U.S. ENVIRONMENTAL PROTECTION AGENCY, REGION 2

July 30, 2020

BY ELECTRONIC MAIL

Robert Law, Ph.D.
de maximis, inc.
186 Center Street, Suite 290
Clinton, New Jersey 08809

Re: Re: Lower Passaic River Study Draft Feasibility Study (FS*)* – Administrative Settlement Agreement and Order on Consent for Remedial Investigation/Feasibility Study (Agreement) CERCLA Docket No. 02-2007-2009

Dear Dr. Law:

The U.S. Environmental Protection Agency (EPA) has reviewed the response to comment file received from the Cooperating Parties Group (CPG) on July 21, 2020 for the Lower Passaic River Study Area (LPRSA). Based on a discussion between the CPG and EPA on July 23, 2020, EPA has provided follow up responses to comment numbers 187, 264, 380, and 397 from the EPA July 16, 2020 response to comment file on the CPG's *Draft Upper 9-Mile Source Control Interim Remedy Feasibility Study (FS) Revision 1* dated May 15, 2020 and comment number 26 from the July 16, 2020 response to comment file on the CPG's *Appendix H (the Interim Remedy Completion Evaluation Framework) of the draft Interim Remedy Feasibility Study Report*. In accordance with Section X, Paragraph 44(d) of the Agreement, EPA has enclosed an evaluation of CPG's response to comment file with this letter.

Please proceed with revisions to the *Draft FS Rev 1* within 30 calendar days consistent with the enclosed comment evaluations. If there are any questions or clarifications needed, please contact me to discuss.

Sincerely,

Diane Salkie, Remedial Project Manager
Lower Passaic River Study Area RI/FS
Enclosure

616117

Cc: Zizila, F. (EPA)
Sivak, M. (EPA)
Hyatt, B. (CPG)
Potter, W. (CPG)
Nickerson, J. (NJDEP)

IR FS Section 7.1.6

#s187, 380, and 397 from July 16, 2020 IR FS comment file:

#	Location	Original EPA Comment	CPG May 21, 2020 Response	EPA July 16, 2020 Evaluation	CPG July 21, 2020 Follow-Up
187	Section 7.1.6	Expand the bulleted list that describes monitoring elements to first provide some level of detail regarding the PDI sampling program and its expected scope (e.g., sampling on a spatially dense grid, approximately 2,000 sampling locations with cores collected to evaluate surface and subsurface conditions, a second round of infill sampling as needed), then to describe the types of construction monitoring that are anticipated, including performance metrics, then to describe the post-IR sediment sampling program (e.g., statistically unbiased sampling, not less than 400 sampling locations), then to describe the general post-IR decision making framework, O&M, and the long-term monitoring. Within the bullet that summarizes the post-IR decision making, ensure that the text describes the post-IR sediment sampling and evaluation of post-IR sediment sampling data as a critical evaluation to determine IR completion. Within the bullet that describes O&M, provide additional information that indicates the types of O&M monitoring that would be expected and the nature of the maintenance that might be triggered.	The text was revised to address this comment, with the caveat that no sampling programs have been scoped at this time, and the details provided in the IR FS are preliminary and subject to change during development of the monitoring programs.	The response and corresponding FS revisions are partially acceptable. Under the PDI bullet, make the following revisions: • Include geotechnical work as a specific component or a supporting survey. • For the bullet that describes the second round of PDI sediment sampling, also indicate that this infill sampling would be performed to better constrain data variability. • For the bullet that describes supporting surveys, delete “to be determined during remedial design”. The PDI will precede completion of the design, and even if certain surveys may arise out of design planning, deleting this phrase eliminates potential confusion. In the construction monitoring bullet, include construction-related sediment sampling. EPA expects some limited scope of construction-related sediment sampling to evaluate potential impact from dredging releases and support the construction certification process. In the post-IR confirmation sampling bullet revise the text to read, “Post-IR confirmation sampling would include sufficient samples to provide a statistically unbiased estimate of the post-IR SWACs, and is anticipated to include on the order of ***not less than 400 and not more than 800*** sediment sample locations ***at which three closely spaced samples will be collected and composited***.” In the O&M monitoring bullet, include chemical monitoring as well, which will be needed to evaluate contaminant isolation. In the long-term monitoring bullet, the description of actual monitoring activities is vague.	Several of the EPA comments [187, 380, 397] relate to sediment sampling during construction, which was noted in a parenthetical in FS Section 7.1.6. Despite its inclusion in the parenthetical, the CPG does not see purpose in sediment sampling during construction. Such sampling is of little utility in the real time reaction to dredging-induced releases, and CPG does not envision that there would be other DQOs for these data. Water quality monitoring is necessary and sufficient to provide the data needed to adjust operations to mitigate releases and associated recontamination. Sediment sampling will not provide the timely information needed to make effective adjustments. Moreover, it is not needed to assess dredging-induced residuals as that assessment will be supported by the comprehensive post-remedy sediment sampling. No change proposed.

				Include a reference to Appendix D, where more information is included related to longer-term data collection.	
380	Section 8.3.X.2	Cap Stability, first sentence: Include sediment sampling along with water column sampling in the first sentence of this paragraph as a component of construction monitoring to ensure minimal resuspension and residual impacts.	N/A	N/A	N/A
397	Section 8.4.2.1	Include sediment sampling as an element of construction monitoring that would be used to ensure minimal resuspension and residual impacts.	N/A	N/A	N/A

CPG proposed language change to IR FS Section 7.1.6 (paragraph 2); from file dated July 27, 2020:

"An IR completion assessment process would be performed to verify that RAO 1 has been achieved. The assessment would consider construction monitoring conducted during remediation to evaluate compliance with the performance requirements specified by the remedial design (i.e., water quality monitoring, bathymetric surveys, discharge monitoring, inspection surveys, sediment monitoring ***sampling solely to evaluate the residual management measures being employed***[1]) and post-remedy confirmatory sediment sampling. These monitoring activities, together with a multi-stage PDI and a robust design process and footprint delineation, comprise the multiple lines of evidence that will be evaluated to verify attainment of RAO 1."

[1] ***Limited sediment sampling would be performed after the completion of the first dredging season, targeting newly deposited sediment on top of capped areas, for the sole objective of evaluating the efficacy of dredging BMPs. The utility of the sediment monitoring would be evaluated and discontinued after the first season if (a) sampling of newly deposited materials on capped surfaces proves impracticable, (b) the concentrations of newly deposited materials are consistent with or lower than near-field water column concentrations measured during active dredging, or (c) the variability and complexity of the system limits the ability to ascertain the cause of any elevated concentrations on the cap and consequently limits the ability to revise BMPs any further than what is concluded using the water column data.***

EPA revisions to CPG proposed language change to IR FS Section 7.1.6 (paragraph 2):

During the July 23, 2020 comment resolution discussion, EPA and the CPG discussed the value of sediment sampling as a component of the construction performance monitoring program aimed at understanding dredging releases and the potential effect of redeposition. The CPG agreed to include sediment sampling as a component of the construction performance monitoring program, and EPA agreed that this sediment sampling would be for the purpose of assessing residuals management practices and the potential to modify construction approaches and BMPs to minimize construction impacts. EPA also reiterated the expectation that this sediment sampling would be limited in scope (with the specific scope to be determined during IR design). EPA reviewed the CPG's suggested revision to Section 7.1.6 in the IR FS, and requests the following edits:

"An IR completion assessment process would be performed to verify that RAO 1 has been achieved. The assessment would consider construction monitoring conducted during remediation to evaluate compliance with the performance requirements specified by the remedial design (i.e., water quality monitoring, bathymetric surveys, discharge monitoring, inspection surveys, sediment sampling to evaluate the residuals management measures being employed[1]) and post-remedy confirmatory sediment sampling. These monitoring activities, together with a multi-stage PDI and a robust design process and footprint delineation, comprise the multiple lines of evidence that will be evaluated to verify attainment of RAO 1."

[1] Limited sediment sampling would be performed after the completion of the dredging season, targeting newly deposited sediment on top of capped areas and/or areas that have received RMC, for the objective of evaluating the efficacy and potential improvement of dredging BMPs. The utility of the sediment monitoring would be evaluated and this monitoring may be discontinued after the first season or a subsequent season if (a) sampling of newly deposited sediment proves impracticable, (b) the concentrations of newly deposited sediment are consistent with or lower than near-field water column concentrations measured during active dredging and the water column monitoring demonstrates compliance with the performance standards, or (c) the variability and complexity of the system limits the ability to ascertain the cause of any elevated concentrations in newly deposited sediment and consequently limits the ability to revise BMPs any further than what is concluded using the water column data.

EPA also expects that comments #187, #380, and #397 will otherwise be addressed in their entirety.

IR FS Table 8-7

#264 from July 16, 2020 IR FS comment file:

#	Location	Original EPA Comment	CPG May 21, 2020 Response	EPA July 16, 2020 Evaluation	CPG July 21, 2020 Follow-Up
264	Section 8.4, Table 8-7	The results in the overall summary, Alternative 2 (4 checks) to Alternative 4 (2 checks) are driven by a few nuanced differences. The checks may suggest that the Alternative 2 scores twice as high as	The summary table was revised to include additional quantitative values, as well as show smaller	The response and corresponding FS revisions are partially acceptable.	Missing direction for some alternatives for some metrics a. For long-term effectiveness and

		alternative 4. Recommend using relative percentages for the metrics that are quantifiable. 1. <u>Reduction of Toxicity, Mobility, or Volume through Treatment (should be weighted more heavily, but the difference among alternatives is relatively small)</u> Alt 2 scores the lowest among 2 to 4. The checks may suggest half as effective as 4, but Alt 2 removes 95% of the mass of 2,3,7,8-TCDD and total PCBs removed in Alt 4. 2. <u>Short-Term Effectiveness, Worker Risk and Community Impact</u> This difference is entirely driven by volume and therefore schedule. Again, the checks suggest Alt 2 is three times better than Alt 4, but the difference is an additional 7 months of remediation (14% longer). The text for these is identical with the exception of the three numbers; area, time, and volume. 3. <u>Implementability</u> This difference is entirely driven by volume and therefore schedule. Again, the checks suggest that Alt 2 is three times better than Alt 4, but the difference is an additional 56,00 CY (15% more volume). In addition, the preliminary footprints suggest that the additional area in the Alt 4 footprint tends to be around the outside of the Alt 2 footprint, suggesting that there may be some economies of scale (e.g., a few additional dredge cycles before having to relocate the dredge). There are no differences in the text for Implementability other than the volume.	differences among alternatives. Note that it is the areas, volumes, and durations that drives the differences between alternatives, these are not nuanced differences. The comparative summary table is just that, a relative comparison and not an absolute result, and in a comparative sense, the alternative do rank in order in terms of volume and duration.	While the general summary information (e.g., attained SWACs, mass removed) and the summary of performance under the threshold criteria are appropriately provided for Alternatives 1 and 5 in Table 8-7, the inclusion of visually comparative information (i.e., the Harvey Balls) for the balancing criteria for Alternatives 1 and 5 confuses the evaluation of Alternatives 2, 3, and 4. Revise Table 8-7 to maintain the summary information for all alternatives but remove the comparative information for NCP criteria for Alternatives 1 and 5 other than for the threshold criteria of overall protection of human health and the environment and compliance with ARARs. Alternative 5 would not attain the threshold criteria and is otherwise not eligible for selection by agreement between EPA, NJDEP, and the CPG. Therefore it is unnecessary to visually demonstrate its relative performance for the balancing criteria in Table 8-7. Alternative 1 would not attain the threshold criteria, and EPA believes that the level of information provided in the narrative of Section 8 and that would remain in Table 8-7 would be consistent with NCP requirements to carry Alternative 1 through the IR FS. This would also be consistent with the fundamental intent of Sections 8.4 and 8.5 in the IR FS, which is to compare Alternatives 2, 3, and 4. Revise footnote a of Table 8-7 to read “Does not achieve the metrics for the threshold criteria for the upper 9-mile interim remedy, and therefore visual comparison of performance for the balancing criteria is not included in this table”. Also, as suggested by other requested revisions to Section 8 in this evaluation file, EPA believes there are more nuanced differences between alternatives than are currently reflected in Table 8-7. Make the following revisions to Table 8-7: • Update the Harvey Balls to reflect 5% increments, which is necessary to objectively and accurately reflect the small-scale but important variation between Alternatives 2, 3, and 4 based on the analyses performed in the IR FS. • To avoid the appearance that there could be relative difference in performance for the threshold criteria, replace the filled circles with “YES” and the unfilled circles with “NO” for overall protection of human health and the environment and compliance with ARARs. • For long-term effectiveness and permanence, use a 95% filled circle for Alternative 2 and Alternative 4. Also make the following revisions for the metrics under long-term effectiveness and permanence: b. For source control and recovery potential, use a 95% filled circle for Alternative 2 [Alt 3 and 4?]. c. For monitoring, maintenance, and ICs, use a 95% filled circle for Alternative 3 and an 85% filled circle for Alternative 4 [Alt 2?].	permanence, use a 95% filled circle for Alternative 2 and Alternative 4 [Alt 3?]. Also make the following revisions for the metrics under long-term effectiveness and permanence: b. For source control and recovery potential, use a 95% filled circle for Alternative 2 [Alt 3 and 4?]. c. For monitoring, maintenance, and ICs, use a 95% filled circle for Alternative 3 and an 85% filled circle for Alternative 4 [Alt 2?].

				metrics under long-term effectiveness and permanence: o For source control and recovery potential, use a 95% filled circle for Alternative 2. o For monitoring, maintenance, and ICs, use a 95% filled circle for Alternative 3 and an 85% filled circle for Alternative 4. • For reduction of toxicity, mobility, or volume through treatment, use a 50% filled circle for Alternatives 2 and 3, and a 55% filled circle for Alternative 4. • For short-term effectiveness, use a completely filled circle for Alternative 2, a 95% filled circle for Alternative 3, and a 85% filled circle for Alternative 4. Also make the following revisions for the metrics under short-term effectiveness: o For time to achieve RAOs, use a completely filled circle for Alternative 2, a 95% filled circle for Alternative 3, and a 90% filled circle for Alternative 4. o For worker risk and community impact, use a completely filled circle for Alternative 2, a 95% filled circle for Alternative 3, and a 85% filled circle for Alternative 4. o For resuspension, use a completely filled circle for each alternative. o For downstream and upstream transport, use a completely filled circle for each alternative. • For implementability, use a completely filled circle for Alternative 2, a 95% filled circle for Alternative 3, and an 85% filled circle for Alternative 4. • For cost, in addition to showing the numerical value, use a completely filled circle for Alternative 2, a 95% filled circle for Alternative 3, and a 90% filled circle for Alternative 4.	

General description from EPA of Table 8-7 scoring process (**for general information only, not to be included in the IR FS**):

Recognizing that Alternatives 2, 3, and 4 represent fundamentally the same IR approach to achieve subtly different SWAC objectives, the relative differences between these alternatives are expected to be and are understandably small. However, to better portray these small-scale but important differences when presented in the format of Table 8-7, EPA believes that a further refinement of the visual representation presented in the May

2020 revised draft IR FS document is warranted. Given the scale of the differences observed between the IR alternatives thorough the evaluations performed for the IR FS, EPA believes 5% increments are appropriate and effective for objectively comparing the alternatives.

EPA performed an independent evaluation of the comparative performance of Alternatives 2, 3, and 4, relying directly on the evaluations presented in the IR FS, to provide the feedback for Table 8-7 that is conveyed in #264 of the EPA's July 16, 2020 evaluations of the CPG's responses to prior comments on the IR FS. In #264 of the July 16, 2020 comment file, EPA indicated the specific visual representation that should be used for each NCP balancing criterion and each metric under each NCP balancing criterion in Table 8-7. As EPA described to the CPG during the July 23, 2020 comment resolution discussion, specific criteria and metrics not explicitly commented on by EPA by way of #264 in the July 16, 2020 comment file would remain the same as in the May 2020 version of Table 8-7 (completely filled circles in these instances). Also as indicated by EPA during the July 23, 2020 comment resolution discussion, the visual representation for Alternative 4 for Short-Term Effectiveness should be a 90% filled circle (as opposed to an 85% filled circle as was conveyed in #264 of the July 16, 2020 comment file).

Footnote for Table 8-7 from EPA (to be included in the IR FS):

The relative ranking of Alternatives 2, 3, and 4 for the balancing criteria, as reflected by circles filled in 5% increments (a more filled circle represents a higher degree of relative performance), is based on the evaluation of the specific metrics (including sub-metrics as relevant) or the measures that are described in the text to assess alternative performance. Where multiple metrics, sub-metrics, and/or measures are used to assess performance, they are aggregated to a total ranking for each criterion. Where comparison to a benchmark is possible (e.g., mass removed on an alternative-specific basis compared to total mass inventory), relative performance reflects this comparison. Where a benchmark does not exist, relative performance is reflected as a completely filled circle for the highest performing alternative and then comparatively diminished performance for the other alternatives. In the absence of a benchmark, a completely filled circle for the highest performing alternative does not necessarily represent all factors that could diminish the performance of even that highest performing alternative, but this methodology is reasonable to demonstrate a comparative evaluation between Alternatives 2, 3, and 4.

10% False Positive Error Rate

#26 from July 16, 2020 IR FS Appendix H comment file:

#	Location	Original EPA Comment	CPG May 21, 2020 Response	EPA July 16, 2020 Evaluation	CPG July 21, 2020 Follow-Up
26	Section 2.4, Paragraph 4	The final sentence in this paragraph states "USEPA considers a level of 95% to be acceptable for the upper bound that will establish the Y value for the post-IR statistical testing." To avoid any confusion between 95% as an appropriate level of statistical certainty for the confidence intervals around the post-IR SWACs and 95% confidence as an expression of control against a false negative declaration, restate this as "USEPA considers ***an error rate*** of ***5%*** to be acceptable for the upper bound ***of a potential false negative outcome*** that will establish the Y ***values*** for the post-IR statistical testing***. This corresponds to a 95% level of confidence that the IR would not be concluded to have not attained the RAO 1 SWAC goals when in fact it did.***" Also note that this portion of Appendix H should also describe the acceptable level of confidence around a potential false positive outcome, where the IR would be concluded to have been successful when the true post-IR 2,3,7,8-TCDD and/or total PCB SWAC(s) is/are actually not statistically equal to or less than the RAO 1 SWAC goal(s). False negative and false positive error rates are controllable through selection of Y values and the post-IR sample size. EPA, NJDEP, and the CPG have agreed on the maximum 5% error rate for a false negative outcome and have discussed the false positive error rate (i.e., 10% as most recently discussed). EPA recognizes that additional discussion may be necessary to arrive at consensus on this false positive error rate level.	The text has been revised accordingly.	The response and corresponding FS revisions are partially acceptable. In the final paragraph of current Section 2.4.2, the final sentence states "The chosen Y value and the post-IR sample size will also reflect a 10% potential false positive outcome, where the IR would be concluded to have been successful when the true post-IR 2,3,7,8 TCDD and/or total PCB SWAC(s) is/are actually above an acceptable level of equivalency to the RAO 1 SWAC goal(s) (defined as Y*RAO 1 SWAC goals)." Revise this sentence to state this more simply as "…is/are actually greater than Y times the RAO 1 SWAC goals." In addition, add language to this paragraph to describe the rationale for selecting unequal false negative and false positive error rates, as this is an important consideration for project stakeholders.	The values of 5% for potential false negative outcome and 10% for false positive outcome were established during the FS Meetings with CPG, EPA, and DEP. That is the primary basis of these values appearing in Appendix H. Regarding justification for 10%, CPG can make the requested text edit and asks EPA to provide their interpretation of the justification for using 10% for the acceptable error rate of false positives to incorporate into the text.

Edits from EPA to final paragraph of current Section 2.4.2 in IR FS Appendix H:

"The Y values will be established based on statistical simulations of post-IR sampling data drawn from concentration maps derived from PDI data and modified such that remediated areas are assigned a residual concentration informed by modeling of IR implementation (see Attachment 1). The Y values will be set such that the expected frequency of false negatives (i.e., concluding that the SWAC goals were not achieved when the true means for the post-IR sediment surface interval are at or below the RAO 1 SWAC goals) derived from the statistical simulations is not more than 5%. USEPA considers an error rate of 5% to be acceptable for the upper bound of a potential false negative outcome that will establish the Y value for the post-IR statistical testing. This corresponds to a 95% level of confidence that the IR would not be concluded to have not attained the RAO 1 SWAC goals when in fact it did. The chosen Y value and the post-IR sample size will also reflect a 10% potential false positive outcome, ***which is also acceptable to USEPA,*** where the IR would be concluded to have been successful when the true post-IR 2,3,7,8 TCDD and/or total PCB SWAC(s) is/are actually ***greater than Y times the RAO 1 SWAC goals*** above an acceptable level of equivalency to the RAO 1 SWAC goal(s) (defined as Y*RAO 1 SWAC goals). ***While the false negative (5%) and false positive (10%) error rates are not equal, the error rates are not statistically required to be equal. The slightly different error rates reflect reasonable and industry-typical rates of error for statistical assessments and support the application of a post-IR sampling program of an appropriate scale to derive statistically unbiased estimates of the post-IR SWACs (see Section 2.4.1). The error rates also reflect appropriate balance between errors that would incorrectly suggest a successful IR was not successful (i.e., false negative, which could lead to a range of unnecessary additional actions to fulfill the intent of the IR) versus errors that would be recoverable (i.e., false positive) through the Adaptive Management Process that would include rigorous evaluation of system response and system recovery following the IR and culminate with the selection, implementation, and demonstration of a final remedy to address remaining risks and attain risk-protective conditions.***"