MEMO

SUBJECT Peer Review of EPA Region 2/CENAN Framework
DATE August 21, 1998
TO Roland B. Hemmett, Ph.D, USEPA Region 2
FROM Tom Dillon, Ph.D, NOAA HAZMAT

CC Alyce Fritz, Ph.D., NOAA HAZMAT

Background: Subject peer reviewed was requested by EPA Region 2. Basis of the review provided by EPA were: 1) Memo for the Record describing a hypothetical testing results and 2) pertinent references. EPA format for the review was a series of questions. Those questions are shown below in italics followed by the peer reviewer's response.

Framework

1. *Is the EPA Region 2/CENAN Framework for evaluating bioaccumulation results scientifically appropriate for determining the suitability of dredged material as Remediation Material? If not, describe deficiencies. (Please see Region 2/CENAN joint evaluation memorandum, Figure 1)*

Many scientifically appropriate approaches designed to evaluate potential effects of environmental contaminants are tiered. Early tiers are usually simple and environmentally conservative. They promote environmentally protective decisions (i.e., avoiding Type II errors) while minimizing cost and effort. Later tiers are more complex and site-specific. They provide greater clarity and certainty for decision-makers.

It is not clear if the subject Framework is tiered. Bioaccumulation bioassay results are compared (more or less sequentially) to: 1) results with reference material, 2) FDA levels, 3) matrix or regional dioxin values, 4) a "risk evaluation" and 5) the eight Green Book factors. It is not clear if these evaluations are tiered. For example, is the initial comparison to reference the most environmentally conservative? If the Framework is not tiered, analytes of concern could be removed from consideration prematurely.

I recognize the larger dredged material evaluation as described in the Green Book is tiered. This concept should be extended to the subject Framework.

2. *Which of the risk-based values derived constitute "true" conservative estimates of risk*

levels (i.e., exceeding the value should be interpreted as sufficient cause to conclude that significant undesirable effects may result through bioaccumulation)? Which of the risk-based values derived constitute conservative screening values (i.e., test tissue concentrations below the value can confidently be interpreted to pose no risk of significant undesirable effects and exceeding should be further evaluated before the probability of significant undesirable effects can be assessed)? How can the "true" risk levels be calculated for those compounds which you believe only to have screening values? How should test concentrations be compared to risk-based levels to determine whether they are exceeded.

Risk is the probability or likelihood of adverse effects. Chemical risk is assessed by determining the probability of exposure (exposure assessment) and the consequences of that exposure (toxicity or effects assessment). Because risk assessment it is a probability-based process, "true" risk as it used in the above question, does not exist. Rather, risk is a gradation of potential outcomes ranging from low to high hazard.

I assume the "risk-based values" in the question refer to the four elements in Framework's Block c labeled "Risk Evaluation for Chemical "X". These four elements are: 1) estimates of steady-state bioaccumulation and food-chain transfer, 2) background comparisons, 3) ecological effects assessment, 4) carcinogenic and non-carcinogenic effects on human health. None of these four considerations can be classified as more or less "conservative". Nor are they "risk-based values" per se. That is, they don't systematically assess exposure and toxicity *vis a vis* specific receptors. The first two are exposure estimates. The last two describe methods for evaluating effects to ecological and human receptors, respectively.

3. *In conducting the integrated effects evaluation using the types of data provided by the applicant, which of the eight factors for LPC compliance listed in the Green Book are appropriate and relevant? How can a quantitative/strategic framework be established to evaluate tissue data for those factors? Considering that comparison to regional Matrix values and site-specific risk values represent case-specific evaluations, is it necessary to conduct the integrated effects evaluation of the bioaccumulation results? (Please see Reference No. 61, page 6-6)*

All eight factors appear very appropriate and relevant. One may wish to keep consideration of the factors qualitative. A simple visual tool such as a matrix with pluses and minuses may be all that is necessary. This visual level of analysis (with accompanying uncertainty analysis) may even be most appropriate given the uncertainties associated with the multiple lines of evidence. Yes, the integrated effects evaluation is very necessary. For example, it is the only time: 1) the potential effects of multiple contaminants are evaluated and 2) results of the sediment toxicity bioassay are considered in conjunction with the bioaccumulation data.

Benchmark and Risk Evaluation Values

4. *Regional Matrix Values*

A. *Are the Matrix values suitable for determining the suitability for placement at the HARS as Remediation Material?*

Matrix values represent 1981 conditions. If the goal of the EPA/CENAN dredging program is to manage towards that level of ambient contamination, then the values appear appropriate.

B. *Regional Matrix values were developed in 1981 by compiling available field data for mercury, cadmium, PCBs, and total DDTs. Were these values derived appropriately for their intended use? Based on current data sets and scientific literature, are these 1981 values suitable for predicting the significant undesirable effect due to bioaccumulation? (Please see Reference No. 57) If not, identify more current references, data sets, and/or actual chemical specific values that would be more appropriate.*

The 1981 matrix values were derived appropriately for their intended use at the time. Whether the 1981 matrix values are appropriate today depends on regional management goals (See peer reviewer's response to 4.A.). If the goal of the EPA/CENAN dredging program is to manage towards present day levels, then the more recent background concentrations contained in reference 98 may be more appropriate.

5. *Regional Dioxin Values* **This issue was not assigned to Dr. Dillon**

A. *Currently, the presence of 2,3,7,8-TCDD at a detectable concentration (i.e., greater than or equal to one part per trillion (pptr)) in tissues of organisms exposed to dredged material precludes its classification as Category I (hence Remediation Material); presence of the remaining dioxin/furan congeners, at concentrations of TEQs equal to or greater than 4.5 pptr, results in a similar conclusion. When 28-day tissue concentrations exceed these values, is there sufficient cause to conclude that placement of the material is not suitable as HARS Remediation Material? If not, what levels indicate sufficient cause for this conclusion? (Please see Reference No. 89)*

B. *Are dioxin values suitable for predicting the significant undesirable effects due to bioaccumulation? If not, should these values be based on a risk analysis paradigm in which the size of the human population subgroup potentially exposed through intentional behavior is compared to the size of the general population in the EPA? Since the primary route of exposure is through consumption of fish and shellfish, should the variability in potential exposure due to differences in fishing behavior (e.g., target species, seasonal preferences) be incorporated in the risk paradigm? How would a benchmark protective of human health compare to benchmarks determined using an ecological risk analysis paradigm for resident fish and piscivorous wildlife?*

6. *FDA Action Levels (Please see Reference No.61, Sec. 6.3)*

A. *Are FDA Action Levels useful as upper limit human health benchmarks?*

See peer reviewer's comment at end of questions regarding human health evaluations.

B. *Would the evaluation be improved by omitting comparison of tissue results to FDA Action Levels?*

See peer reviewer's comment at end of questions regarding human health evaluations.

7. *Human Health Risk, Cancer and Noncancer*

A. *Are the risk values suitable for determining the suitability for placement at the HARS as Remediation Material? If there are better alternatives for human risk, specifically what are they?*

For possible alternatives, see peer reviewer's comment at end of questions regarding human health evaluations.

B. *Benthic tissue levels for cancer protection were derived using assumptions focused on attaining a cancer protection at the 10^{-4} risk level. Is this risk appropriate for a determination of ocean placement of Remediation Material? (Please see Region 2/CENAN joint evaluation memorandum, Appendix for Table 1, Page A-4, A-5)*

Setting levels of protection for cancer risk is a matter of policy. For example, EPA's policy in the Superfund program is to use the risk range of 10^{-4} to 10^{-6} as a "point of departure". Generally speaking, risks within the range require a site-specific baseline risk assessment, risks less than the range require little to no further evaluation while risks above the range require immediate attention (i.e., removal/remediation). On the other hand, the State of Florida has promulgated its risk policy as a single deterministic point of compliance; 10^{-6}.

C. *Benthic tissue levels for noncancer protection were derived using Reference Dose (RfD) of several organic and inorganic contaminants for the protection of human health. Are these values appropriately and consistently derived? Is the whole body/fillet conversion factor of 1.35 an appropriate factor for all of the contaminants considered if human exposure is assumed to be primarily via consumption of the fillet portion of the fish? (Please see Region 2/CENAN joint evaluation memorandum, Appendix for Table 1, Attachments B and C) If not, what factors would be appropriate? For the lead noncancer value, since there is no RFD for lead the value was derived differently than the other metals. Was the value derived appropriately? (Please see Reference No. 88)*

Not peer reviewer's area of expertise.

D. *Are the risk values suitable for predicting the significant undesirable effects due to bioaccumulation? Since the primary route of exposure is through consumption of fish and shellfish, should the variability in potential exposure due to differences in fishing behavior (e.g., target species, seasonal preferences) be incorporated in the risk paradigm?*

Suggestions in peer reviewer's comment at end of questions would provide for a way to account for variability in site-specific differences in fishing behavior, ingestion rates, ingested seafood items, target human populations, etc.

8. *Ecological Risk*

A. *Ecological effects benchmarks include the Water Quality Criteria Tissue Level (WQCTL), Critical Body Residue (CBR) associated with narcotic responses, and certain mutagenic/teratogenic effects. Is it valid to use the CBR effect end point for evaluating significant undesirable effect? Are there other ecological end points that should be used to measure ecological risk that are protective of marine benthic and fish life via trophic transfer, particularly for PAHs? If so, identify. With regard to a narcotic effect for chlorinated organic compounds, should an additive approach be considered to include the contribution of chlorinated hydrocarbons against this narcotic (CBR) endpoint.*

Critical body residues, when expressed on a molar basis, is an appropriate endpoint for chemicals where the mode of toxicity is narcosis. CBR is not appropriate for chemicals with other modes of action (e.g., receptor-mediated, neural disruption). Additivity of effects is an appropriate assumption for narcotic chemicals if expressed on a molar basis. However, almost all our knowledge of narcotic chemicals is based on acute exposures and lethality endpoints. Environmental exposures rarely approach acutely lethal levels. On the other hand, chronic narcosis (especially effects on behavior) may be more insidious and environmentally relevant. This effect has received scant attention.

PAHs are generally thought to be directly toxic to benthos. Risk Swartz's sigma PAH model is a good evaluative tool for direct toxicity. The risks associated with PAHs entering the food web is less certain. Most uncertainties are associated with PAH metabolites. Analysis of metabolites tends to be difficult and expensive, they are ephemeral and organisms' ability to metabolize PAHs range widely both quantitatively and qualitatively. Many consider PAHs in the food web a low risk scenario because they are rarely detected at noticeably levels (except sediments and mollusks). However, this belief is based on parent PAH compounds, not their metabolites. We know relatively little of PAH metabolites' environmental fate and effects on survival, growth and reproduction.

B. *Is the EPA 2 WQCTL approach (i.e., multiplying the Water Quality Criteria Chronic Value by the Bioconcentration factor) appropriate for determining ecological effects levels of the contaminants for which they were developed? Specifically, are the appropriate BCFs used (for*

fish, bivalves, etc)? (Please see Region 2/CENAN joint evaluation memorandum, Appendix for Table 1, Page A-1)

The WQCTL approach is one tool for evaluating the toxicological significance of bioaccumulation results. It is probably appropriate if one accepts certain assumptions: 1) chemicals of concern are not metabolized , 2) internal sequestering mechanisms are of minor importance, 3) benthic organisms' sensitivity is similar to nektonic forms upon which CVs were based, 4) benthic exposure approximates that for nektonic forms, 5) there is a mechanistic explanation linking residue and effect, 6) the CV is based on effects, not bioaccumulation, 7) there are few toxicity drivers in the sediment. The WQCTL is not appropriate (i.e., is under protective) when direct exposure to sensitive life stages (e.g., early life stages of fish) is anticipated. A more direct assessment of chronic toxicity via sediment bioassays may be less problematic. See peer reviewer's comment at end of questions regarding ecological receptors.

C. *BCFs reported for fish were used in the calculations of WQCTLs for organics; is this derived level appropriate for setting benthic tissue ecological effects levels? If the fish tissue levels are used, should adjustments be made to the derived levels to reflect the higher lipid contents of the benthic organisms used in the testing program?*

See peer reviewer's response to 8.B.

D. *Are the WQCTLs calculated for metals using bivalve BCFs appropriate for setting levels for polychaetes or vice versa?*

See peer reviewer's response to 8.B.

E. *Are the uncertainty factors applied while deriving ecological effects levels for PAH contaminants appropriate? Does this adequately address the uncertainty around the derived values? Can uncertainty be accounted for using these order of magnitude adjustments? Should they be applied elsewhere to the other risk-based values?*

Uncertainty factors are a legitimate method for quantifying uncertainty. The problem comes when multiple UF are used. Then, quantitative risk estimates are driven absurdly low and the value to decision-making falls precipitously. My preference is to limit UF to extrapolations which have a quantitative basis (e.g., acute to chronic ratio) and provide a narrative description for other uncertainties, especially those with a mechanistic basis. See peer reviewer's response to 8.A for PAH portion of question..

A distinguishing characteristic of environmental risk assessment is its explicit (not implicit) treatment of uncertainty. If EPA/CENAN intends for the Framework to emulate a risk-based approach, it must have specific sections devoted exclusively to uncertainty analysis. An explicit recognition of uncertainty promotes, not hinders, environmental decision-making.

F. *Are the risk values suitable for predicting the significant undesirable effects due to*

bioaccumulation; are there better alternatives for ecological nonspecific risk?

See peer reviewer's response to 2. and comment at end of questions regarding ecological receptors.

G. *If you believe that these values are over- or under- conservative, what do you believe to be an appropriate way to improve them.)*

See peer reviewer's response to 2. and comment at end of questions regarding ecological receptors.

Calculations

9. *Should total PCBs continue to be estimated by doubling the total of 22 congeners or should it be quantified directly using another measure of quantification? What method is most appropriate for sediments in the NY/NJ Harbor area? (Please see Reference No. 60, Table 4-4B)*

Could not evaluate the PCB doubling method with materials provided. Reference 60, Table 4-4B did not describe the method in sufficient detail. The original reference cited in the table (NYSDEC, 1991) was not provided.

10. *Currently, 28-day tissue concentrations of certain organic contaminants are adjusted by some multiplier to estimate the concentrations of those compounds had the exposure been of sufficient duration to allow attainment of steady state levels. (Please see Reference Nos.5 and 46) Are these adjustments appropriate? Should steady state corrections be applied to any other of the listed contaminants? Are there other compounds for which we test that are not expected to approach steady state within the 28-day period?* **This question was not assigned to Dr. Dillon.**

11. *Is the calculation and use of BaP toxicity equivalence an appropriate way to estimate the potential carcinogenicity of PAHs? (Please see Region 2/CENAN joint evaluation memorandum, Appendix for Table 1, Section C.)*

PAH toxicity to aquatic organisms, not a BaP toxicity equivalence approach, is discussed in Appendix for Table 1, Section C. Carcinogenicity is an important endpoint for human health but generally not appropriate for ecological receptors.

12. *Similar to PCBs, only a subset of those PAHS present in New York Harbor are measured for testing evaluation. How should the remainder be considered?*

See peer reviewer's responses to 8.A and 16.

13. *Is the assumption of a trophic transfer coefficient of one appropriate for use in evaluating the potential for human health and ecological impacts associated with metals in Remediation Material? Are the trophic transfer factors calculated for organic compounds correct? (Please see Region 2/CENAN joint evaluation memorandum, Appendix for Table 1, Attachment C.)*

A TTC of 1 is appropriate for metals with the exception of organo-metals which would have higher values. Published TTC values for organics vary with study, chemical and trophic level. The "correct" TTC is one developed for the specific chemical and trophic level of concern.

14. *Is the assumption of a fish consumption rate of 6.5 g/day appropriate for use in evaluating the potential for human health impacts associated with metals in Remediation Material? (Please see Region 2/CENAN joint evaluation memorandum, Appendix for Table 1, Page A-5) Would it be appropriate that the evaluation focus on a higher consumption population?*

6.5 g/day is a standard ingestion rate EPA uses in its risk assessment guidance (RAGS) for Superfund. When available, documented site-specific ingestion rates based on target populations are preferred. (See peer reviewer's comment at end of questions regarding human health.)

General

15. *Is it plausible to replace any other risk assessment assumptions with assumptions specific to the HARS site? (Please see Region 2/CENAN joint evaluation memorandum, Appendix for Table 1, Attachment C and Reference Nos. 88) Is it appropriate to consider the HARS intended use to be factored into an evaluation of effects at the community or population level?*

Yes to both questions.

16. *Is use of the Squibb et al. (1991) report appropriate for identifying the contaminants of concern? Are there contaminants which should be added to or deleted from the list of contaminants for which we presently test? Please see Reference No. 51)*

Squibb et al. is one source which is appropriate for identifying contaminants of concern. The analytes listed in Tables 4-4A and 4-4B of the RIM are also appropriate. Consider EPA method SW 846-8310 to achieve lower detection limits for PAHs.

17. *Should risks from synergistic effects, from exposure to multiple contaminants, be evaluated using results from tissue analyses? If so, how? If not, why not?*

These risks should be evaluated. However, I'm less convinced a residue-based approach is optimal. See peer reviewer's comment at end of questions regarding ecological receptors.

18. *Is test tissue concentration exceeding reference tissue concentration by less than 10X a meaningful evaluative criterion? (Please see page 9 of the Region 2/CENAN joint evaluation memorandum)?* **This question was not assigned to Dr. Dillon.**

19. *Are the studies from which background tissue concentrations were calculated weighted appropriately? If not, what method is recommended? Is the use of the mean the most appropriate measurement of central tendency? If not, what measure should be used? (Please see Reference No. 98) Are the assumption, presented on page 14 pertaining to comparisons of bioaccumulation in test tissue to tissue concentrations in organisms from the vicinity of the remediation site, valid for evaluating undesirable effects?* **This question was not assigned to Dr. Dillon.**

20. *Can baseline tissue concentrations, from appropriate benthic organisms resident to the HARS, be used as standards to determine suitability for Remediation Material as defined above?* **This question was not assigned to Dr. Dillon.**

Peer Reviewer's Comment:

It may be time to rethink the paradigm. The current Framework reflects an evolution in dredged material management which began many years ago with the matrix values. Since that time, EPA Region 2 and CENAN have worked diligently to incorporate appropriate emerging technologies into their dredging program. My observation is that they have done this objectively and with enthusiasm. Affecting these ongoing regional activities was a national dredging program with evolving policies and new Federal regulations. As a result, the Framework today appears cobbled. It is not unlike a single family dwelling which has acquired multiple additions over the years. For the most part, the individual components of the Framework are technically sound. However, it may be time to rethink the paradigm. Here are my suggestions.

1. **Human Health** - Clearly segregate the human health evaluation from ecological risk assessment. The receptors are clearly different, the evaluative tools are different and the public perception is certainly different. I recommend following EPA's risk assessment guidance for Superfund (RAGS). RAGS has well developed toolbox of exposure models and toxicity values for evaluating both

carcinogens and non-carcinogens. RAGS could be "regionalized" and renamed to avoid the potential stigma of Superfund in the dredging program. The EPA/CENAN dredging program would then have a true, risk-based evaluation of bioaccumulation bioassay data specific to human receptors in the New York Bight area.

2. **Ecological Receptors** - The focus here should be twofold; direct toxicity to benthos and toxicity of chemicals entering the food web.

a. **Direct Toxicity to Benthos** - Experience has shown that interpreting bioaccumulation bioassays *vis a vis* toxicity to benthos has been problematic. The multitude of tools and interpretive guidance in the current Framework are symptomatic. It may useful to remember that bioaccumulation bioassays were originally developed to assess *bioaccumulation potential,* not toxicity. The shift to a toxicity interpretation came about as an attempt to assess chronic sublethal toxicity. Since we now have chronic sublethal sediment bioassays transitioning from the research to the regulatory community, we no longer have to jump that chasm. As soon as feasible, EPA and CENAN should incorporate chronic sublethal toxicity tests and de-emphasize the use of bioaccumulation bioassays for toxicity assessments. Their value is in assessing exposure, not toxicity, remains essential.

b. **Toxicity of Chemicals Entering the Food Web** - After direct toxicity to benthos, the risks of sediment-associated chemicals shifts to chemicals possibly entering the food web. As a first step, EPA and CENAN should develop a consensus for what receptors in the food web they are trying to protect. Certain species of finfish (flounder) and crustaceans (lobsters, crabs) come to mind; perhaps certain bivalve species (also humans, see below). Exposure models linking sediment to specific receptors must be develop or refined. Many models currently exist and range in tiers from the simple (BSAF) to the complex (Gobas/Thomann-type kinetics). An exposure dose and/or tissue residue would be the models' output depending on how constructed. Dredged material bioaccumulation bioassays would then be used to reduce model uncertainty with project-specific bioavailabilty information. To complete the assessment, one would couple exposure model output with established toxicity information (dose- or residue-based) to yield estimates of ecological risk. A discussion of the uncertainties would accompany this risk estimate.

The food web models would also accrue significant benefits to the **Human Health** risk assessment described above. Instead of generic estimates of fish and shellfish tissue residues, model predictions would be specific to the New York Bight and to seafood ingestion profiles typical of area residents. With time, persistent bioaccumulative chemicals such as mercury, pesticides, PCBs would likely emerge as the only chemicals posing significant risks.

None of the above ideas are particularly new or innovative. They really just cobble existing technologies into a simpler Framework. The new paradigm, I believe, would simplify and improve how bioaccumulation results are interpreted in the EPA/CENAN dredging program.