A picture is worth a thousand data points: Making videos and images from marine environmental monitoring available to all

Markus Lindh, Swedish Meteorological and Hydrological Institute (Sweden), markus.lindh@smhi.se
Patrick Gorringe, Swedish Meteorological and Hydrological Institute (Sweden), patrick.gorringe@smhi.se
Johan Stål, Swedish Meteorological and Hydrological Institute (Sweden), johan.stål@smhi.se

"*A picture is worth a thousand words*" is a common proverb to describe the advantage of using a picture over words in advertisement. In China there is also the famous proverb: "*it is better to see something once, than to hear about it a thousand times*". In marine environmental monitoring, video and images are often the foundation of data, where one video or image can be worth thousands of data points. Yet, in many cases only analysed data is publicly available and the video and image data are kept locally inaccessible for potential users and sometimes even discarded after analysis. There are now also increasing international and national demands to make video and image data publicly available. For example, the INSPIRE directive, FAIR principles and the public sector are putting high demands on marine data collected. Taken together, there is a fundamental gap between data sampling and data access where video and image data from the marine environmental monitoring community needs to be made publicly accessible with good metadata descriptions.

MYEAR	SDATE	LATIT	LONGI	LATNM	SFLAG	COVER%
2019	2020-04-07	56.94000	12.21170	*Zostera marina*	SPP	85
2019	2020-04-07	56.94000	12.21170	*Polysiphonia*	SP	10
2019	2020-04-07	56.94000	12.21170	*Mytilus edulis*	SP	5

Figure 1: A picture of an eelgrass meadow (*Zostera*; top panel) and a simplified mock example of analysed data from the same picture in a table format (bottom panel). In this example, a diver performs regional or national monitoring and aims to observe and identify different species by collecting videos and images. Analysed data (species observations) is sent to the national archive for oceanographic data for quality control and eventually becomes publicly available. However, the original raw data, i.e. images and video that forms the basis of analysed data have up until now not been archived or published.

The Swedish Meteorological and Hydrological Institute (SMHI) is the National Oceanographic Data Centre (NODC; UNESCO/IOC/IODE) and hold the Swedish archive for oceanographic data (commissioned by the Swedish agency for marine and water management; SwAM). We collect, archive, perform quality controls and publish data from national and regional marine environmental monitoring. SMHI also provide data for international data aggregators and regional sea conventions.

SMHI have now also developed a state-of-the-art system for archiving and publishing image and video data from marine environmental monitoring. Here, we will present our solution for a system that is capable of receiving, archiving, quality controlling and packaging data.

By making the video and image data available for the public we aim to:

- Increase data quality -> Data quickly gets outdated where analysis and re-analysis of data will keep data updated so the user has the highest available quality.
- Meet future expected and unexpected needs -> Changes in analysis methods and interpretation of data, including taxonomic re-annotation is problematic where data quickly becomes outdated.
- Utilize emergent technologies (machine learning and other AI methods) -> State-of-the-art video and image analyses can automate workflows and free resources while decreasing human bias.
- Easy data interpretation aimed toward decision-makers -> Stakeholders and decision makers involved in marine monitoring are continuously met by data reports often including complicated tables and graphs of analysed data; easy access to collections of video and image data may be a way to simplify interpretation of environmental change.
- Increase data availability and transparency -> One of SMHIs most important missions is to increase the availability and use of data including its transparency, i.e. making the underlying original raw data that is the basis for the table of analysed data publicly available.

The novel system and data management will be documented and used to create an entry for Ocean Best Practices within the JERICO-RI project and hence contribute to the Ocean Best Practices System Repository (OBPS-R). Video and image data collected at SMHI include monitoring of epibenthos, phytoplankton and megafauna (seals and harbour porpoise). We are using a metadata description applied from the INSPIRE directive and include the use of digital object identifiers (DOI) for published data packages so that data is findable. All data, including metadata, goes through a high-throughput quality control and administration system using software developed in-house at SMHI with open source code (MIT license) and is available both for Microsoft, Linux and Apple users. The data will also be made available to ongoing international projects such as JERICO-RI and for data aggregators such as EMODnet, SeaDataCloud and ICES. We anticipate that the data will be used by the regional sea conventions HELCOM and OSPAR and potentially also the European Environmental Agency.