COMMENT

Check for updates

A vision for FAIR ocean data products

Toste Tanhua [1✉], Siv K. Lauvset[2], Nico Lange[1], Are Olsen[3], Marta Álvarez [4],
Stephen Diggs[5], Henry C. Bittig [6], Peter J. Brown [7], Brendan R. Carter[8,9],
Leticia Cotrim da Cunha [10], Richard A. Feely[9], Mario Hoppema [11],
Masao Ishii [12], Emil Jeansson [2], Alex Kozyr[13], Akihiko Murata[14],
Fiz F. Pérez [15], Benjamin Pfeil[3], Carsten Schirnick [1], Reiner Steinfeldt [16],
Maciej Telszewski[17], Bronte Tilbrook [18], Anton Velo [15], Rik Wanninkhof[19],
Eugene Burger[9], Kevin O'Brien[8,9] & Robert M. Key[20]

The ocean is mitigating global warming by absorbing large amounts of excess carbon dioxide from human activities. To quantify and monitor the ocean carbon sink, we need a state-of-the-art data resource that makes data submission and retrieval machine-compatible and efficient.

Human activities such as combustion of fossil fuel, land use change, and cement production increased the atmospheric carbon dioxide (CO_2) concentration to 418 ppm in April 2021. This level is almost 50% higher than at the beginning of the industrial age. The greenhouse effect of atmospheric CO_2 and other gases has led to significant warming and increased stratification in the ocean, and has consequences for ecosystems and marine ecosystem services. Notably, atmospheric CO_2 concentrations would now be around another 76 ppm higher than current levels[1] if the ocean had not taken up a significant fraction of our emissions from the atmosphere[2].

The ocean is one of the largest carbon pools on the planet, second only to the Earth's crust. The ocean contains about 38,000 Gigatonnes of carbon and thereby dwarfs the cumulative emissions of fossil CO_2 since the Industrial Revolution from fossil fuel combustion (about 440 GtC to 2019) and land-use change (about 210 GtC)[1]. As such, the accumulation rate of carbon in the surface ocean of about 1 $\mu mol\,kg^{-1}\,year^{-1}$ driven by anthropogenic CO_2 emissions is much smaller than the natural variations in dissolved inorganic carbon content, over a range of 500 $\mu mol\,kg^{-1}$ regionally and 100 $\mu mol\,kg^{-1}$ seasonally[3]. Thus, any emission-driven trends in ocean carbon concentrations or changes in biogeochemical cycles are expressed amid large

[1] GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany. [2] NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway. [3] Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research, Bergen, Norway. [4] Instituto Español de Oceanografía, A Coruña, Spain. [5] UC San Diego, Scripps Institution of Oceanographyi, San Diego, CA, USA. [6] Leibniz Institute for Baltic Sea Research Warnemünde, Rostock, Germany. [7] National Oceanography Centre, Southampton, UK. [8] Cooperative Institute for Climate, Ocean, and Ecosystem Studies, University Washington, Seattle, WA, USA. [9] Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, Seattle, WA, USA. [10] Faculdade de Oceanografia/PPG-Oceanografia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro (RJ), Brazil. [11] Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany. [12] Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan. [13] NOAA National Centers for Environmental Information, Silver Spring, MD, USA. [14] Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology, Yokosuka, Japan. [15] Instituto de Investigaciones Marinas, IIM – CSIC, Vigo, Spain. [16] University of Bremen, Institute of Environmental Physics, Bremen, Germany. [17] International Ocean Carbon Coordination Project, Institute of Oceanology of Polish Academy of Sciences, Sopot, Poland. [18] CSIRO Oceans and Atmosphere and Australian Antarctic Program Partnership, Hobart, Australia. [19] Atlantic Oceanographic and Meteorological Laboratory, National Oceanic and Atmospheric Administration, Miami, FL, USA. [20] Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ, USA. ✉email: ttanhua@geomar.de

natural variability in these seawater properties across a range of spatial and temporal scales. Accurately quantifying a small change against a large and variable background requires precise and accurate measurements made over decades.

The GLobal Ocean Data Analysis Project (GLODAP)[4,5], initiated in 2004 and subsequently updated[6–8], has been instrumental in delivering carbon-relevant interior ocean data that support well-quantified estimates of the ocean carbon sink. The project delivers near-global data coverage; standardized quality control procedures; a high degree of internal consistency; common data formats; and open and free access to the available data. Compared to its first version, the GLODAP data inventory has more than tripled in size (Fig. 1).

In order to continue to serve its purpose, GLODAP needs to advance both its data ingestion systems and its data extraction systems to become more streamlined and automated. In order to decrease the amount of routine manual work as well as the potential for errors, data submission workflows must become uniform, semi-automated, and compatible with machine-learning techniques for quality control. The data extraction system also needs to accommodate a wider range of filtering to fine-tune requests from users.

Global ocean carbon data

Faced with the challenge of quantifying the ocean's storage of anthropogenic carbon, the ocean community began to systematically measure marine inorganic carbon concentrations in the 1970 and 1980's[4]. These efforts ramped up significantly during the World Ocean Circulation Experiment and the Joint Global Ocean Flux Study (WOCE/JGOFS) during the 1990's, and have later been continued along selected WOCE lines in the repeat hydrographic programs including the Global Ocean Ship-based Hydrographic Investigations Program (GO-SHIP)[9].

The primary focus of GLODAP is synthesizing seawater inorganic carbon chemistry data from these global cruise campaigns. However, data for ocean hydrography, dissolved oxygen, transient tracers, inorganic nutrients, and a range of other variables are included to facilitate interpretation. A unique feature of

GLODAP is the addition of several layers of quality control and adjustments conducted to minimize inconsistencies and biases in the data[10] using a range of tools such as comparison of deep water values at nearby locations. GLODAP offers uniform data at three levels; (1) data from individual cruises in a uniform format with coherent quality control and unit conversion applied, (2) a bias adjusted data product, and 3) a global 1° × 1° mapped climatology[11].

The GLODAP data product has supported more than 2000 articles (and counting) since the year 2000, evidencing its extensive use by the scientific community and the trust placed in it. Seminal contributions on the oceanic anthropogenic carbon content and temporal evolution would not have been possible without GLODAP[2,12,13]. The knowledge from these studies informs, for instance, the Intergovernmental Panel for Climate Change (IPCC) assessments, and the Sustainable Development Goals (SDG) of the UN Agenda 2030 and the Global Climate Observing System (GCOS) indicators on ocean acidification. GLODAP is also an essential reference data set for autonomous observing networks, such as Biogeochemical-Argo: "The long-term success of a global chemical sensor observing system will depend on support from an ongoing, shipboard hydrographic program to produce a high-quality data set for deep waters at the global scale."[14].

With the growth in the amount of data, the ongoing need to provide information on the ocean carbon sink to inform global carbon emission-reduction efforts, and the emerging need to monitor impacts of initiatives in geoengineering and sustainable use of the oceans, the importance of GLODAP will only increase. However, despite receiving short-term funds from a range of projects, GLODAP is a largely unfunded community effort organized and executed by the GLODAP team. Such a situation is unsustainable, and there is significant risk that the effort will diminish or disappear in the next few years. The building and supporting of infrastructure will be critical to ensure that GLODAP continues to provide a valuable service to the global community.
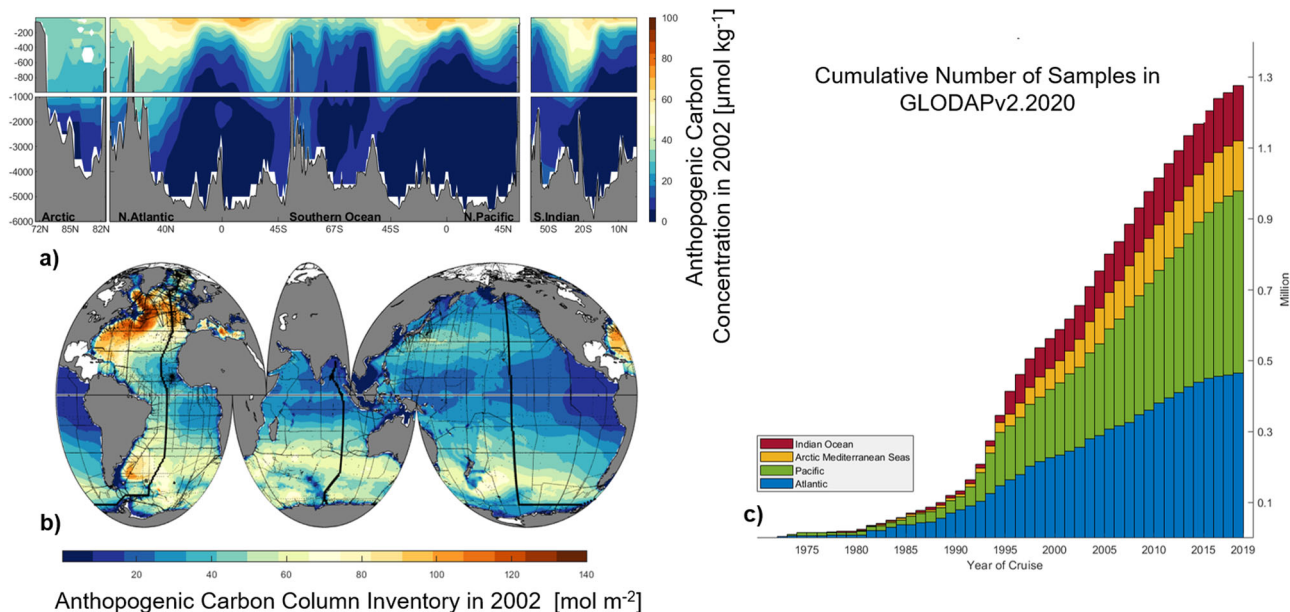

Fig. 1 Key outputs and metrics of GLODAP. a Interior ocean concentration of anthropogenic carbon along a section indicated with a black line in panel (**b**). **b** Integrated column inventory of anthropogenic carbon[22]. Both panels used transient tracer data and the Transit Time Distribution method to calculate anthropogenic carbon[23] content. **c** Cumulative number of samples in GLODAPv2.2020 over time.

Improved efficiency and service

The current GLODAP workflow requires substantial manual work that necessitates dedicated time from, and funding for, data experts, and that introduces opportunities for data handling errors. GLODAP has matured over the last decade with a set of well-documented protocols and development of dedicated software, as well as a backbone of data management support. However, the GLODAP team now strives for advancements on both data input and output, toward a semi-automated system that will reduce the manual work intensity and associated errors.

First, the team aims to implement a uniform, semi-automatic, and standards-compliant data ingestion system that will facilitate the data submission and quality control procedures. This will enable direct interaction with data providers, leading to improvements in data handling, data quality control, and documentation. The envisaged changes will also enable rapid application of novel quality control approaches using machine-learning techniques.

Second, we want to upgrade to a versatile data extraction system. Such a system will provide more flexibility and options to users, such as requesting output with originally submitted data (without adjustments), or only sub-sets of the data in various formats.

These upgrades will streamline repository workflows to insure the data products are FAIR (findable, accessible, interoperable, and reusable)[15], while reducing the burden of data management on scientists. Nevertheless, there will remain a need for experts to spend time on quality control and internal consistency adjustments.

Branch out to keep data accessible

We expect that the improvements will encourage submission of data through building a community of data providers, and will simplify and streamline the process of providing regular updates of the GLODAP products. At the same time, access to GLODAP data will increase. Workflow improvements would allow for enhanced data access systems supporting machine-to-machine services, and better integrated data visualization products[16,17].

The GO-SHIP repeat hydrography effort currently provides the backbone of GLODAP thanks to its high data quality and rapid availability. However, many other datasets reach GLODAP through the extensive network of the GLODAP team; some of these datasets will be functionally lost if not collated by GLODAP. An automated system can aid rescue these data for reuse, by providing a streamlined process for scientists to submit data and metadata, and for users to access and visualize the data.

Upgrades of GLODAP will benefit from the data system that has already been developed for the Surface Ocean CO_2 Atlas (SOCAT)[18]. SOCAT successfully streamlined data submission, quality control, and release of an annual synthesis product, but faces the same resourcing challenges as GLODAP to sustain regular updates. Leveraging an existing, and proven, workflow translates to a significant reduction in both cost and labor of developing a similar system for GLODAP.

An investment for the planet's future

GLODAP needs continued support from the scientific community, but also needs support from funding agencies and stakeholders. Without the updated infrastructure and adequate sustained resourcing in place, GLODAP services may not be able to be maintained on a regular basis.

While the ocean currently takes up about 2.6 Gt of anthropogenic carbon annually, we must understand the evolution, efficiency, and regional patterns of the ocean carbon sink if we want to be able to predict the climate effect of future emissions, as well as to quantify and assess mitigation efforts. Furthermore, human activities affect ocean biogeochemistry in other ways as well, such as de-oxygenation[19], changes in nutrient supply[20], and ocean acidification[21], issues that all need high quality, consistent ocean biogeochemical data to quantify trends, and variability.

Co-located high-quality measurements of physical and biogeochemical parameters that allow for the separation of natural variability from anthropogenic changes—as delivered by GLODAP—are a key component to monitoring, understanding, and mitigating the human influence on the Earth's climate.

References

1. Friedlingstein, P. et al. Global carbon budget 2020. *Earth Syst. Sci. Data* **12**, 3269–3340 (2020).
2. Gruber, N. et al. The oceanic sink for anthropogenic CO_2 from 1994 to 2007. *Science* **363**, 1193–1199 (2019).
3. Broullón, D. et al. A global monthly climatology of oceanic total dissolved inorganic carbon: a neural network approach. *Earth Syst. Sci. Data* **12**, 1725–1743 (2020).
4. Key, R. M. et al. A global ocean carbon climatology: results from Global Data Analysis Project (GLODAP). *Global Biogeochem. Cycle* **18**, GB4031 (2004).
5. GLODAP. https://www.glodap.info/ (2021).
6. Olsen, A. et al. The Global Ocean Data Analysis Project version 2 (GLODAPv2)—an internally consistent data product for the world ocean. *Earth Syst. Sci. Data* **8**, 297–323 (2016).
7. Olsen, A. et al. GLODAPv2.2019—an update of GLODAPv2. *Earth Syst. Sci. Data* **11**, 1437–1461 (2019).
8. Olsen, A. et al. An updated version of the global interior ocean biogeochemical data product, GLODAPv2.2020. *Earth Syst. Sci. Data* **12**, 3653–3678 (2020).
9. Sloyan, B. M. et al. The global ocean ship-based hydrographic investigations program (GO-SHIP): a platform for integrated multidisciplinary ocean science. *Front. Mar. Sci.* **6**, https://doi.org/10.3389/fmars.2019.00445 (2019).
10. Tanhua, T. et al. Quality control procedures and methods of the CARINA database. *Earth Syst. Sci. Data* **2**, 35–49 (2010).
11. Lauvset, S. K. et al. A new global interior ocean mapped climatology: the 1°×1° GLODAP version 2. *Earth Syst. Sci. Data* **8**, 325–340 (2016).
12. Sabine, C. L. et al. The Oceanic sink for Anthropogenic CO_2. *Science* **305**, 367–371 (2004).
13. Khatiwala, S. et al. Global storage of anthropogenic carbon. *Biogeosciences* **10**, 2169–2191 (2013).
14. Johnson, K. S. et al. Biogeochemical sensor performance in the SOCCOM profiling float array. *J. Geophys. Res. Oceans* **122**, 6416–6436 (2017).
15. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
16. webODV. https://hdl.handle.net/20.500.12085/webodv-glodap (2021).
17. digitalearthviewer-glodap. https://hdl.handle.net/20.500.12085/digitalearthviewer-glodap (2021).
18. Bakker, D. C. E. et al. A multi-decade record of high-quality fCO_2 data in version 3 of the Surface Ocean CO_2 Atlas (SOCAT). *Earth Syst. Sci. Data* **8**, 383–413 (2016).
19. Schmidtko, S., Stramma, L. & Visbeck, M. Decline in global oceanic oxygen content during the past five decades. *Nature* **542**, 335–339 (2017).
20. Moon, J.-Y., Lee, K., Tanhua, T., Kress, N. & Kim, I.-N. Temporal nutrient dynamics in the Mediterranean Sea in response to anthropogenic inputs. *Geophys. Res. Lett.* **43**, 5243–5251 (2016).
21. Feely, R. A. et al. Impact of anthropogenic CO_2 on the $CaCO_3$ system in the oceans. *Science* **305**, 362–366 (2004).
22. Lauvset, S. K. et al. Processes driving global interior ocean pH distribution. *Global Biogeochem. Cycle* **34**, e2019GB006229 (2020).
23. Waugh, D. W., Hall, T. M., McNeil, B. I., Key, R. & Matear, R. J. Anthropogenic CO_2 in the Oceans estimated using transit-time distributions. *Tellus* **58B**, 376–389 (2006).

Author contributions

T.T. and S.K.L. co-lead the writing team. S.K.L. and N.L. made the figure, A.O., M.A., N.L., S.D., H.C.B., P.J.B., B.R.C., L.C.C., R.A.F., M.H., M.I., E.J., A.K., A.M., F.F.P., B.P., C.S., R.S., M.T., B.T., A.V., R.W., E.B., K.O.B., and R.M.K. all contributed to the GLO-DAP project and the writing.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.T.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.