysis was also conducted using just the Message ID metadata field, where this was available. Manual Custom Hash is a sub-set of this Message ID analysis because Manual Custom Hash relies on Message ID and other fields.

(d) Textual Near Duplicate (TND). This analyses the textual content of the documents to determine a percentage similarity across documents (e.g., 90-100%) and group similar documents together based on textual content.

5 The following table provides a summary of the duplicate analyses conducted by KPMG across the four workspaces, with an explanation for the reasons where it was not possible or desirable:

Workspace	Relativity Processing Duplicate Hash	Manual Custom Hash	Message ID	Textual Near Duplicate analysis
BSFF Processing	Yes: Step 1	Yes: Step 2	Yes: Step 3	Yes: Step 4
POHIT Processing	Yes: Step 5	No <i>(Given the size of the workspace, Message ID was selected as a priority for analysis)</i>	Yes: Step 6	Yes: Step 7

<b>GLO</b>	Yes*: Step 8  <i>(* Relativity Processing Hash not available, MD5Hash used instead)</i>	Not possible  <i>(Message ID not available)</i>	Not possible  <i>(Message ID not available)</i>	Yes: Step 9
<b>CCRC Processing</b>	Yes: Step 10	Yes: Step 11	Yes: Step 12	No  <i>(Given the low results from other analyses – see below - decision taken not to prioritise TND)</i>

**Findings from KPMG 12-step analysis**

6 The number of parent emails for the 13 custodians over the relevant time period extracted from Exchange was 391,775. KPMG's analysis was

performed at the parent level to maintain family context and align how Relativity applies its deduplication logic.

#### **Step 1**

- 7 This data was initially processed into the BSFf Processing Workspace in Relativity and deduplicated using global deduplication at parent level using the Relativity Processing Duplicate Hash.
- 8 This deduplication reduced the overall number of Exchange parent emails to 363,841.

#### **Step 2**

- 9 The balance of 363,841 Exchange parent emails was compared to emails that already existed in the BSFf Processing Workspace using the Manual Custom Hash. The analysis was able to match an additional 137,339 Exchange emails and brought the overall population to circa 226,502 documents.

#### **Step 3**

- 10 KPMG then conducted an analysis using just the Message ID in the BSFf Processing Workspace. This matched a further 2,787 Exchange emails.

#### **Step 4**

- 11 The remaining Exchange emails of 223,715 were then analysed for textual similarity with other emails in the BSFf Processing Workspace. Only an additional 29 emails were able to be matched at 100% textual similarity.

### **Step 5**

- 12 For step 5, KPMG widened out the analysis to cover the POHIT Processing Workspace. The analysis of the Relativity Processing Duplicate Hash was able to match a further 1,044 Exchange emails, reducing the remaining balance to 222,642.

### **Step 6**

- 13 KPMG conducted an analysis using just the Message ID in the POHIT Processing Workspace. This matched a further 20,772 Exchange emails. Given the very large size of the POHIT Processing Workspace (19.4 TB), Message ID was selected as a priority for analysis; this required less manual data extraction than the Custom Hash and was expected (based on the findings from steps 2 and 3 above) to generate very similar numbers of matches. Also, the time period for POHIT Processing Workspace matching was limited to 1 January 2010 to 31 December 2015 to reduce machine time for the analysis whilst still covering over 90% of the emails from Exchange in this date range.

### **Step 7**

- 14 The remaining Exchange emails of 201,870 were then analysed for textual similarity with other emails in the POHIT Processing Workspace, also in the date range 1 January 2010 to 31 December 2015. Only an additional 482 emails were able to be matched at 100% textual similarity, leaving 201,388 unmatched Exchange emails.



### **Step 8**

15 KPMG's analysis was extended to cover GLO, also using the date range 1 January 2010 to 31 December 2015. In the absence of Relativity Processing Duplicate Hash, the MD5Hash of the remaining Exchange emails was compared to emails that already existed in the GLO Workspace. Zero matches were found.

### **Step 9**

16 The Exchange emails of 201,388 were then analysed for textual similarity with other emails in the GLO Workspace, also using the date range 1 January 2010 to 31 December 2015. Zero matches were found at 100% textual similarity.

### **Step 10**

17 Finally, KPMG widened out the analysis to cover the fourth main workspace, the CCRC Processing Workspace. The analysis of the Relativity Processing Duplicate Hash was able to match only 1 Exchange email.

### **Steps 11 and 12**

18 Zero matches were found for Manual Custom Hash but an analysis using just the Message ID in the CCRC Processing Workspace matched a further 39 Exchange emails.

- 19 The residual balance of Exchange parent emails was therefore 201,348 after the 12 stage deduplication analysis process. A total of 190,427 of the Exchange emails were matched using the techniques above, which represents 49% of the starting Exchange population.

### **BSFf Sampling**

- 20 In order to perform a level of quality control over the main KPMG deduplication matching techniques set out above, KPMG created six sample sets for BSFf to review, which were designed to cover the four techniques across the workspaces with the largest identification of duplicate, or materially duplicative, documents:
- 21 Sample 1: BSFF Processing – Manual Custom Hash (Step 2): This consisted of a comparison of 100 Exchange and 100 Mimecast documents in the BSFF Processing Workspace.
- 22 Sample 2: BSFF Processing - Message ID (Step 3): Sample 2 consisted of a comparison of 100 Exchange and 100 Mimecast documents in the BSFF Processing Workspace that have the same message ID and did not match on Manual Custom Hash.
- 23 Sample 3: POHIT Processing - Message ID (Step 6): Sample 3 consisted of a comparison of 100 Exchange documents and 100 Mimecast documents in the POHIT Processing Workspace that matched the same Message ID.

- 24 Sample 4: POHIT Processing - Processing Duplicate Hash (Step 5): Sample 4 consisted of a comparison of 100 Exchange documents and 100 Mimecast documents in the POHIT Processing Workspace that matched based on the Processing Duplicate Hash.
- 25 Sample 5: POHIT Processing - TND (Step 7): Sample 5 consisted of a comparison of 100 Exchange documents and 100 Mimecast documents in the POHIT Processing Workspace that matched as 100% TND duplicates.
- 26 Sample 6: GLO – TND (Step 9): There were zero matches found at 100% textual similarity in GLO. Sample 6 therefore consisted of a comparison of 100 Exchange documents and 100 emails in the GLO Workspace that matched with a 95-99% textual similarity.
- 27 Sample 7 is also TND: comparing Exchange with emails in the GLO Workspace. The sample has selected the GLO document at 100% and the Exchange doc as 95-99%.

#### **Findings from BSFf sampling**

- 28 BSFf are finalising their work with the samples. BSFf reviewed an Exchange document against the potential equivalent Mimecast document using the Relativity document compare function. Preliminary views are that sample methods 1-5 identify documents in Exchange which have no apparent differences with the potential equivalent in Mimecast or have differences identified by the Relativity compare function which appear to be differences in how email addresses, URLs, images or email headers

are rendered (albeit in such a way that any relevant text remains visible).  
Post Office Cyber and KPMG will consider the observations for relevance (if any) to their work. Post Office will write to the Inquiry should it consider that any of the duplication analysis methods are suitable for Post Office's data.

- 29 The names of the relevant custodians are not given here to avoid going into specific details of s21(03)

**APPENDIX 3: CURRENT EDM STRUCTURAL REVIEW ACTIVITIES**

<b>Category</b>	<b>Activity Description</b>	<b>Anticipated Completion Date of Current Activity</b>
Electronically stored information (ESI) excluding eMedia	Identification of the extent of Post Office's live data universe	Identification work is currently anticipated to complete before the next disclosure hearing. Preservation, collection and processing work is possible beyond that.
	Activities relating to the Mimecast/Exchange issue	Work to understand the issue is now largely complete insofar as is practicably achievable. Further work is required to check/validate preservation steps taken to date. Collection and processing work has commenced, but the full scope of work

		remains to be established. Further details are contained in the body of my witness statement.
	Validation of historic preservation activity across other ESI data sources	Post Office is hopeful that this work will be completed within the next 2 months but will update the Inquiry once a more certain timeframe can be established or if that changes.
	Review of ESI received in the past from third parties to establish whether further collection of ESI is required	Post Office hopes that this work will be completed within the next 2 months but will update the Inquiry once a more certain timeframe can be established or if that changes.

ESI stored on eMedia	Investigation of the 5 servers and 12 back-up tapes located at Chesterfield	Imaging of the servers and back up tapes is significantly progressed and work is under way to understand the data stored on them. This will necessarily be an iterative process, but Post Office anticipates writing to the Inquiry before the next disclosure hearing with a substantive update.
	Confirmation of understanding relating to the NAS Drive data (further to BSF's letter to the Inquiry dated 17 November 2023)	Post Office hopes that this work will be completed before the next disclosure hearing but will update the Inquiry if that changes.

	Validation of historical assumptions about the likely probative value of eMedia	Post Office hopes that this work will be completed within the next 2 months but will update the Inquiry once a more certain timeframe can be established or if that changes.
	Review of custodian disclosure questionnaires to establish whether further collection of eMedia is required	Post Office hopes that this work will be completed within the next 2 months but will update the Inquiry once a more certain timeframe can be established or if that changes.
Hard copy documents	Post Office enhanced self-certification process supplemented by an on-site search by Innovo Law	BSFf is not involved in this work but understands from Post Office that it is likely to complete within the next few



		weeks and before the next disclosure hearing.
	Post Office reindexing of hard copy documents stored in Oasis archives	BSFf is not involved in this work but understands from Post Office that it is likely to complete within the next few weeks and before the next disclosure hearing.
	Review of custodian disclosure questionnaires to establish whether further collection of hard copy documents is required	Post Office hopes that this work will be completed within the next 2 months but will update the Inquiry once a more certain timeframe can be established or if that changes.

---

#### **APPENDIX 4: De-duplication**

- 1 This appendix refers to productions of documents made by or on behalf of Post Office as the “**Productions**”, regardless of whether they were made by BSFf, HSF or P&P (although some of the options below are easier across just BSFf productions and more difficult across BSFf, HSF and P&P productions collectively).
- 2 This appendix explains processes that the Inquiry, aided by its eDisclosure provider and information rights management team, might wish to consider to assist it in identifying potential exact or near duplicates of documents produced by Post Office to the Inquiry. Doing so might help the Inquiry prioritise documents for review and analysis (e.g., where near or exact duplicates appear in different families).
- 3 This appendix is based on information provided to BSFf by KPMG. I understand from KPMG that one process may be used on its own or in combination with other processes specified here or that the Inquiry identifies. I have sought to note below where a process and/or data might be limited in its use to a production by BSFf and such advantages and disadvantages known to KPMG and/or BSFf. Should the Inquiry require further technical information, I would need to defer to KPMG.
- 4 While KPMG has provided this information for the purposes of assisting me in preparing this statement and, ultimately, of assisting the Inquiry, KPMG has asked me to make clear that neither KPMG nor BSFf know how the

Inquiry is managing the documents produced by or on behalf of Post Office.

This appendix is not legal or technical advice to the Inquiry.

- 5 This appendix is separated into two sections:
  - (a) options using the existing data available in the Productions; and
  - (b) additional options available using data from the original documents.
  
- 6 A result of applying one or more options is that the Inquiry might have questions about why specific documents were provided as duplicates or apparent duplicates and/or in a specific format (e.g., image, placeholder). BSFf and Post Office will endeavour to answer any such questions the Inquiry might have (including following such engagement with HSF, KPMG or P&P as is required).
  
- 7 I understand from KPMG that, if the Inquiry and its eDisclosure provider adopt the textual near duplicate analyses (TND) discussed below it would be necessary to re-run that process or combination of processes each time Post Office produces additional documents if the Inquiry wishes to ensure that the near/exact duplication analysis is fully up to date (as the TND score may change as additional documents are added to the TND data set).

**Part A: Options using the existing data available in the Productions**

- 8 I understand the following steps may be used on their own or in combination with each other (not necessarily in this order). These options are available to the Inquiry presently (and its disclosure provider) with the

data produced to it by Post Office. KPMG is available to assist with any questions if needed or take the steps below and provide the necessary information to the Inquiry's disclosure provider. We anticipate that the Inquiry may have its own view on which, if any, of the following it is comfortable with being used and in what order.:

- (a) MD5 Hash
- (b) TND analysis
- (c) Manual custom#

9 MD5 Hash: Identify forensic duplicates using the MD5 Hash field:

- (a) Summary: use the MD5# field provided in the Productions to identify duplicative documents within the document population.
- (b) Advantage(s):
  - (i) The MD5 Hash field provides a unique and forensically accurate digital fingerprint of a document. Deduplication using MD5# therefore gives the highest level of confidence in identifying exact documents. This # field is representative of the document in its native format and allows for comparison of exact duplicate versions of documents disclosed in different formats, such as a redacted version of a document versus a native version of a document.

- (ii) Identifies exact duplicate documents amongst the Productions where they appear in different family groups.
  - (c) Disadvantage(s) / Points to note: Documents that appear the same to a reviewer but have small textual differences will have different MD5 Hash values. and will therefore not be identified as duplicates. Further, identical documents that are processed using different eDisclosure software applications may also have differences in their MD5 Hash values.
- 10 TND analysis: perform TND analysis using the extracted text field:
- (a) Summary: This process analyses the textual content of all documents and is used to determine a percentage similarity across documents (e.g., 90-100%) and group similar documents together based on textual content. In addition, the Inquiry may find it useful if its eDisclosure provider builds an analytics index using extracted data in all Productions so that the Inquiry can run TND across all BSFf Productions.
  - (b) Advantage(s): the Extracted text of documents is included for all Productions and TND analysis is available in most review platforms, for example, Relativity which is used by the Inquiry's eDisclosure provider Anexsys. At 100% TND, documents can be identified where the textual content of 2 or more documents is the same (even though the metadata may not be).

(c) Disadvantage(s):

- (i) TND is not a forensically accurate way to identify exact duplicate documents. There may be small textual differences between duplicate documents because of how those documents were processed by different disclosure processing software having no substantive impact on the document but making them appear as textually different.
- (ii) I should point out that TND below 100% should be used with particular caution because small textual differences may also be substantive (for example, a different draft adds only the word “not” to a relevant sentence).
- (iii) TND may not be effective for placeholder and/or redacted documents.
- (iv) TND analysis may also group similar placeholder documents together, even though the underlying originals will be different.

11 Create a manual custom hash:

- (a) Summary: using a concatenation of fields such as Email From, Email To, Email CC, Email BCC, Unified Title and Sort Date/Time, it might be possible to identify similar emails. I understand some of these fields might need to be cleaned, for example, removing seconds from the Sort Date/Time because emails may have a time

sent/received recorded fractionally differently in Post Office source systems for the sender and recipient.

- (b) Advantage(s): The fields listed are available in the Productions provided to the Inquiry and can be concatenated together to approximate how similar documents (whether a parent email or attachment) can be grouped together to identify potential duplicates that share the same fields.
- (c) Disadvantage(s): the resultant manual custom hash has a lower degree of accuracy compared to methods listed above. The method's accuracy will depend on the data quality and consistency of the produced documents' metadata.

12 The deduplication methods identified above would be at a document level.

**Part B: Additional options available using data from the original documents (i.e., with additional assistance from the Post Office)**

13 The following steps may be used on their own or in combination with each other (not necessarily in this order). I understand "original documents" to mean the documents as collected and processed by KPMG and which may be different to the format in which they are produced (e.g. with redactions and consequently reduced extracted text). These are steps which would require KPMG's assistance. I anticipate that the Inquiry may have views on which, if either, of the methods are used and, if both, the order.

- (a) develop custom hash for analysis; and/or
- (b) perform TND analysis using the extracted text field from the original data.

14 Develop Custom has for document analysis:

- (a) Summary: KPMG have developed an approximate custom hash value for top level emails using the Message ID (this was not provided in the Production loadfiles), Unified Title and Sort Date/Time. BSFf are testing samples of documents to confirm the effectiveness of this method (based on the Exchange vs. Mimecast data) and will update the Inquiry separately once the testing is complete.
- (b) Advantage(s): KPMG can identify the Message ID in the BSFf and, subject to additional time, for HSF review Relativity workspaces that accommodated the original production(s) and can generate loadfiles that can be shared with the Inquiry as additional data to generate the custom hash themselves or KPMG can calculate the hash and provide that to the Inquiry. The custom hash will be static per document, thus once provided to the Inquiry with respect to a Bates numbered document, this information will not change.
- (c) Disadvantage(s): The custom hash method is not a 100% forensically accurate technique and is subject to potential inaccuracy compared to MD5 Hash analysis. If the Inquiry wishes to



adopt this method, all future Productions sent to the Inquiry will include the Message ID field (or include the custom hash). Further, KPMG will need to produce a custom hash for all productions to date.

- 15 Perform TND analysis using the extracted text field from the original data:
- (a) Summary: This process analyses the textual content of original documents that were produced to the Inquiry to determine a percentage similarity and group similar documents together based on textual content. KPMG have provided this data as part of the Additional Loadfile for BSFf Productions.
  - (b) Advantage(s): The TND Analysis can be shared with the Inquiry such that they have access to the same information as KPMG. The Percentage Similarity e.g., between 90% - 100% similarity allows flexibility on the documents to consider as duplicates (or very similar).
  - (c) Disadvantage(s): The TND analysis needs to be updated for any newly produced documents will require TND to be run again and the resultant analysis per document could potentially be different. Thus the analysis shared with the Inquiry would need to be updated each time. Further, it is likely to take significant time to conduct this across all productions because it would require all productions to be consolidated into one workspace for TND analysis to be conducted.

**Index to First Witness Statement of Christopher Michael Jackson**

<b><u>No.</u></b>	<b><u>URN</u></b>	<b><u>Document Description</u></b>	<b><u>Control Number</u></b>
1	WITN10810101	Letter from BSFf to the Inquiry dated 13 October 2023	WITN10810101
2	WITN10810102	Letter from BSFf to the Inquiry dated 16 October 2023	WITN10810102
3	POL00126339	Letter from BSFf to the Inquiry dated 1 September 2023	POL00126339
4	WITN09950100	Witness Statement of Gregg Rowan dated 23 August 2023	WITN09950100
5	POL00114170ds	First Interim Disclosure Statement dated 27 May 2022	POL-0113558
6	WITN10810103	Letter from Post Office to the Inquiry dated 10 September 2021	WITN10810103
7	POL00118164ds	Witness Statement of Ben Foat dated 21 June 2023	POL00118164ds

<b>8</b>	WITN10810104	Letter from HSF to the Inquiry dated 15 October	WITN10810104
<b>9</b>	POL00298235	Email dated 14 August 2013 3.37pm	POL-BSFF- 0136285
<b>10</b>	POL00298236	Letter to Post Office dated 2 August 2013	POL-BSFF- 0136286
<b>11</b>	POL00124516	Letter from HSF to the Inquiry dated 18 August 2023	POL00124516
<b>12</b>	WITN10810105	Letter from BSFf to the Inquiry dated 6 October 2023	WITN10810105
<b>13</b>	WITN10810106	Letter from HSF to the Inquiry dated 20 October 2023	WITN10810106
<b>14</b>	WITN10810107	Email from Inquiry to BSFf dated 1 November 2023 9.27am	WITN10810109
<b>15</b>	POL00165906	Letter from BSFf to the Inquiry dated 2 November 2023	POL00165906

<b>16</b>	POL00000657	GLO EDQ (WBD) dated 6 December 2017	VIS00001671
<b>17</b>	POL00142261	P&P DMD dated 19 August 2020	POL-0143530
<b>18</b>	POL00039560	P&P DMD (Annex 1) dated 19 August 2020	POL-0036042
<b>19</b>	POL00142414	P&P Addendum dated 13 January 2021 (erroneously marked 2020)	POL-0143646
<b>20</b>	WITN10810108	P&P Addendum Annex dated 13 January 2021	WITN10810108
<b>21</b>	WITN10810109	P&P Second Addendum dated 19 December 2022	WITN10810109
<b>22</b>	POL00114173ds	Second Interim Disclosure Statement dated 18 October 2022	POL-0113561
<b>23</b>	POL00114176ds	Third Interim Disclosure Statement dated 30 November 2022	POL-0113564

<b>24</b>	POL00114177ds	Fourth Interim Disclosure Statement dated 12 January 2023	POL-0113565
<b>25</b>	WITN10810110	Letter from BSFf to the Inquiry dated 10 November 2023	WITN10810110
<b>26</b>	WITN10810111	Letter from BSFF to the Inquiry dated 17 November 2023	WITN10810111
<b>27</b>	WITN10810112	Email from POL to the Inquiry dated 16 October 2023	WITN10810112
<b>28</b>	WITN10810113	Letter from the Inquiry to BSFf dated 09 October 2023	WITN10810113
<b>29</b>	WITN10810114	Letter from BSFf to the Inquiry dated 11 September 2023	WITN10810114
<b>30</b>	WITN10810115	Letter from HSF to the Inquiry dated 14 August 2023	WITN10810115
<b>31</b>	WITN10810116	Letter from HSF to the Inquiry dated 25 August 2023	WITN10810116

<b>32</b>	WITN10810117	Letter from BSFf to the Inquiry dated 24 November 2023	WITN10810117
<b>33</b>	WITN10810118	Letter from BSFf to the Inquiry dated 15 December 2023	WITN10810118
<b>34</b>	WITN10810119	Letter from BSFf to the Inquiry dated 15 December 2023	WITN10810119

Witness Name: Christopher Michael Jackson

Statement No.: WITN10810100

Dated: 19 December 2023

**POST OFFICE HORIZON IT INQUIRY**

---

**EXHIBIT TO THE FIRST WITNESS STATEMENT OF CHRISTOPHER  
MICHAEL JACKSON**

---