Document title	Issues observed in zoobenthos data extracted from COMBINE
Code	4-1
Category	INF
Agenda Item	4 - Specific issues by biodiversity data type
Submission date	
Submitted by	Finland and Secretariat
Reference	

Background

For HELCOM core indicator "State of the soft-bottom macrofauna community", COMBINE data base should be the source of data and Contracting Parties have been invited to report zoo-, phytoplankton and zoobenthos data on an annual basis to COMBINE and clearly communicate the issues encountered in reporting to the Secretariat and the data host ICES (State & Conservation 9-2018, 3MA.32). However, COMBINE data was supplemented with national data in latest State of the Baltic Sea reporting (HOLAS 2 in 2018) to improve the coverage and because of some observed issues in reported data.

This document outlines some issues observed with the reported data.

Attachment 1 to this document contains examples of reported benthos data.

Action requested

 The meeting is invited to take note of the information and suggest solutions.

Issues observed in zoobenthos data extracted from COMBINE

This document outlines below issues observed with the reported data on zoobenthos and the usability of data for the "State of the soft-bottom macrofauna community" core indicator.

Issues observed:

1. Data from all countries is not available

2. Inconsistency in taxonomy used in reported data

3. Differences in reporting of units and replicate samples
   - For indicator calculation, it is problematic if units and replicate samples are not reported consistently. Some data has been reported per samples, some were calculated to m2 values and some were averages for several replicates. (See attachment 1 for examples).
   - Attachment 1 displays following examples in separate sheets:
     1. An example with three replicate samples and numbers per sample. **This is how the data is most useful for the benthic indicator**.
     2. An example with two replicates, but reported as numbers per square meters.
     3. An example where data is reported as numbers per sample, but the numbers are not integers. It is unclear where this number comes from. It seems like sampled area has been something else than 1000 cm2, but numbers are calculated to 1000 cm2. Assuming that the smallest number here (3.41355) is one individual the sampled area would be 293 cm2. That again doesn't fit with van Veen... Is there information on how many replicate samples are taken in the reported data?
     4. Here the sampled area is 0.1, likely m2. I'm not 100% sure, but likely sampled area should be reported as cm2?
     5. Some other inconsistencies are also found, such as this where probably wrong unit has been reported (?)

Suggestions for improvement:

1. Improve reporting practices / clarify reporting needs and formats

2. Consistent use of WoRMS AphiaID in reporting species taxonomy.
   AphiaID is used in ICES, but not required when reporting?

3. The indicator is calculated per sample, so data need to be reported per sample.
   Species richness is sensitive to sample size (number of replicates), **so averages of replicates cannot be used**. Preferably, the raw data, i.e. actual counts/weight in the sample should be reported. Together with the sampled area (grab size), this is what is needed for indicator calculation.