

EXHIBIT 2

Declaration of Josh Woods

IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF NORTH CAROLINA
SOUTHERN DIVISION
No. 7:23-CV-897

IN RE:)
)
CAMP LEJEUNE WATER LITIGATION)
)
This Document Relates to:)
ALL CASES)

DECLARATION OF JOSHUA WOOD

1. I, Joshua Wood, make the following declaration in lieu of an affidavit as permitted by 28 U.S.C. § 1746. I am aware that this declaration will be filed with the United States District Court for the Eastern District of North Carolina, and that it is the legal equivalent of a statement under oath.

2. I am Chief of the Office of Litigation Support (“OLS”), Office Management Programs, Civil Division, United States Department of Justice. OLS provides technical support to case teams from all the litigating offices within the Civil Division, including the Environmental Torts Litigation (“ETL”) office.

3. I have 29 years of litigation support experience, including experience providing litigation support to some of the largest civil litigation that the Department of Justice has handled over the past 20 years. I have held my current position for three months. Previously, I was the Director of OLS for a little over 11 years. Prior to that I was Chief of Information Technology for OLS for five years. I have worked on Native American trust account

litigation, complex tobacco litigation, Hurricane Katrina litigation, Deepwater Horizon litigation, and Residential Backed Mortgage Securities litigation, as well as thousands of other litigations involving the Civil Division and Federal government agencies. I have significant knowledge and experience with Federal agencies' capabilities to respond to litigation, and the role that Congressional appropriations play into that response. Litigation response involving Federal agencies differs from corporate litigation due to funding considerations and constraints.

4. In my current role as Chief of OLS, I manage staff that provides litigation support to the Civil Division, and act as a liaison with agency staff regarding an agency's preparation for litigation as well as its management of existing litigation requirements. I have 10 Government staff who oversee the management of the litigation support contract for the Civil Division, known as MEGA 5, which supplies litigation contractor support services to the Civil Division, as well as other parts of the United States Government at large. My staff oversee all cases that require e-discovery work. This work can take many forms, from data processing all the way through full case support with paralegals, law clerks, e-discovery experts, and presentation specialists for court hearings and trials.

5. In the course of carrying out these duties, I am involved in and oversee the ongoing technical support for ETL's Camp Lejeune Justice Act ("CLJA") case team. A significant part of this oversight involves ensuring that OLS has adequate resources—in terms of time, manpower, and technical tools and capabilities—to meet the needs of the case team in a timely manner. This declaration explains the necessary OLS resources to collect, process, search, review, and produce the proposed discovery in the CLJA litigation. At this early stage, it is difficult to precisely predict how a search and review will progress. However, I

have made reasonable estimates that quantify the processes, timeframes, and costs based on currently available information.

I. E-DISCOVERY OVERVIEW

6. OLS follows industry-standard e-discovery practices. E-discovery is commonly conceptualized using a framework called the Electronic Discovery Model (“EDRM”). More information on the EDRM is available online.¹

7. E-discovery under the EDRM proceeds in several phases: (1) Identification; (2) Preservation; (3) Collection; (4) Processing; (5) Review; (6) Analysis; (7) Production; and (8) Presentation. While OLS can and does advise on each of these phases, most of our work focuses on ensuring complete and usable collection; processing data using specialized technical tools; supporting case teams’ review and analysis (which can include leveraging search tools); and production of data to opposing parties.

II. SCOPE OF FEDERAL INFORMATION IMPLICATED IN CLJA LITIGATION

8. The CLJA litigation involves collecting both custodial and non-custodial discovery from multiple federal agencies.² I understand that those agencies include:

- a. United States Marine Corps
- b. United States Navy
- c. Environmental Protection Agency
- d. Department of Veterans Affairs
- e. Agency for Toxic Substances and Disease Registry (a sub-agency of the

¹ See <https://edrm.net/edrm-model/current/>.

² For purposes of this declaration, “custodial data” refers to data collected from individual agency employees from sources such as Outlook and One Drive. “Non-custodial data” refers to data that is not closely associated with an individual employee and collected from sources such as shared drives and archived federal records. Both categories are part of “Other Discovery” as defined by Case Management Order No. 8, ECF No. 52.

Centers for Disease Control and Prevention of the Department of Health and Human Services)

f. National Archives and Records Administration.

9. Federal agencies are generally responsible for collecting their own data for use in litigation. The tools and methods these agencies use to complete such collections varies significantly depending on the agency, necessitating coordination from OLS to ensure that data is collected in a defensible and usable manner.

III. SCOPE OF DISCOVERY

A. COLLECTION

10. As part of OLS's coordination with CLJA client agencies, I have gathered estimates of the custodial data to be collected for the CLJA litigation. This early case assessment permits OLS to estimate the type and amount of resources that will be necessary to support discovery efforts.

11. Initial evaluations estimate that the following volumes of custodial data may be subject to collection under the United States' proposed scope of discovery:³

³ I understand that the United States and Plaintiffs are in ongoing meet and confer discussions to determine a final list of custodians and search terms to be applied at the agency level prior to data retrieval. The estimates included in this Declaration are provided for purposes of understanding the case's landscape as it currently exists, and are subject to change following additional negotiations among the parties.

Agency	Custodians	Total Estimated Custodial Data Volume (GB)	Total Estimated Custodial Items (including unindexed items, where available)
USMC	3	127.87	222,490
DON	3	184	296,293
EPA	3	66.09	124,314
ATSDR	4	176.65	114,754
TOTALS:	13	554.61	757,851

12. To arrive at these preliminary estimates, the following collection search terms⁴ were applied at the agency level:

- a. "Lejeune"
- b. "Benzene"
- c. "Chloride"
- d. "Tetrachloroethylene"
- e. "Trichloroethylene"
- f. "Cancer"
- g. "Leukemia"
- h. "Lymphoma"
- i. "Parkinson" or "Parkinson's" or "Parkinsons"
- j. "pce"
- k. "tce"
- l. "dce"

⁴ These are preliminary collection results based on the provided search terms that account for searchable items within e-mail systems at the applicable agencies. The government may need to conduct additional analysis of unsearchable and unindexed items within those systems to determine whether further collection, separate workflows, or additional evaluation is required.

m. “VOC”

No other filters were applied.

13. Broader collection involving additional custodians will increase these estimates roughly proportionally. Thus, Plaintiffs’ requests to collect data from approximately 50 custodians—i.e. approximately 3.85 times more custodians—would likely result in the following overall approximate volumes of data for collection (assuming the same collection search terms are applied):⁵

Total Custodians:	50	TOTAL EST. ITEMS:	2,917,726
TOTAL EST. VOLUME (GB):	2,135.25	TOTAL EST. VOLUME (TB):	2.09

14. As mentioned above, while agency capabilities may vary, what remains consistent is that the efforts agencies will need to expend to collect this volume of data will take a significant amount of computer processing time that cannot be significantly shortened by simply adding personnel or resources.

15. Moreover, transmitting this volume of data from the collecting agencies to OLS will also take time that cannot be significantly shortened by simply adding personnel or resources. I estimate that data transmittal following agency collection for just the United States’ 13 proposed custodians will take approximately three days. Extrapolated out for Plaintiffs’ 50 custodians, transmission of data will take approximately 9 days, as there are some efficiencies achieved when transmitting larger datasets.

⁵ If collection search terms are not applied, these estimates will be astronomically larger, and are not covered in this declaration. As drafted, Plaintiffs’ discovery requests in this matter range back to the first use of e-mail in the government, which occurred approximately around the year 2000.

B. PROCESSING

16. Once the data has been received by OLS, it must be processed for use in the Civil Division's review and production platform, Relativity.

17. The form of collecting significantly impacts the steps and time needed to process data for review. If data is received in load ready format, processing may be fairly straightforward, consisting of loading information for review and conducting quality control. If, on the other hand, data is received in raw, unprocessed formats, OLS must take steps to first process that data either in-house or via contractors.

18. Most of the data slated for collection in the CLJA litigation is expected to be transmitted to OLS in unprocessed form due to various agencies' collection capabilities. Based on receiving the estimated data volumes described in paragraphs 11 and 13 above in largely unprocessed format, I estimate that processing—i.e. preparing the data for the case team's review in Relativity—will take approximately five working days for the United States' 13 proposed custodians. The estimated cost for processing this collection is \$76,047.80.

19. For the larger estimate reflecting Plaintiffs' proposed scope of discovery, I estimate that processing the data would take approximately 20 working days. The estimated cost to process this larger collection would be \$301,191.20.

20. While some of the processing can happen concurrently or on a rolling basis as data is received, computing and resources will necessarily limit how much data can be processed at any given time. This cannot be significantly impacted by simply adding personnel or resources.

C. REVIEW

21. Once the data is available for review, it must be searched and reviewed (the various

review approaches are discussed further below). Generally speaking, search and review can include any of, or some combination of the following approaches, which the Government intends to employ to meet its legal obligations under the Federal Rules of Civil Procedure:

- a. Linear—Review is conducted entirely by human reviewers reviewing each page;
- b. Search-term based—Relevant documents are searched for review using search terms or other filtering tools (such as date ranges), with any documents not meeting the filters not part of the primary review; and/or
- c. Technology Assisted Review (TAR)—Review is streamlined and prioritized by leveraging artificial intelligence to assist human reviewers. (Note that TAR does not work effectively on privilege review.)

22. A strictly manual review without using any additional search terms on the collected data in the United States' proposed scope of discovery as described in ¶ 12 would require 24,381 human hours/3,047.63 workdays, assuming that the average rate of review is 20 documents per hour (including a simultaneous privilege review). The estimated cost of this review is \$3,714,288.05.

23. Using additional negotiated search terms is likely to fairly substantially reduce the volume of documents requiring review. Generally, we expect that agreed-upon pre-review search terms will reduce review volume by at least 30%. Based on this estimate, I expect that using search terms here will result in a review universe that would require 16,537 human hours/2,067.13 workdays to complete a linear review of the United States's proposed scope of discovery. This would reduce the estimated cost of review to \$2,519,302.14.

24. Reviewing the approximately four-times-larger set of discovery proposed by Plaintiffs

would take proportionally longer. I expect that review of the data described in ¶ 13 would require 93,866.90 human hours/11,733.36 working days. The estimated cost to review this increased scope would be \$14,856,912.21. Applying search terms as described in ¶ 22 to this larger set would require 63,668 human hours/7,959 working days to review.

25. Leveraging TAR after application of negotiated pre-review search terms is also a possible review option. Assuming the corpus of documents is appropriate for review, TAR can potentially expedite the review process. TAR is a computer program that uses artificial intelligence to assist humans to review documents. In TAR, humans will make judgments about a sample set of documents that the computer will then apply to larger set of documents (i.e. the AI model). In my experience, TAR does not work effectively to assist in privilege review, and that review must occur separately, as noted below.

26. TAR's effectiveness depends on many variables that must be analyzed after the documents are processed and, in some cases, during the review itself. More specifically, the documents or information should be the type that will work with TAR algorithms. Typically, these are documents with sufficient narrative text for the software to identify various conceptual, semantic, and linguistic relationships. For example, e-mails and Word documents will likely support TAR use, whereas image files, scanned documents, or information from databases and Excel spreadsheets with little text will not. Additionally, the TAR population must have appropriate richness to be effective (e.g. the estimated percentage of responsive documents out of a total TAR population). Further, many variables in the data itself will impact how long it will take for the project to be successful, such as the complexity of issues and documents within the corpus. Thus, while the Government offers its best estimate for the purposes of planning and scheduling, the Government cannot make any guarantees regarding

the projected time to complete a TAR project of this magnitude.

27. I project it would take approximately 1,488 human hours/186 workdays for the 13 custodians and 5,730 hours/716.27 workdays for the approximately 50 custodians for the TAR AI model to stabilize (i.e. to defensibly complete the TAR project). I am making this estimate assuming 18% of the TAR eligible documents will need to be manually reviewed for the AI model to defensibly and reliably apply review decisions to the whole TAR-corpus, and that pre-review search terms have been applied prior to TAR as discussed in ¶ 23. The estimate is based on the historical average (which can vary widely depending on the richness and complexity of the TAR corpus). In this case, based on the total size of the estimated collected data, adjusted for projected search term culling, I estimate it would take at least 47 days with four reviewers⁶ to complete a successful TAR stabilization and initial review project for the collection of 13 custodians, and 179 days for the collection of approximately 50 custodians. The estimated cost of running a TAR review would be \$226,737.19 for 13 custodians and \$906,948.77 for approximately 50 custodians.

28. Even with the use of TAR for TAR-eligible documents, we must still account for non-TAR documents in separate workflows. For the purpose of estimating the portion of non-TAR-eligible documents, we have observed collections to have non-TAR-eligible volumes between 10% and 30% of the total volume.⁷ Assuming we could statistically sample the data and not apply a linear review, the sampling would add another five days to the project based on our review statistics above.

⁶ Importantly, adding more reviewers will not necessarily speed up the TAR project. More specifically, TAR requires some consistency with reviewers to train the AI to work effectively, and thus a TAR project has diminishing returns if there are too many reviewers, which can sacrifice consistency in decision-making and confuse the model, leading to a longer project.

⁷ Non-TAR document volumes could be impacted by any additional unsearchable and unindexed items identified from the collection sources, as noted above.

29. Once responsive TAR & non-TAR documents have been identified, the United States must review for privilege. TAR processes are not sufficiently advanced to reliably identify privileged materials, including Government-specific privileges. As a result, the Government will need to review the responsive documents for privilege prior to production, the volume of which is dependent on the richness of the document collection. Estimates of the total responsive document collection can vary wildly, but for the purpose of this affidavit, I used 20% responsive rate (which is within the normal rate of 8% to 40%). This would equate to 66,149 documents for the collection of 13 custodians and 254,674 for the collection of approximately 50 custodians. The estimated privilege review time for the collection of 13 custodians is 1,654 human hours/207 workdays and the collection of approximately 50 custodians is 6,367 human hours/796 workdays. The estimated cost would be \$324,743.20 for the collection of 13 custodians and \$1,298,972.80 for the collection of approximately 50 custodians.

D. PRODUCTION

30. Once the case team has determined which documents should be produced and marked them for any necessary endorsements (such as redactions or protective order warnings), OLS can begin preparing the data for production.

31. Production times depend on the data volume and types at issue. Generally speaking, production time for the 13 custodians would take 3 days, and production time for approximately 50 custodians would take 12 days. Additional time is required to conduct quality control checks of the outgoing data. I estimate that quality control for the 13 custodians to be 7 days and 14 days for the 50 custodians.

32. As with incoming data, producing data requires time to transmit the data from the

United States to the receiving party. I estimate that data transmittal will take approximately 1 day.

E. SUMMARY

33. To summarize the above analysis, please see the following. These estimates assume 7 attorneys working exclusively and full time on the review:

	Working Days for Complete Production of 13 Custodians ⁸	Working Days for Complete Production of 50 Custodians
Strict Linear Review Only	454	1,732
Search-Term Based Review	314	1,193
TAR Review (with search terms and privilege review)	102	354

I declare under penalty of perjury that the foregoing is true and correct.

JOSHUA WOOD Digitally signed by JOSHUA WOOD
Date: 2023.12.21 15:58:51 -05'00'

⁸ The total number of days for all estimates includes: data transmittal estimates (¶ 15), processing time (¶¶ 16-20), review time (¶¶ 21-29), production time (¶ 31), quality control time (¶ 31), and transmittal time (¶ 32).