February 15, 2014

Honorable David G. Campbell, Chair
Advisory Committee on Civil Rules
Committee on Rules of Practice and Procedure Administrative
Office of the United States Courts
One Columbus Circle, NE
Washington, DC 20544

RE: Comments to Proposed Amendments to the Federal Rules of Civil Procedure

My name is Patrick Oot. I am co-founder of the 501(c)(3) non-profit Electronic Discovery Institute (EDI). EDI's core mission is to conduct research in the litigation and technology space as well as to educate judges, lawyers and technologists about technology solutions where they intersect with the practice of law. At EDI's annual leadership summit, we gather hundreds of in-house practitioners, outside counsel, government attorneys, plaintiff's attorneys and judges that volunteer their time to benchmark and discuss emerging issues in law and technology. For a sample of EDI's good work, I have attached a special report from our recent discourse on Rule 37(e) led by EDI's President, Robert Owen.

I am also Senior Special Counsel in the Office of the General Counsel at the United States Securities and Exchange Commission (SEC). I am one of two attorneys responsible for the agency's electronic discovery obligation and have testified twice on behalf of the agency as a 30(b)(6) witness defending the agency's discovery guidelines that I co-authored. While my views and comments are my own and not those of the SEC or federal government, I do bring the unique perspective of a former in-house counsel that is now a government attorney in an office of the general counsel responsible for managing the perpetual problems of a large data holder.[1]

I commend the advisory committee for its hard work and long hours dedicated to a continued effort to provide just, speedy and inexpensive resolution of the law. Our country truly appreciates the sacrifices of personal time made by the committee members and their families.

While I support continued research into more effective limitations on cost, I write in support of the amendments and believe that they are a positive step toward balancing the scales. The amendments will provide litigants with much needed tools to alleviate risk and mitigate significant cost. I have five major points to this comment:

I. Discovery costs hinder a party's access to justice, promote settlement leverage and favor remedies outside the federal justice system.

In 2007, I joined a panel at Georgetown Law, moderated by Arthur Miller, debating whether or not the cost of electronic discovery threatened to skew the justice system. Panelists included a

[1] The Securities and Exchange Commission disclaims responsibility for any private publication or statement of any SEC employee or Commissioner. Statements made express the author's views and do not necessarily reflect those of the Commission, the Commissioners, or members of the staff.

cross section of legal professionals and jurists such as Hon. John Facciola of the District Court of the District of Columbia and Justice Stephen Breyer of the United States Supreme Court. During the discussion I disclosed the millions of dollars in electronic discovery service provider fees my then employer paid on a single large-scale matter.

In response, Justice Breyer commented that, "if it really costs millions to do that, then you're going to drive out of the litigation system a lot of people who ought to be there. They'll go to arbitration …. They will go somewhere where they will write their own discovery rules, and I think that is unfortunate in many ways."[2] Justice Breyer framed the problem that the ambitions of FRCP Rule 1 sought to address: justice should be administered to "secure just, speedy and inexpensive determination of every action and proceeding."[3]

Yet, growth in data volume at large organizations spells friction to these Rule 1 tripartite goals. Discovery costs are massive in federal litigation. While requesting parties argue that storage is cheap, the careful and precise services required to manage litigation data are extremely expensive.

II. Growth in "junk data" will continue to burden producing parties as evidentiary scope creeps across emerging technology and platforms. Limitations on scope and stronger proportionality guidance are crucial to counter rising discovery costs that result from general data growth and over-preservation.

The mere act of storing data, even the unknown or unintentional storage of data, should not create instant preservation obligations. Yet texts, IMs, tweets, voicemail and Facebook posts ("junk data") are fair game for preservation and production; even if they have very little value in an organization's mission.

Today, wholesale data recording technology captures every aspect of our life – right down to online chats that replace water cooler conversation. Explosive growth in new storage platforms, new methods of communication and life-recording technology create futuristic discovery burdens that will only increase over time. We are on the verge of an avalanche as on-the-record technology, such as Google Glass and wearable video cameras, come online. Cameras like these are "small enough to attach anywhere on your person that, thanks to its built in Wi-Fi, allows you to live stream everything to your Facebook wall."[4] We must consider meticulous, thought-out limitations to discovery data presented in our future justice system, or we will simply be overrun by the cost to categorize and produce it. Proportionality is crucial; as are bright line rules wherever possible.

We need not review futuristic technologies to see discovery data growth. An observer can open the daily e-mail inbox of an average corporate employee to see the surging data volumes. The Radicati Group, a market research firm focused on the computer and telecommunications

---

[2] John Bace, *Cost of E-Discovery Threatens to Skew Justice System*, Gartner Report No. G00148170, April 20 2007, Available at https://www.gartner.com/doc/503935?ref=SiteSearch&sthkw=justice%20breyer&fnl=search (last accessed February 15, 2014)

[3] Fed. R. Civ. P. 1

[4] Andrew Hoyle, *Live stream your life in HD to your Facebook friends with the Looxcie HD,* CNET, February 24 2013, Available at http://reviews.cnet.com/digital-camcorders/looxcie-hd/4505-6500_7-35619090 html (last accessed February 15, 2014).

industry, recently reported that in 2013 the average business user sent and received 115 daily e-mail messages and consumed an average of 13.1 MB of storage per day.[5] Figure 1 presents this statistic annually; with the average custodian storing 28,750 e-mails or 3.275 gigabytes of e-mail data, per year.[6]

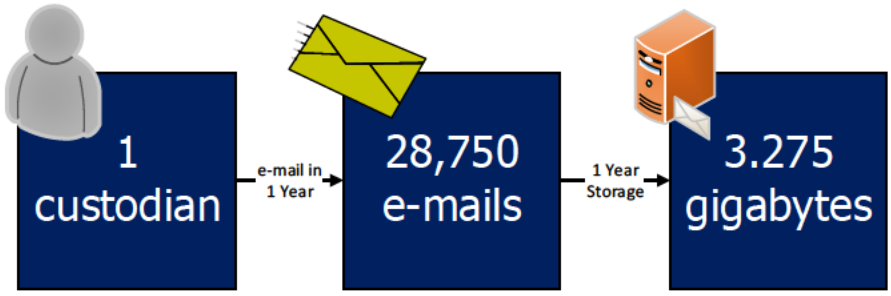

Figure 1: E-mail Sent and Received by a Single User in 1 year

While the Radicati report suggests that e-mail growth will slow to a mere 6% year-over-year increase, slow-moving litigation and perpetual litigation holds on custodians can cause a user's mailbox to swell.[7] With this growth, an organization must maintain an e-mail box sized over 20 gigabytes containing an average of 178,250 e-mails – all before consideration of other sources (loose files on PCs, network drives, SharePoint sites, databases and handheld devices, not to mention paper materials). *See Figure 2*.

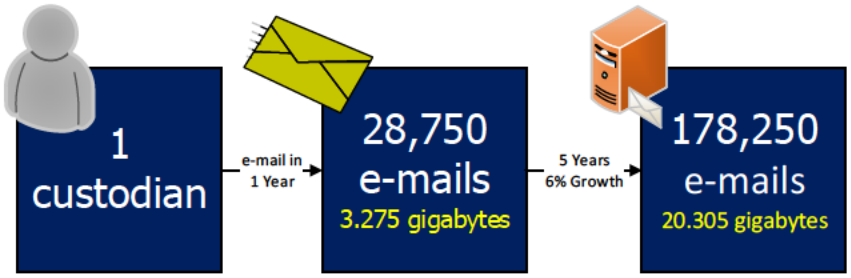

Figure 2: E-mail Retained for a User on Perpetual Litigation Hold for 5 year

While these explosive statistics might seem alarming, consider the heightened demand on an individual employees and start counting messages in your own Inbox – they add up.

Coincidentally, the e-mail projections extracted from the Radicati report are in-line with the all-source custodian data volume of 30.1 gigabytes presented by Jon Palmer, Assistant General Counsel at Microsoft, at the EDI Leadership Summit.[8] Litigants should anticipate 30 gigabytes of user data for the "average user" at a large organization.

---

[5] Sara Radicati and Quoc Hoang, *Email Statistics Report, 2011-2015,* The Radiciati Group, Inc. Available at http://www radicati.com/wp/wp-content/uploads/2011/05/Email-Statistics-Report-2011-2015-Executive-Summary.pdf (last accessed February 15, 2014). *See also*, Openwave Messaging Announces Universal Messaging Suite (TM) Version 9.0, Press Release on Wall Street Journal May 22 2013, Available at http://online.wsj.com/article/PR-CO-20130522-905546 html?mod=googlenews_wsj (last accessed February 15, 2014).

[6] Assuming a 2000 hour work year.

[7] Radicati, *supra* note 5.

[8] *See Proceedings of the 2013 EDI Leadership Summit: At the Crossroads of Bad Faith & Negligence: How Sekisui Shows We Need New Rule 37(e) Adapted from a presentation at the 2013 EDI Leadership Summit* Santa Monica, California (October 2013) page 4. Attached and available online at http://www.ediscoveryinstitute.org/Sekisui.

Without the proposed rules, large data holders will have very little help at the crossroads between sanctions risk and discovery expense. Litigants have two options: either protect against sanction risk by extending preservation to large swaths of tangentially relevant data while incurring massive discovery expenses; or maintain tight destruction schedules and more precise litigation holds to decrease discovery costs but risk potential sanctions for failure to preserve data.

III. Scope of data preservation and production directly correlates to expensive e-discovery data services – let alone review costs. Broad requests and fishing expeditions have significant financial consequences on producing parties.

Costs are rising as a result of data growth. Opinions and surveys are valuable in shaping the debate, but I have always been a proponent of quantifying the cost and volume problems presented by electronic discovery. During my tenure as Director of Electronic Discovery at Verizon, I provided data points from a half-dozen matters to Nick Pace of the RAND Corporation for his report that the advisory committee has already received. The RAND report provides insight into discovery costs at large organizations from 2005-2010.

I highlight the importance of the cost generated by the data deluge in 2014 by correlating the data volume discussed above with publically available sources that list the fees charged by widely used electronic discovery service providers. Generally, most electronic discovery service providers charge their clients under a volume-driven model – usually per gigabyte, in addition to per-seat software licenses and add hourly rates for their consultants, attorneys and project managers. To help understand these charges, I point to the openly available pricing schedules on The General Services Administration's (GSA) website.[9]

GSA acts as a products and services wholesaler to federal agencies. Other federal entities, local governments and large organizations use GSA pricing as a guide to help understand the market in the procurement process. From GSA's website, its priorities include "using the purchasing power of the federal government, [GSA] will drive down prices, deliver better value, and reduce costs to [its] customer agencies." Arguably, GSA prices on electronic discovery services represent real-world pricing for repeat-customer litigants and provide a reliable tool to investigate litigation focused electronic discovery costs.

The volume data presented by Mr. Palmer, and more broadly by the Radicati Group, both support the notion that large data holders must manage between 20 and 30 gigabytes of data per custodian placed on litigation hold. My goal is to combine this volume evidence with pricing information to illustrate the discovery costs spent for e-discovery of a small 10-custodian matter by linking custodian volume to the fees on a typical GSA electronic discovery contract.[10]

---

[9] *See GSA Mission and Priorities* at http://www.gsa.gov/portal/content/100735 (*last accessed* February 15, 2014).
[10] *See GSA Federal Acquisition Service eLibrary Contractor Listing* at http://www.gsaelibrary.gsa.gov/ElibMain/sinDetails.do?executeQuery=YES&scheduleNumber=36&flag=&filter=&specialItemNumber=51+508 (*last accessed* February 15, 2014). (Three dozen service providers appear under the GSA schedule for litigation support services (36-51-508). For this illustration, I randomly selected contract # GS-25F-0004P. While this was a random selection, I invite the committee to visit the GSA website to concur that the rates for services under GS-25F-0004P are representative of other contractors on the schedule).

For this exercise, I selected the pricing schedule from GSA contract number GS-25F-0004P, the contract for Deloitte Financial Advisory Services LLP ("Deloitte Discovery").[11] From its website, "Deloitte Discovery is one of the industry's largest global discovery services providers, with highly skilled professionals in more than 35 countries, serving multinational corporations, law firms and government agencies in complex litigation and regulatory matters."[12] While I am sure other options exist, I believe Deloitte Discovery's rates for services under the GSA schedule to be representative of the pricing charged to government agencies and large corporations based on the hard work of GSA employees to secure good pricing for its clients. Further, this comment is not an evaluation of the procurement process of any organization. Using a well-known, highly developed organization as a bell-weather should not cause major disruption in the comment.

As stated previously, many litigation services providers charge for their services by either data volumes or flat software licenses. In GSA contract GS-25F-0004P, we see both. For example, as illustrated in Figure 3 below, upon receipt, a provider will decompress and deduplicate files to avoid redundant work; under GS-25F-0004P the price is $272 per gigabyte.[13]

Second, the vendor must extract metadata and text from e-mail and loose computer files and populate a searchable database with the extracted text. This step is required whether a client uses advanced analytics (computer assisted review, predictive coding, etc.) or keyword searching; under GS-25F-0004P, the price is $817 per gigabyte.

Third, counsel would select some documents for human review and would need to access them via an online platform; under GS-25F-0004P a $2543 flat fee per month provides the client with software licenses for 30 individual reviewers and 4 gigabytes of online storage, additional gigabytes are $126 per month and additional user licenses are $50 per month.[14]

Finally, native file documents selected for production must be extracted from the hosted system; under GS-25F-0004P, that price is $752 per gigabyte.[15] As previously mentioned, all these technology services are typically bundled with billable hours from consultants and project managers. Outside counsel feels would also apply to the management of the process.

---

[11] See *GSA Federal Supply Service Authorize Federal Supply Schedule Price List Contract Number: GS-25F-004P* (Page 9) (*last accessed* February 15, 2014). Available at https://www.gsaadvantage.gov/ref_text/GS25F0004P/0MFF92.2R0VDT_GS-25F-0004P_DFASGS25F0004PPRICELISTMODCMA254120413.PDF. http://www.gsaelibrary.gsa.gov/ElibMain/sinDetails.do?executeQuery=YES&scheduleNumber=36&flag=&filter=&specialItemNumber=51+508 (*last accessed* February 15, 2014). Three dozen service providers appear under the GSA schedule for litigation support services (36-51-508). For this illustration, I randomly selected contract # GS-25F-0004P. While this was a random selection, I invite the committee to visit the GSA website to concur that the rates for services under GS-25F-0004P are representative of other contractors on the schedule.

[12] See http://www.deloitte.com/view/en_US/us/Services/Financial-Advisory-Services/Deloitte-Discovery-Financial-Advisory/index.htm (*last accessed* February 15, 2014).

[13] GSA, *supra* note 10.

[14] *Id* at 10.

[15] It should be noted that native file production is more cost effective that image based productions, even so, some documents must be converted to images for redaction purposed. I have not included these image charges given the variability in redaction needs.

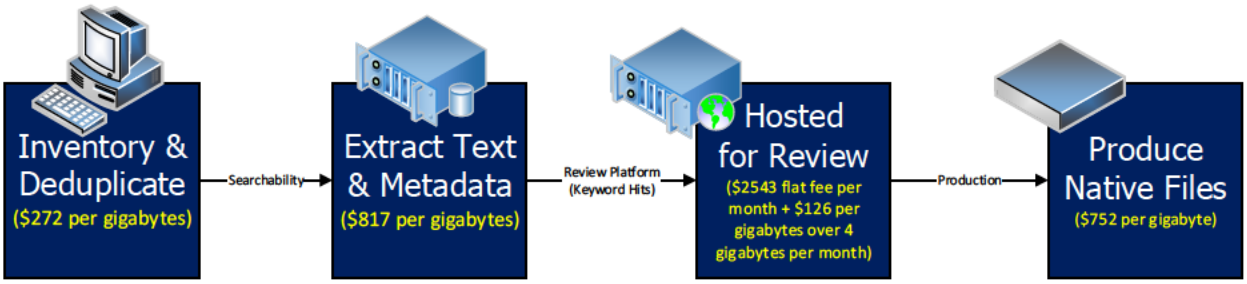

Figure 3: Typical Litigation Services Workflow With Pricing from GSA contract # GS-25F-0004P

Applying the prices from GSA contract GS-25F-0004P to the average 30 gigabyte custodian, inventory and deduplication charges would cost $8160. Data volume reduction using aggressive deduplication can reduce data volume by as much as 40%.[16] Thus, the text and metadata extraction of the reduced 18 gigabytes would cost a client an extra $14,706 for a total of $22,866.[17] Under GSA contract GS-25F-0004P, initial data handling charges for a matter with ten custodians could cost as much as $228,660 without document review.

At this stage in the process, the litigant can cull the data using keyword search terms, date ranges or other filters. The filtered set can be migrated to a hosted platform, with a minimum monthly charge of $2,543 for 4 gigabytes and 30 user licenses. Assuming aggressive culling, so the data set never grows larger than the 4 gigabyte minimum, a matter that lasts just ten months can cost a litigant $25,430 in hosting using the pricing from GSA contract GS-25F-0004P. The data services expense for a ten custodian matter is now at $254,090, without anyone looking at the documents or using predictive coding to determine relevancy.

Litigants must now apply some sort of analysis on the documents for relevancy and privilege. While there are many strategies to analyze the data (keyword searching, computer assisted review, predictive coding, linear review, etc.), all require some human input into software to achieve an outcome that an attorney can attest to. If a responding party is able to aggressively cull the matter using technology to only look at a small sample of 15% of the documents and an outside contract attorney staffing firm would charge $50 per hour for attorney review, as illustrated in figure 4, initial review cost would be $222,812. This does not include management and supervision by a law firm associate, typically $250-$350 per hour.

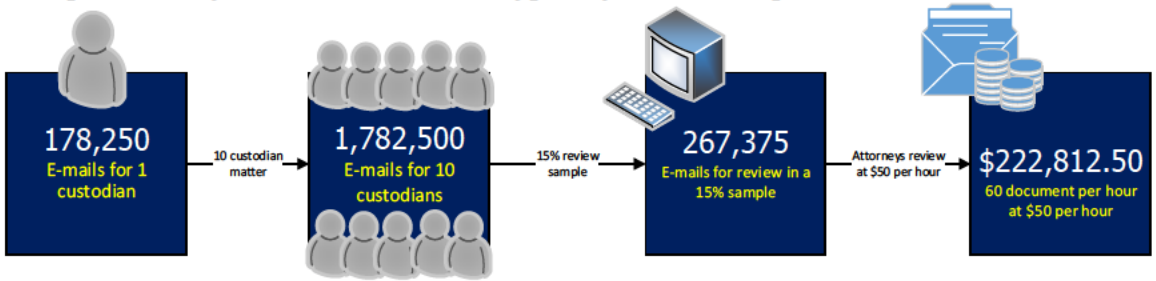

Figure 4: Typical Document Review Pricing on a 15% sample of a ten custodian matter at $50 per hour

[16] See Patrick Oot, et al., *Ethics and Ediscovery Review*, ACC Docket Vol. 28, Issue 1 (Jan/Feb 2010) Pages 46-57. http://www.ediscoveryinstitute.org/publications/acc_docket_ethics, (*last accessed* February 15, 2014);(39.1 percent reduction in data volume by deduplication). *See also* Kroll Ontrack Thought Leadership Team, *3 Deduplication Options: How Do You Choose?* (March 19, 2012) at http://www.theediscoveryblog.com/2012/03/19/3-de-duplication-options-how-do-you-choose/; (*last accessed* February 15, 2014). ("Effectively, de-duplication can reduce the number of documents to be reviewed by as much as 90 percent, and, on average, 30 or 40 percent").

[17] GSA, *supra* at FN 10.

Under this model, adding the cost of litigation support data management for a ten custodian matter to the review costs associated with reviewing a 15% sample of documents would cost a litigant $476,902 without any input from outside counsel.

This sum is supported by outside sources. For example, Kroll Ontrack, an electronic discovery service provider recently reported the results of an advisory board survey that revealed the cost of electronic discovery data services and document review for one gigabyte of data averages at $20,000.[18] Applying the Kroll results to our Radicati report average custodian email data volume, with a 15% review sample suggests a matter cost of $609,200. *See figure 5.*

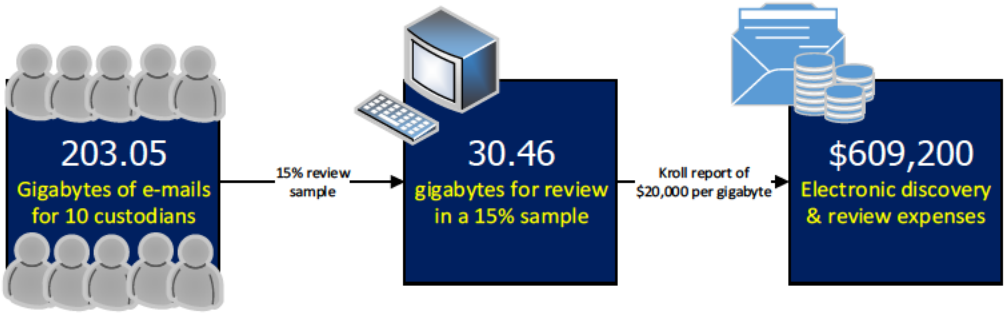

Figure 5: Cost Using Radicati Report Volumes and Kroll Ontrack Survey Results

These results suggest that large organizations see e-discovery costs in a ten custodian matter in the hundreds of thousands of dollars. A significant admission price to enter the justice system. We should all have the same concern that Justice Breyer vocalized seven years ago at Georgetown Law. While data volumes continue to grow, our practice must consider some type of limitations on data entering the electronic discovery process.

I support the new rules, as they will provide a good first step to allow responding parties to tighten their preservation practice as a result of lowering sanctions risk in proposed rule 37(e) while preventing fishing expeditions by narrowing the scope of 26(b)(1) through moving the proportionality requirement of 26(b)(2)(C)(iii) to 26(b)(1) and removing a requesting party's ability to obtain information "reasonably calculated" to lead to the discovery of admissible evidence.

I have spent considerable time thinking about the data problem, I hope these figures promote appropriations from our legislature for continued research by the Federal Judicial Center on how innovation can make our judicial process more cost-effective and fair for both parties in our adversarial system. Not every litigant is well-financed, not every litigant is tech-savy and not every litigant can meet the demands of data-intensive litigation.

IV. Technology is not keeping pace with volume growth; emphasis must be placed on defining and educating attorneys on reasonableness, not tools.

Predictive analytics and computer assisted review (CAR) will be helpful in managing growth, but are by no means the wide-spread panacea to the discovery cost problem that proponents make claim to. While I am a well-known supporter of CAR technology, and have

[18] *See E:discovery, what it means for big data*, at http://www.ediscovery.com/CMS/PDF/infobite-ediscovery-big-data_krollontrack2014.pdf (*last accessed* February 15, 2014).

used it on many matters, I believe it is just one arrow in an emptying quiver of resources available to the reasonable attorney to cost-effectively meet discovery obligations.

I currently lead a research project on this very subject with Pallab Chakraborty of Oracle Corporation, and Professors Gerd Infanger and Peter Glynn from Stanford University. Our research evaluates the cost performance of multiple document review systems. Our initial report reveals great variance in both cost and responsiveness.[19] I believe technology is crucial in solving the problem of data growth, but I have concern about the current motion practice generated from a producing party's choices when conducting their response. Much more effort should be placed on educating parties on how to act reasonably and how good is good enough.

V. The rules must protect reasonable decisions

Finally, and most importantly, we must place more emphasis on the FRCP 26(g) "reasonable inquiry" certification standard.[20] As a legal community we must do a better job at promoting reasonableness. Requesting parties have far too much adversarial oversight into the discovery practices of the producing party, and are demanding a close-to-perfect standard of performance in discovery when the actual standard is a reasonable inquiry. Reasonableness is far from perfection.

I am concerned with requests, responses and court orders that demand "every and all" document(s) produced "completely and entirely" to the requesting party.[21] Certifying a kitchen sink response is unadvisable and arguably impossible in matters with millions of documents. Education initiatives mandated by state bars and law schools could help alleviate this problem. I suggest emphasis on this standard in the committee notes. Perhaps we can avoid a few fire drills.

I appreciate the opportunity to make these comments. I look forward to the continued great work of the committee. Thank you for your consideration.

Sincerely,

Patrick Oot

---

[19] Monica Bay, *EDI-Oracle Study: Humans Essential in E-Discovery*, http://www.lawtechnologynews.com/id=1202642098011/EDI-Oracle-Study%3A-Humans-Essential-in-E-Discovery (*last accessed* February 15, 2014).

[20] *See Mancia v. Mayflower Textile Servs. Co.*, 253 F.R.D. 354, 357 (D. Md. 2008) ("The duty to make a 'reasonable inquiry' is satisfied if the investigation undertaken by the attorney and the conclusions drawn therefrom are reasonable under the circumstances.")

[21] *See Starr International Company v. The United States*, No. 1:11-cv-00779-TCW Docket No. 172, (Fed. Cl., October 22, 2013). ("…so that both parties can prepare for the rest of discovery and ensure that all relevant documents have been produced").