N60028_000456
TREASURE ISLAND
SSIC NO. 5090.3.A

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
REGION IX
75 Hawthorne Street
San Francisco, CA 94105-3901

December 18, 1995

Ernesto M. Galang (Code 1832.5EG)
EFA-West
Naval Facilities Engineering Command
900 Commodore Drive
San Bruno, California 94066-2402

Re: Draft Final Phase II Ecological Risk Assessment Work Plan and Field Sampling Plan for Naval Station Treasure Island dated November 8, 1995

Dear Mr. Galang,

The U. S. Environmental Protection Agency (EPA) has received and reviewed the subject document. EPA's comments are enclosed.

If you have any questions, please call me at (415) 744-2383 or Clarence Callahan at (415) 744-2314.

Sincerely,

Rachel D Simons

Rachel D. Simons
Remedial Project Manager
Federal Facilities Cleanup Office

Enclosures

cc: Jim Sullivan, NAVSTA TI
Mary Rose Cassa, DTSC
Gina Kathuria, CRWQCB
Timo Allison, PRC
Clarence Callahan, EPA
H-9-2 File
Admin Record (3 Copies)

DRAFT FINAL PHASE II
ECOLOGICAL RISK ASSESSMENT WORK PLAN
AND FIELD SAMPLING PLAN
FOR NAVAL STATION TREASURE ISLAND (NSTI)
DATED NOVEMBER 8, 1995

General Comments:

A. Before the work plan is revised, EPA recommends that the Navy meet with the Regulators to discuss these comments.

B. Overall, the document has incorporated many of EPA's previous comments. Some of the remaining inconsistencies are related to the on-going discussions for developing the San Francisco Bay specific screening values and the toxicity reference values (TRVs). EPA recommends that these two activities do not delay finalizing the work plan.

Specific Comments:

1. Section 1.2 Modification of Draft Phase II Work Plan, page 2

The fourth bullet is not clear. Will the leachate tests be included in or excluded from the risk characterization process? Please clarify.

2. Section 6.1 Aquatic Assessment Endpoints, page 22

The second paragraph in this section begins,

"The peregrine falcon (representing the raptor guild), the black-belied plover and spotted sandpiper (representing the shorebird guild), and the mollusks, annelids, and crustaceans of the intertidal and subtidal areas were selected as assessment endpoints."

EPA understands that these receptor species will be used to collect data to quantify the measurement endpoints that are logically related to the assessment endpoints. Examples of data included presence-absence, tissue and sediment concentrations, toxicity, sediment grain sizes and water quality factors. Examples of measurement endpoints that the above data address include sample associations (dendrograms), food chain effects, contaminant distributions, acute and chronic toxicity and habitat quality (both substrate and water quality). The suggested assessment endpoints are as follows:

Assessment Endpoints	Measurement Endpoints	Data Requirements
Diversity	presence-absence	density, biomass,etc
Special status species	presence-absence	density, biomass,etc
Food chain effects	tissue concentrations	tissue residues
Habitat quality I	sediment concentrations	sediment chemistry
Habitat quality II	sediment quality	grain sizes,TOC, pH
Habitat quality III	water quality	chemistry, pH, etc
Acute exposure effects	toxicity	bioassay testing
Chronic exposure effects	toxicity	bioassay testing

This approach emphasizes the broader more general, yet identifiable and logical assessment endpoints that are "above" the individual species level where the assessment is more general rather than restrictive. The assessment strategy at the individual level is generally directed at toxicity, whereas the suggested approach is directed at a biological and in some instances closer to an ecological assessment.

3. Section 6.2.1.1 Total chemistry, page 24

Since the ER-Ls are the target values for screening, ER-Ls should be cited as at least the minimum detection limits for the contaminants.

4. Section 6.2.1.2 Measurements of Bioavailability, page 25

Pore water appears to be the focus for the measurement of bioavailability. Although the four bullets contain some truth e.g. bullet #2 "The bioavailability of contaminants depends on their ability to penetrate biological membranes,...", they also contain statements that are best described as non sequitur e.g. bullet #1 "The bioavailability of a contaminant is a more realistic measurement of toxicity than total chemistry." Why not just measure toxicity directly?

If the goal of this effort is to predict toxicity by measuring

bioavailability, then bioassays should be used since they directly measure toxicity. First, a testing strategy should be described that outlines how toxicity will be predicted from bioavailability e.g. what data will be collected and what decisions will be made using these data. Second, a data validation strategy should be designed e.g. what data will have to be validated and on what basis i.e. decision points will the validation be acceptable. One of the most important components of these steps include the literature that support the decisions. Basically, EPA needs to be presented with clear decision points to determine what is toxic and where.

The first assumption in this section should state the importance of pore water in relationship to toxicity. For example, "Pore water is believed to be a better indicator of bioavailability and subsequently toxicity than total bulk chemistry."

What measures of "chemical speciation" in the second bullet will be made to predict bioavailability?

The last bullet is suggesting that bond strength can be inferred, i.e. predicted or ranked, by using differential extraction procedures. Extraction techniques using particular solutions to predict bioavailability are much closer to reality than predictions of bioavailability as a result of "partitioning" estimates using equilibrium methods. EPA suggests an emphasis on measurements e.g. extraction rather than equilibrium modelling.

5. Section 6.2.1.2 Measurements of Bioavailability, page 26

What will data such as total organic matter, grain size in appropriate size classes, pH, Eh and indicators of biological activity such as ATP be used for and what will they tell us? How will these data be incorporated into the assessment?

6. Section 6.2.2 Selected Bioassays, page 27

EPA agrees that replicated acute toxicity tests using the adult amphipod, *Eohaustorius estuarius*, should be used to determine the abundance and quality of food for feeding birds and as a general indicator of habitat quality (along with other sediment characteristics). The pore water test organism, *Strongylocentrotus purpuratus*, seems appropriate for further defining the toxicity of chemicals in sediments and as a specific indicator of the abundance of benthic organisms and a general indicator of habitat quality. EPA believes that the pore water tests can provide a little more than "...information on the toxicity of pore water and will act as a check, on interpretation of acute effects from tests using whole sediment." The pore water test is not only a "check" or validation of the whole sediment test. The pore water test will also provide a general evaluation of sediment toxicity (food production) and habitat

quality (food availability).

7. Section 6.2.3 Tissue Residue Studies. page 29

In addition to the great blue heron and willet, other resident species that should be considered as possible backup aquatic avian assessment endpoints are the mallard, common teal, pied-bill grebe, ruddy duck and snowy egret. In addition to resident species, seasonal birds could be use if appropriate considerations made for the season.

8. Section 7.1 Characterization of Ecological Risk to Benthic Aquatic Receptors, page 30

Beginning with this section, the description in the text should follow Figure 7-1 "Risk Characterization to Benthic Receptors" by explaining each step in the process and the criteria for each decision point.

9. Section 7.1.1.2 Pore Water Chemistry Screening Criteria, page 32

In Figure 7-1, the purpose of the box titled "Compare to San Francisco Bay Water Quality Objectives" is not clear. If the sites are already between the LSV and the HSV, then why do they need to be compared to S.F. Bay Water Quality Objectives? What is to be gained by this comparison?

Please provide a copy of the reference for the San Francisco Bay Water Quality Objectives (RWQCB 1995) to EPA.

10. Section 7.1.1.2 Pore Water Chemistry Screening Criteria, page 32

If pore water chemistry is a check or validation of whole sediment chemistry, then the pore water chemistry and whole sediment chemistry should be performed on split sediment samples not on separate samples from the same location. These two pieces of data should then be used to characterize the sites.

11. Section 7.1.2.1 No Further Action, page 33

The "No Further Action (NFA)" seems to be adequately described.

12. Section 7.1.2.2 Feasibility Study, page 33

The "Feasibility Study (FS)" as described in this section needs to be more clearly defined. The FS is really a continuation of the risk evaluation process. The FS must incorporate all of the available impact and exposure information (the risk analysis) when evaluating the no-action option, the remediation options and the cost-benefit analysis.

13. Section 7.1.2.3 Toxicity Evaluation, page 33

The purpose of the risk assessment process is to define the conditions or the level of the contaminants and their distribution that are directly or indirectly related to the risk of effects to the resources. This is a critical point in the process. To have the Navy, or anyone else arbitrarily decide "...how many times concentration of specific contaminants must exceed these values in order for the sediments from a particular location to undergo bioassay testing." defeats the purpose of risk assessment. This issue must be discussed with the Regulators in the next meeting before the work plan is revised (see General Comment A).

14. Section 7.1.2.3 Toxicity Evaluation, page 34

Bioassay Sediment Holding Times

The Navy must provide the data to show that holding times for sediment can be extended beyond two weeks. Although the Navy cites a conversation with Gary Ankley, no data is provided to justify this position. The citation alone is insufficient to justify an increase in holding time. It is logical that bioassays be performed pre- and post-storage. However, will pore water undergo the same considerations for pre- and post-storage bioassays? How will a decision be made for significant differences between these pre- and post- storage tests? How will the five locations for the pre- and post-storage tests be chosen? What are the criteria for selection of these locations? Will the locations be chosen based on low or high contamination or the presence of metals or organics?

If a relatively fast turnaround time can be obtained for metals e.g. ICP and PAH e.g. SIM analyses, potentially the problem of holding times can be avoided. This presumes that both metals and PAH analyses can be done in substantially less than 8 weeks.

Bioassay Data Analysis and Interpretation

It appears that the use of Johnston et al (1994) is not based on fully documented data. There is little information presented or available that demonstrates the direct or indirect connection between amphipod mortality of 20 percent and "amphipod population ecology."

15. Section 7.1.3 Further Evaluation, page 35

EPA recommends that the following step be added to the process shown in Figure 7-1 as part of the "Is more information needed?" decision point within the "Further Evaluation" phase. After the initial risk analysis, the managers may decide that some locations are still with insufficient information to categorize

them as NFA or FS based on not having fully defined the extent of contamination. The risk assessment up to this point, including bioassays and chemical samples, would have defined the concentrations of contaminants that were significant in terms of effects, however the sampling would not have been sufficient to define the extent of contamination. At this point, the cost of cleaning up the entire area must be weighed against the cost of more sampling for sediment chemistry to delineate the contaminated area.

16. Section 7.2.1 Exposure Assessment, page 38

For an estimate of the dose, the site use factor (SUF) is listed on page 38 and discussed on page 39. It seems reasonable that the dose is composed of estimated on-site and off-site exposures. The on-site exposure being SUF, the off-site exposure being 1-SUF and the total dose being SUF plus 1-SUF. Ambient or background concentration for each contaminant would be the easiest calculation to determine the off-site exposure.

17. Section 7.2.3 Characterization of Risk, page 47

The table showing dose and TRVs seems straight forward, however the application is still poorly defined. For instance, although, EPA agrees that "If all COPCs and assessment endpoints for a site fall into this category (high dose and low TRV), the whole site would be recommended for no further action." But how will the site boundaries be defined? Will Theissen polygons and/or some other process i.e. further sampling be used? This part of the risk assessment falls under the category or phase called "risk characterization" and will take some more discussions and clarification.

18. Section 8.2 Terrestrial Screening Assessment, page 50

The bullet at the top of the page seems to ignore the possibility that the bioaccumulation of some contaminants may be greater than the soil concentration i.e. BCF > 1. This would only be true if there were no such bioaccumulative compounds in the terrestrial setting. Please explain.

19. Section 8.3 Factors Influencing Risk and Its Characterization, page 53

This comment is relevant for both the aquatic and the terrestrial component of this work.

The three components of "risk characterization" include first, the organizational level of impact i.e., individual, population, or community, second, the intensity or severity of effects, how much response (of the receptor) is a result of how much contaminant, and third, the extent of the effect, how wide of an

area are the significant effects observed for the whole site, Treasure Island in this case.

Although the discussion includes the "level" of effects of contamination, it does not include the "intensity" and "extent" of the risk characterization. The material presented is a general discussion of the possible effects to levels of organization e.g., trophic levels, individuals, populations, and communities, however, there is little discussion on "how" the levels will be assessed. What data that are planned for collection will be used to discuss the intensity of effects? How will these data be integrated into the overall risk assessment for the aquatic and the terrestrial environments? What will be the criteria for determining the extent of remediation? How will the Navy determine to remediate certain areas and not others?

20. Section 11.3.1.3 Surface Sediment Analysis, page 63

How will locations where VOCs are suspected be determined?

EVALUATION OF THE NAVY'S RESPONSES TO EPA'S COMMENTS ON INTERIM DRAFT FINAL PHASE II ECOLOGICAL RISK ASSESSMENT WORK PLAN DATED JULY 18, 1995

1. Response to General Comment No. 1, page 21

Although we have had several discussions about the original material presented in the first document, the BTAG request was that a case study be used to evaluate the Navy approach compared to the BTAG approach. Treasure Island and Alameda NAS were two locations suggested as possible case studies. The document states that "The Navy will not present a case study..." The present material is a lot closer to many of our suggestions, however there are some differences which are cited in the above comments. What is the basis for not providing a case study?

2. Response to General Comment No. 2b, page 22

The Navy agrees to determine the screening values "...for sediments using the past 2 years of results from the San Francisco Bay Protection Plan Program" which are the numbers that the BTAG has agreed to use in place of the NOAA numbers from Long and Markel (1992). What is the purpose of using the "San Francisco Bay Water Quality Objectives" for screening also?

3. Response to General Comment No. 2c, page 23

The Navy has agreed to use pore water testing, but will further evaluate the leachate test on 15-20 samples. What is the rationale for this?

4. Response to General Comment No. 3a, page 24

Figure 7-1 does not completely reflects the approach agreed upon in the meeting. The purpose of the San Francisco Bay Water Quality Objectives was not discussed and needs to be clarified.

5. Response to General Comment No. 3b, page 24

Why not perform a TIE at several locations in the TI investigation to determine the exact cause of toxicity rather than the 15-20 samples proposed for a leachate test? TIEs would be more helpful and directly applicable to this work compared to the leachate testing. Again, the citation of a telephone conversation is not sufficient to justify a decision for a technical question. The EPA would like to see data to justify this decision.

6. Response to General Comment No. 3c, page 24

Where are the tests described to evaluate "growth" and "production" as assessment endpoints using polychaetes?