Data recycling. How to get new information out of existing data?

Everaert Gert[1], Vandegehuchte Michiel[1], Anders Ruus[2], Hjermann Dag øystein[2], Borgå Katrine[3], Green Norman[3], Boitsov Stepan[4], Jensen Henning[5] and Poste Amanda[5]

[1] Flanders Marine Institute (VLIZ), InnovOcean site, Wandelaarkaai 7, 8400 Ostend, Belgium
    E-mail: gert.everaert@vliz.be
[2] Norwegian Institute for Water Research, NO-0349 Oslo, Norway
[3] Department of Biosciences, University of Oslo, NO-0316 Oslo, Norway
[4] Institute of Marine Research, NO-5817 Bergen, Norway
[5] Geological Survey of Norway, NO-7491 Trondheim, Norway

The use of monitoring data in scientific research is often limited to the purpose for which they have been collected. However, integrating different sets of data can significantly increase our ability to investigate and quantify processes and dynamics observed in the marine environment. In this research, we illustrate the relevance of existing data for answering long lasting environmental research questions. We performed a meta-analysis of three existing datasets to infer spatial patterns of anthropogenic chemicals in Norwegian marine sediments. We related these patterns to different potential pollution sources. To do so, we integrated 5036 existing data entries of concentrations of chemicals of the Norwegian Institute for Water Research, the Institute of Marine Research, and the Geological Survey of Norway that were collected at 333 sampling locations between 1986 and 2014. These data were analyzed with regression-based models and we found that atmospheric inputs are the main sources of pollution for most organic chemicals considered. However, north of the Arctic circle, we found that concentrations of polycyclic aromatic hydrocarbons increased from south to north which was most likely related to a combination of coal-eroding bedrock and the biological pump. For metals, a predominantly hotspot-driven spatial pattern was found as historical pollution hotspots (e.g., Sørfjord in western Norway) for mercury, zinc, cadmium, and lead were identified. The knowledge acquired here is essential for developing effective remediation strategies that are consistent with international conventions on pollution control. We have shown that the large amount of marine data that is available gives excellent opportunities to tackle a variety of environmental research questions. As the computational time for transferring environmental data into scientific knowledge has drastically decreased, we are confident that big data analyses will become increasingly relevant for answering long lasting research questions in marine sciences.

Keywords: big data; chemicals; meta-analysis; Norway; spatial patterns