



A sea glider being hoisted onto a ship in the Arctic Ocean after taking various measurements, including of conductivity, temperature and depth.

Ocean data need a sea change to help navigate the warming world

Annie Brett, Jim Leape, Mark Abbott & the authors of a High Level Panel for a Sustainable Ocean Economy Blue Paper on *Technology, Data and New Models for Sustainably Managing Ocean Resources*

Open up, share and network information so that marine stewardship can mitigate climate change, overfishing and pollution.

The ocean covers about 70% of Earth's surface, regulates the climate and is home to countless species of fish, a major source of protein for more than one billion people. It is now under threat from climate change, overfishing and pollution.

To respond to these threats, those who use, safeguard and study our seas need real-time information. Too often, ocean management has been undermined by the lack of data on human activity and on the waters themselves. Pirate fishers have plundered the high seas with impunity, knowing they cannot be traced. Crew members on legitimate fishing boats have been tortured and even murdered, out of sight. Stocks have been overfished because most quotas are set only annually, using last

year's data. Illegal fishing has proliferated, devastating ecosystems and undermining global food supplies.

Happily, new technology platforms collected more data on the oceans in 2018 than was gathered during the entire twentieth century¹. Data from satellites, autonomous underwater vehicles and other platforms have come together with emerging data streams from social media, smartphones and low-cost distributed sensors. This enables a new understanding of the impact of human activity on the ocean (see 'Data tsunami').

For instance, fishing vessels worldwide can now be tracked in near-real time using the website Global Fishing Watch (<https://globalfishingwatch.org>). This partnership between Google, the international

Comment

ocean-conservation organization Oceana and SkyTruth, an environmental watchdog in Shepherdstown, West Virginia, uses satellite data to monitor planetary threats. The partnership combines GPS location data from fishing vessels with machine-learning analytics.

Since 2016, it has provided information on activities such as the transfer of fish between intermediate carrier vessels – a technique often used to disguise smuggling. It has also helped to catch boats that illegally dip in and out of marine protected areas. This service is possible because of advances in communication, such as 5G technology for mobile-phone networks, as well as improved capabilities in artificial intelligence and machine learning.

This and other tools combine data from increasingly robust observation networks worldwide. More than 6,000 floating sensors, satellites and other remote-sensing technologies generate a real-time understanding of ecosystems and the risks they face (see go.nature.com/3c8jcs).

Connecting disparate data sets can vastly boost our knowledge. For example, finding and combining existing maps of the ocean floor more than doubled the proportion that has been mapped globally – from about 6% in 2014 to 15% in 2019 – without any new surveys (see go.nature.com/3gthqno).

A major stumbling block to universal data synthesis is ownership. Petabytes of ocean data are under the control of government agencies, researchers and private companies, such as those in oil and shipping². This information must be made available – fast – to enable sustainable management of marine resources.

Here we call for two things. First: federated data networks to connect disparate ocean databases. Second: new incentives and business models for data sharing. These can create an open, actionable and equitable digital ecosystem for the sustainable future ocean. The upcoming United Nations Decade of Ocean Science for Sustainable Development (2021–30) must end data segregation and usher in a new era of automated access for all.

Four problems

The swelling of ocean information in the past decade has not been accompanied by a rethinking of how data are collected, shared and accessed. Historical data-management methods have created a highly fragmented landscape that is resistant to integration. There are four big problems.

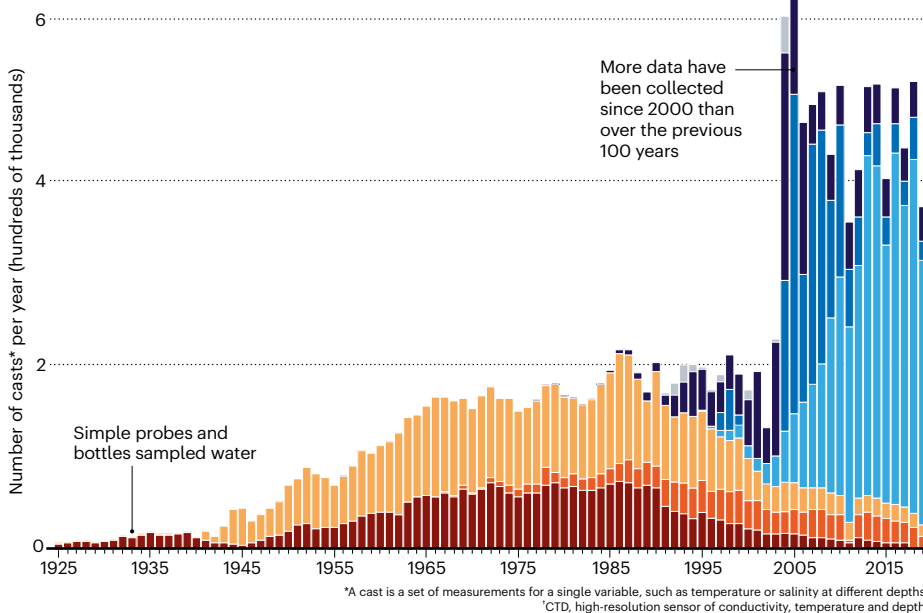
Silos. Government agencies, companies, researchers and resource users keep vast stores of data that are collected and managed for their own specific purposes. These troves are inaccessible and invisible to others. For example, the US Navy holds extensive oceanographic data from areas that are

DATA TSUNAMI

The rapid growth in ocean information in the past decade has not been accompanied by a rethink of how data are collected, shared and accessed. Historical data-management methods prevent a comprehensive understanding of the impact of human activities on the ocean.

Type of instrument:

- Bottle
- Bathythermograph
- CTD*
- Buoy (moored or drifting)
- Attached to a marine animal
- Autonomous vehicle
- Other



SOURCE: WORLD OCEAN DATABASE

rarely reached by research vessels. Private fishing and shipping vessels have reams of information on oceanographic conditions that remains locked away. Silos have serious consequences: illegal fishers, for example, can land their catches unimpeded, knowing that nations don't normally share information on vessel identity or routes. Scientists have few incentives to expend the effort necessary to make their data sets available.

Control. Even when individual data holders realize that their assets might be useful to others, they are often reluctant to share them with centralized repositories, because

“Vast amounts of scientific data collected by defence departments worldwide remain classified.”

they want to control how the information is accessed and used. The private sector keeps its data close, fearing competition or public scrutiny. For instance, aquaculture farms record detailed information on local ocean conditions. They do not share it because of concerns over a backlash from environmentalists about the effects of their operations on nutrient levels and other conditions. Meanwhile, vast amounts of scientific data collected by defence departments worldwide remain classified.

Format and quality. Data are often not interoperable. Inconsistent reporting practices, a lack of funding, concerns over sharing and a lack of attribution in publications have had two effects. First, these problems have hampered community efforts to create universal standards. Second, they have prevented the uptake of portals such as the Ocean Data Standards and Best Practices Project (www.oceandatastandards.org) or the World Ocean Database (<http://wod.iode.org>).

Fragmentation. Attempts to bring data together often drive fragmentation – of data sets, communities and data norms. Centralized catalogues do increase the visibility of data sets, as happened with the Intergovernmental Oceanographic Commission's Ocean Data and Information System (<http://odis.iode.org>). They do not always solve the problem of access. And the proliferation of these lists makes matters worse. There are more than 70 overlapping catalogues for polar ocean data alone¹. Time after time, such bespoke solutions have evolved to meet only the temporary needs of managers and scientists. Scaling up conventional approaches won't work.

Three fixes

Ocean data are dispersed. So are the teams of experts that must make sense of them. These 'many-to-many' networks will evolve as collaborations change. Therefore, new data architectures must enable flexible access,

usage, analysis and cooperation. Here, we outline three key ingredients.

Federated networks. A fast track to interoperability is networking of existing data sets. Global tagging standards and metadata protocols specify when and how data can be stored, transmitted and used, and by whom. They also describe the suitability of the data for management and enforcement decisions. These standards support the connection of disparate repositories through trusted data brokers and they streamline access; data-holders retain control. Data that meet criteria specified in the tags can be made available automatically to users. This provides efficient and timely access for a broad array of managers and users.

Federated networks are used in other fields, such as health, to help overcome confidentiality concerns. In these networks, the information itself is not shared – instead, queries are submitted to gather the needed information while protecting patient privacy.

If users are willing to relinquish some control over storage, data lakes of information held in a raw format can be nodes in larger federated networks. Data lakes move unstructured data onto cloud architecture. This improves access and lowers the costs of analysis. Data lakes can also enable the development of services. For instance, when the US National Oceanic and Atmospheric Administration (NOAA) published data from its Next Generation Weather Radar in the cloud, the information was used to analyse and track bird migrations³.

The UN Educational, Scientific and Cultural Organization (UNESCO) should lead an effort to develop standards for tagging and metadata. The organization should require that all data collected use these protocols and are made publicly available.

Open data. Sharing must be established as a new default unless there are compelling security, proprietary or privacy constraints. New standards could allow industry and military data holders to define data tags that make robust, long-term data sets automatically available after any embargoes have expired.

Governments must lead the way by aggressively declassifying and sharing data that are relevant to ocean science and management. The technical obstacles can now be overcome, as we have set out here. Companies and researchers can be incentivized to share data by making it a condition for access to public resources, such as funds for ocean research, permits for coastal development or licences for oil exploration or fishing.

Companies that sell or process fish and seafood are already moving to introduce greater transparency into their supply chains. For instance, 65 major retailers, processors, marketers, traders and harvesters signed the Tuna 2020 Traceability Declaration in 2018,

committing to full traceability of all tuna products by the end of 2020 (see go.nature.com/3c4wrpv). They are responding to growing consumer demand for better information on the goods they sell and their provenance, legality and social and environmental sustainability. Recognition is growing that better visibility across supply chains allows companies to understand and manage risks. They must build on these efforts.

Researchers must commit to collecting data using standardized protocols and metadata. They can build on the existing standards of the ocean-science community, including the US Integrated Ocean Observing System and Ocean Best Practices system (see go.nature.com/3ebwjtc). Work is needed to integrate these standards and incentivize their adoption.

And what of credit? Assigning DOIs (digital object identifiers) to data sets can ensure that scientists are recognized for contributing to them. Networks that allow researchers to publish their data in one place with global access

“Sharing must be established as a new default unless there are compelling constraints.”

will remove many of the other logistical barriers to sharing. Embargo windows give researchers time to publish scientific findings before the data are made available to all. For example, the UK Natural Environment Research Council requires those it funds to publish all data within two years of the end of data collection.

Business models. The provision of ocean data is an important public good, supporting sustainable management globally. However, the costs associated with standardizing and disseminating data are high: about 10–20% of the budget of oceanographic research projects. Most existing research databases rely on public funding. There is an urgent need for new revenue models that can make data more broadly available. This requires innovation in business models.

As private data sources proliferate, investors, philanthropic organizations and governments should invest in approaches that combine commercial viability with support for data management. For example, private satellite and drone providers already sell data to governments and large companies. They can make these data available in delayed or slightly degraded form to coastal communities and developing countries that can't afford to pay for them.

Businesses should also find ways to incentivize better data collection throughout their supply chains. Thai Union, a major seafood company in Samutsakorn, Thailand, for instance, is piloting the use of a blockchain ledger with a technology company called Fishcoin to reward small-scale fishers in mobile-phone minutes

for collecting and sharing data. Of course, such data must be shared in a useful and actionable format, consistent with the standards used by scientists and governments.

Governments should continue to provide free access to raw oceanographic data for all users. Insurance companies, weather forecasters for precision agriculture, and others use raw ocean and climate data to develop lucrative knowledge products.

Payments for these analytical services create opportunities to support research databases in the long term⁴. NOAA's Big Data Program provides a model. NOAA has partnered with Amazon and other technology firms to put data sets in the cloud so they are freely available to the public. Amazon gains insights into future knowledge-based services by analysing data-usage patterns. In return, NOAA has seen access to some data more than double, increasing their utility at no cost to the government.

However, it is risky to rely on public-private partnerships to reduce the costs of network infrastructure. Companies such as Amazon could decide that hosting ocean data is no longer economically viable, or a firm might disappear from the market. Moreover, such partnerships can be significantly influenced by financial and political pressures beyond the scientific and stewardship needs of the data. Public entities can protect against these eventualities by partnering with a variety of companies, and by ensuring that data remain backed up on publicly funded servers.

Global coordination and commitment is needed. The UN Decade of Ocean Science for Sustainable Development is an opportunity to revolutionize how ocean data are collected, stored and used.

The authors

Annie Brett is a fellow at the Center for Ocean Solutions, Stanford, California; and at the World Economic Forum Centre for the Fourth Industrial Revolution, San Francisco, California, USA. **Jim Leape** is a fellow at and co-director of the Center for Ocean Solutions at the Stanford Woods Institute for the Environment, Stanford, California, USA. **Mark Abbott** is president and director of Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA. A full list of the authors of a High Level Panel for a Sustainable Ocean Economy Blue Paper on *Technology, Data and New Models for Sustainably Managing Ocean Resources* accompanies this Comment online (see go.nature.com/3cvtczd). e-mail: abrett1@stanford.edu

1. Tanhua, T. et al. *Front. Mar. Sci.* **6**, 440 (2019).
2. Pendleton, L. H. et al. *ICES J. Mar. Sci.* **76**, 1415–1423 (2019).
3. Vance, T. C. et al. *Front. Mar. Sci.* **6**, 211 (2019).
4. Organisation for Economic Co-operation and Development. *Business Models for Sustainable Research Data Repositories* (OECD, 2017).