



## **SOCAT - Surface Ocean CO<sub>2</sub> Atlas - a showcase for transparent data management and international collaboration**

B. Pfeil (1,2,3), A. Olsen (4,2), A. Kozyr (5), S. Hankin (6), H. Koyuk (7), D. Bakker (8), C. Sabine (6), N. Metzl (9), and M. Telszewski (10)

(1) University of Bergen, Bergen, Norway (benjamin.pfeil@gfi.uib.no), (2) Bjerknes Centre for Climate Research/Center for Climate Dynamics, Bergen, Norway, (3) World Data Center PANGAEA, Bremen, Germany, (4) Institute for Marine Research, Bergen, Norway, (5) Carbon Dioxide Information Analysis Center, Oak Ridge, USA, (6) NOAA/PMEL, Seattle, USA, (7) University of Washington/JISAO, NOAA/PMEL/TMAP, Seattle (USA), (8) University of East Anglia, Norwich, UK, (9) LOCEAN/IPSL, Paris, France, (10) UNESCO/IOCCP, Paris, France

The Surface Ocean CO<sub>2</sub> Atlas (SOCAT) is a global synthesis of surface ocean carbon dioxide (CO<sub>2</sub>) measurements collected on research vessels, voluntary observing ships and moored as well as drifting platforms. The first public release of SOCAT took place at UNESCO in September 2011. This version consists of 6.3 million quality controlled, uniform format and recalculated surface water fCO<sub>2</sub> (fugacity of CO<sub>2</sub>) data from 1851 voyages in the global oceans between 1968 and 2007 - making it the world largest database in its field.

SOCAT was initiated in 2007. At that time surface ocean CO<sub>2</sub> data were archived in a wide range of formats and at numerous sites around the world, each with its own rules for access, and documentation of the data was frequently poor. This made it virtually impossible to generate comprehensive data synthesis products e.g. for large scale – long term studies.

The international ocean carbon community decided to initiate SOCAT as a community driven effort to assemble, harmonize, quality control and document surface ocean CO<sub>2</sub> data into one open access database. It took four years of hard work by marine carbon scientists and data managers around the world to assemble and quality control the first version of SOCAT.

Two SOCAT products are available via <http://www.socat.info/>:

1) A global data set of recalculated surface water fCO<sub>2</sub> values in a uniform format, which has undergone 2nd level quality control

2) A global, gridded product of monthly mean surface water fCO<sub>2</sub>, with no temporal or spatial interpolation. Individual cruise files with recalculated fCO<sub>2</sub> values can also be retrieved.

The SOCAT products and cruise files can be accessed via the World Data Centre PANGAEA -Data Publisher for Earth & Environmental Science (<http://www.pangaea.de/>) and the Carbon Dioxide Information Analysis Center (CDIAC, <http://cdiac.ornl.gov/oceans/>).

The data products are also available via a sophisticated, online data visualization and analysis tool, called the Live Access Server, and as an Ocean Data View Collection (access via <http://www.socat.info/>).

The methods in SOCAT are fully transparent and documented. The products and cruises are citable through Digital Object Identifiers (doi-s). Cruises in SOCAT have been assigned a standardized cruise identifier called Expocode (a code containing Cruise ID, Year, Month and Day of Cruise), a doi-number and detailed information on the data, so-called metadata (e.g. investigator name, calibrations). Every data point in SOCAT has a link to its cruise of origin via the doi-number giving detailed access to input data and data treatment.

SOCAT is a good example on how to publish global distributed data as a data product while giving credit to the data originator, scientists and data centers involved.