IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MASSACHUSSETTS

ALEXANDER STYLLER, INTEGRATED COMMUNICATIONS & TECHNOLOGIES, INC., JADE CHENG, JASON YUYI, CATHY YU, CAROLINE MARAFAO CHENG, PUSHUN CHENG, CHANGZHEN NI, JUNFANG YU, MEIXIANG CHENG, FANGSHOU YU, and CHANGHUA NI, Plaintiffs, vs. HEWLETT-PACKARD FINANCIAL SERVICES COMPANY, HEWLETT-PACKARD FINANCIAL SERVICES (INDIA) PRIVATE LIMITED, HP INC., HEWLETT PACKARD ENTERPRISE COMPANY, and DAVID GILL Defendants.	Civil Action No. 1:16-CV-10386 (LTS) **Hon. Leo Sorokin**

REPORT AND RECOMMENDATION NO. 1 OF COURT-APPOINTED DISCOVERY EXPERT REGARDING PLAINTIFF's PRODUCTION OF 775,000 PAGES

Alexander Styller, Integrated Communications & Technologies, Inc. ("ICT"), Jade Cheng, Jason Yuyi, Cathy Yu, Caroline Marafao Cheng, Pushun Cheng, Changzhen Ni, Junfang Yu, Meixiang Cheng, Fangshou Yu and Changhua Ni (collectively, "Plaintiffs") and Hewlett-Packard Financial

Services Company, Hewlett-Packard Financial Services (India) Private Limited, HP Inc., Hewlett Packard Enterprise Company and David Gill (collectively, the "Defendants") have disputed whether Plaintiff's production is deficient in that it contains many documents that are not relevant to this matter. On September 13, 2019, this Court appointed Tara Emory of Driven, Inc. as Discovery Expert to assist the court in resolving this issue. Dkt. 232. The Discovery Expert and the parties have worked together successfully, and the issue is now resolved.

Background

Plaintiffs represent that emails from individual plaintiffs Cheng, Yuyi and Yu are unavailable because those emails were not migrated with other data from ICT's servers in 2014, and were subsequently destroyed. In addition, these former employees' computers were seized by Chinese officials, and then later disposed of in China. Stipulation Regarding Electronic Discovery Protocol, Dkt. 183 ("Discovery Protocol"). On a September 16, 2019 call ("9.16.19 Call") with the Discovery Expert and the parties' counsel, Plaintiffs' counsel Mr. Dimitry Joffe explained that because some evidence was not available, Plaintiffs tried to find as many alternate data sources as possible, in order to include as much relevant information as possible. For example, Mr. Joffe explained that one computer backup

contained a prior backup, which included ten computers' data. All were loaded into Plaintiffs' review database. After collecting data, Plaintiffs then removed system files, applied search terms that had been negotiated with Defendants (which were renegotiated several times to narrow the results), and produced the resulting set of 227,425 documents, which totals 791,141 pages. 9.16.19 Call.

The parties agree that as a result of the broad collection from many sources, and culling (narrowing the set) only by keyword, this production included many documents that are not responsive to discovery requests. In light of the many documents that are not responsive and are included in Plaintiffs' production, Defendants' counsel, Mr. Paul Saso, argued that Plaintiffs' production would be overly burdensome to review.

To separate responsive and not responsive documents, Plaintiffs proposed to review, and began the process of reviewing, a spreadsheet that listed metadata for documents, such "as the subject matter, sender, recipients, and date." Joint Report, Dkt. 206 ("Joint Report"), at 3. In the Joint Report, Defendants objected to this process, and proposed alternatives.

Although the Court has not ruled on Plaintiffs' and Defendants' proposals, Mr. Joffe completed the review, almost entirely using a chart with metadata. The review did not include the full text of documents, although for

certain documents, Mr. Joffe did review the full text in the Plaintiffs' review database. Mr. Joffe explained that he used the chart to sort by sender and use metadata to identify spam. With this process, he represented he was able to identify 54,842 responsive documents, and 172,583 not responsive documents ("Null Set"). 9.16.19 Call.

Regarding the set Plaintiffs identified as responsive, Defendants' counsel has accepted this set as adequate, though imperfect. However, Defendants disputed the sufficiency of Plaintiffs' production of the Null Set, arguing that the Null Set included documents that were responsive and significant for Defendants' case. Plaintiffs represented that the Null Set had few responsive documents, and those would not be significant in light of the documents already included in the set identified as responsive. Id. On a September 17, 2019 call ("9.17.19 Call"), the Plaintiffs and Defendants stated they had previously agreed that Plaintiffs would not use documents in the Null Set as evidence for their case, but that Defendants could use these documents as evidence to prove their own case.

Plaintiffs' Process and Obligations

Plaintiffs' discovery efforts included two atypical processes, which contributed to Defendants' concerns. First, Plaintiffs included large amounts

of data that could have been culled based on sources and location.[1] Overinclusive collection can dilute the prevalence of relevant documents, making them harder to find. Overinclusiveness of collected data is not a problem as long as efforts are made later to cull out the unneeded data. Often, such culling is done by narrowing sources based on custodial interviews, sampling results, and other search techniques to eliminate data from nonresponsive sources and locations. Here, Plaintiffs did not undertake any efforts to narrow their overinclusive production, other than the spreadsheet review.

Second, Plaintiffs' spreadsheet review was an atypical way of identifying responsive and nonresponsive documents. On one hand, some strategies utilized by Mr. Joffe in the spreadsheet review are functionally equivalent to culling strategies commonly executed inside review databases, such as excluding emails from internet domains known to send spam. Mr. Joffe put much effort into his process, and also reviewed some documents in full. On the other hand, while a metadata-only review may be used to eliminate spam, it is not generally a reliable way to identify responsive documents. Other than exclusions of clearly not responsive documents based

[1] The fact that the search was overinclusive, and likely included sources without relevant data, does not mean that it could not also have been underinclusive and potentially omitted significant data sources.

on metadata, a text review is generally necessary to test, develop, and implement any search strategy. As described below, Plaintiffs were unable to explain or recreate Mr. Joffe's spreadsheet review search strategy to show how a metadata-only review could have been effective.

However, despite Plaintiffs' process appearing potentially deficient, sampling has shown that Plaintiffs' identification of the Null Set was mostly accurate, as described below.

The Discovery Protocol states that in order to cull documents that are not relevant, "the parties may filter" documents by using search terms, narrowing custodians, and removing duplicates. It does not provide any steps that must be taken to eliminate documents that are not relevant to the matter.[2] Discovery Protocol at 13-14. Therefore, without any additional obligations in the Discovery Protocol, the standard of reasonableness of Plaintiffs' production arises only from the Fed. R. Civ. P. 26(g) obligation to make a "reasonable inquiry" before certifying that their production complies with discovery requirements. The signing attorney certifies that the discovery response is "(ii) not interposed for any improper purpose, such as to harass, cause unnecessary delay, or needlessly increase the cost of

[2] Examples of steps that could be taken include a culling by metadata (such as email sender domains), responsiveness review, and/or application of Technology Assisted Review ("TAR").

litigation; and (iii) neither unreasonable nor unduly burdensome or expensive, considering the needs of the case, prior discovery in the case, the amount in controversy, and the importance of the issues at stake in the action."

Discovery Expert Process

At the Discovery Expert's request, both parties took separate samples of 385 documents from the Null Set, representing a statistical sample with a 95% +/- 5% confidence level[3] for the relevance of the Null Set. In the interests of transparency, each side disclosed which documents were in their random sample, which documents were identified as responsive, and the nature of the responsive documents.

Plaintiffs found seven relevant documents, and stated one was responsive but duplicative of other documents identified as responsive, and six were marginally relevant to the matter but not responsive to any document requests. Defendants' Null Set sample similarly yielded seven responsive documents. Plaintiffs' counsel deemed all seven of these to be marginally relevant. Defendants deemed as highly relevant the documents Plaintiffs referred to as marginally relevant and not responsive. These documents related to H3C equipment from sources other than Defendants.

[3] This means there is a 95% chance that the responsiveness for each sample will be with be within 5% of the true responsiveness of the Null Set.

Also at the Discovery Expert's request, Plaintiffs' counsel prepared a list of exclusionary searches used by Mr. Joffe to identify spam. The list contained email addresses that Plaintiffs asserted could be searched to remove spam, and would have excluded 66,265 documents. However, as discussed on the 9.17.19 Call, while all of the email addresses were contained in the Null Set documents, some of the email addresses (some of which belong to parties in the case) were also contained in responsive documents. Therefore, the list could not be used to exclude any documents, and Mr. Joffe conceded he could not recreate the searches he had used to generate Responsive and Not Responsive designations during his review using a chart. 9.17.19 Call.

Based on the samples from both Defendants and Plaintiffs, 7/385, or 1.8% of the Null Set is responsive.[4] This 1.8% responsiveness of the Null Set is known as an Elusion Rate. This Elusion Rate is low, meaning identification of additional responsive documents potentially carries a large burden. However, while a low Elusion Rate weighs against additional searches for responsive documents, the burden of such searches should be considered in light of quantity and quality of the missed responsive documents.

[4] Taking their samples together, the confidence level for the combined sample is 95% +/- 3.5%.

First, an Elusion Rate of 1.8% in 172,583 documents means that an estimated 3,106 responsive documents remain hidden in the Null Set. The 54,842 responsive set was not sampled because the parties did not dispute its sufficiency. Even assuming this set was only 50% responsive, with only 27,421 responsive documents, this would be a 90% recall, meaning that 90% of all responsive Plaintiffs' documents containing the stipulated keywords have been identified in Plaintiffs' production.[5] 90% is generally accepted in discovery practice as a high recall rate.

Second, the quality of missed documents should be considered. In this case, the missed documents largely related to H3C equipment from sources other than Defendants ("Disputed Issue Documents"). The testing revealed that Plaintiffs deemed the Disputed Issue Documents to be not responsive, while Defendants claim those documents are important.

Based on the sample results, the Discovery Expert recommended that Plaintiffs design a search, with adequate testing to ensure its effectiveness, to identify the Disputed Issue Documents. While maintaining these documents are not responsive, Plaintiffs agreed to design the search in order to resolve the discovery dispute. 9.17.19 Call.

[5] 27,421 responsive documents in the set identified responsive plus 3,106 responsive documents in the Null Set totals about 30,527 responsive documents. 27,421 identified responsive documents out of (divided by) a total of 30,527 documents is 90% recall.

Plaintiffs and Defendants then cooperated to design and implement a search for Disputed Issue Documents. After some initial discrepancies in results of the searches when executed in Plaintiffs' and Defendants' separate review platforms, the parties had their eDiscovery providers discuss the search on a call, and were able to come to similar results. In the latest iteration of the search designed by Plaintiffs, Defendants' execution of the search yielded 4,077 documents. Defendants reviewed a sample and estimated these documents are 91% responsive as Disputed Issue Documents.[6]

On a September 19, 2019 call, Mr. Joffe and Mr. Saso agreed that their dispute regarding relevancy of the Null Set is now resolved. They also agreed to modify their prior agreement regarding use of evidence, so that Plaintiffs may use as evidence for their own case any of the 54,842 documents Plaintiffs previously identified, as well as any of the 4,077 documents in the Defendants' version[7] of the search capturing the Disputed Issue Documents. Defendants will send Bates numbers of their 4,077 search results to Plaintiffs. By agreement of the parties, Plaintiffs will not use as evidence in their case documents that are not in the previously identified

[6] Given margins of error, this roughly corresponds to the number of estimated responsive documents remaining in the Null Set.
[7] Small differences remain between the number of results for Plaintiffs and Defendants when executing the search.

responsive documents or in the 4,077 results from the Disputed Issue Documents search. Defendants are free to continue to search, review, and use documents from the entirety of Plaintiffs' production.

Discovery Expert Recommendations

1. By agreement of the parties, Defendants, but not Plaintiffs, may present as evidence documents from the 168,506 Null Set documents not included in the Disputed Issue Documents.

2. The Discovery Expert has noted to the parties that the Clawback provision of the Discovery Protocol contains language that may not sufficiently protect the parties from privilege waiver because it will only apply to disclosures that were "inadvertent," which in some cases can be a contested issue. *Compare Irth Sols. LLC v. Windstream Commc'ns LLC,* No. 2:16-CV-219, 2017 WL 3276021 (S.D. Ohio Aug. 2, 2017) (holding defendant waived attorney-client privilege claims for inadvertently produced documents) *with Winfield v. City of New York,* 15-cv-05236, 2018 WL 2148435 at *4 (S.D.N.Y. May 10, 2018) (explaining importance of Rule 502(d) non-waiver orders in providing predictable, cost-effective protection from inadvertent privilege waiver). It is recommended that the parties enter an additional stipulated protective order under Federal Rule of Evidence 502(d), to apply to disclosures whether inadvertent or otherwise.

/s/ *Tara S. Emory*

Dated: September 20, 2019

Tara S. Emory

Court-appointed Discovery Expert