

TARA OCEANS

Data Release Policy

December 2012

The Tara Oceans Consortium Data Release Policy is based on early, pre-publication release of raw and analyzed data sets according to the following guidelines :

Most data obtained during the Tara Oceans expedition originate from instruments and environmental samples that must be validated to avoid propagation of erroneous information. For that purpose, all samples and data will be validated by the Consortium for quality and unique identification prior to their release. The consortium will make its best to ensure that validation steps are of the highest quality. However the released pre-publication data should still be considered preliminary and may contain errors.

Sequencing data :

-Raw short reads will be deposited to the EBI short read archive and released immediately after validation.

-More advanced data (assemblies, annotations ...) will be released immediately after validation and before publication.

-Improved versions of assemblies and annotations will be released when available.

Oceanographic data :

-Raw (unvalidated) and quality-checked data from all oceanographic instruments will be deposited to the PANGAEA (Data Publisher for Earth and Environmental Science) and released immediately after validation.

-Optical, chemical and biological (including taxonomy) data generated from environmental samples will be deposited to the PANGAEA (Data Publisher for Earth and Environmental Science) and released immediately after validation.

All released data are subject to the following usage policy :

The Tara Oceans Consortium has established a policy for pre-publication data usage with the aim to make quickly and broadly available the global data generated during the project for advancing science, while maintaining a priority for analyses by the laboratories that participate in the data production effort. During one year after release in a public database, the data will be considered usable with the following restriction: specifically, intended analyses may not compete with those reserved for the Tara Oceans Consortium, including global analyses at the genome and community level, and comparisons between samples. Therefore, before starting any study on these reserved data sets, investigators are required to contact the Tara Oceans Consortium to obtain an agreement. This restriction will be lifted after the publication of analyses on the considered data sets by the Tara Oceans Consortium, or automatically one year after the release of the data. If a publication is made

from the released data sets, it is expected that the Tara Oceans Consortium is cited in the acknowledgment section. Raw data sets have the same limits for use than the advanced data sets derived from them. Download of data from the Tara Oceans project implies a full acceptance of the policy usage described in this paragraph.