

**Credit Where Credit is Due: Authorship of Open Ocean Data
Workshop Report**

**6 - 7 October 2022
Ditchley Park, UK**

By: Corinne Bassin, Allison Miller, Jyotika Virmani, Icarus Allen, Louis Demargne,
Liz Greenhalgh, Katrin Linse, Graham Moore, Ralph Rayner, Graham Smith,
Martin Smith, Helen Snaith, Fabio Trincardi, Martin Visbeck

Table of Contents

Background	2
Event	2
Recommendations	3
Cross-Sectoral Recommendations and Actions	4
Change the Value Proposition of Data Repositories	4
Persistent Identifiers (PIDs)	5
Cross-Sector Training & Funding Opportunities	5
Conferences & Meetings	5
Sector-Specific Recommendations and Actions	6
Publishing	6
Academia	7
Data Centers	7
Funders	8
Industry	9
UN Decade of Ocean Science for Sustainable Development/ Intergovernmental Oceanographic Commission	9
Appendix A: Workshop Programme	10
Appendix B: Workshop Participants	12

Background

In recent years the shift towards rapid, large-scale open sharing of data has slowly gained traction among the science community. Organizations such as [UNESCO](#), the [U.S. National Academies of Science, Engineering and Medicine](#), and [NASA](#) suggest that the next few years will see a renewed effort towards open data sharing.

The benefits of openly sharing data have been enumerated—shares knowledge, enables more widespread and efficient use of the data, enhances scientific integrity, increases confidence, solves multiple problems, increases the value-to-cost ratio, and enables scientific research to continue beyond the researcher’s career, to name a few. The quicker data is made available, the faster understanding and knowledge are gained, and action can be taken. As seen during the coronavirus pandemic, the [rapid dissemination of data and knowledge](#) led to the widespread development of vaccines and care in a timely and world-saving manner.

There are still [challenges](#) to openly sharing data—data needs to be [findable](#) and accessible, [many say they’ll share data but don’t](#), and [institutional policies and training are outdated](#). Sharing data effectively takes both time and financial resources. Delaying the accessibility of data until, after, or simultaneously with a scientific publication reduces timely access to data.

Researchers may be more inclined to exert the effort to share their collected data if contributing data is acknowledged and valued similarly to publications. The change will occur with a cultural shift to recognizing data sharing as a form of productivity and giving proper credit to those that contribute.

Event

Schmidt Ocean Institute, in partnership with The Ditchley Foundation, hosted *Credit where credit is due: Authorship of open ocean data* at Ditchley Park, in Chipping Norton, UK, on October 6-7, 2022, to identify actionable and implementable solutions to recognize and reward the dissemination of acquired data and knowledge (Appendix A). The workshop [gathered 13 experts](#) (Appendix B) from around Europe with various perspectives on the data pipeline: funding, data access, publishing scientific knowledge, and academic research. The workshop conversation focused on practical steps that could be taken to enable better data sharing and how to shift the values and culture.

The Schmidt Ocean Institute is a philanthropic foundation established in 2009 by Eric and Wendy Schmidt to boldly explore our unknown ocean through advanced technologies, intelligent observations, and the open sharing of information. A central principle that guides Schmidt Ocean Institute is the rapid and open sharing of acquired scientific observations, data, and knowledge to stimulate the growth of data applications and amplify further ocean exploration, discovery, understanding, and management.

The Ditchley Foundation, a privately-funded charity founded in 1958 by Sir David Wills, aims to host conversations that build new common ground to move toward solutions to complex problems. A central tenet of Ditchley is that change is driven and managed by personal relationships, not reports or declarations, and therefore it engages the local and global communities to create space and time for deep reflection, sharing of ideas, and making new connections.

In 2019, Ditchley hosted [*Transforming Ocean Data*](#), a workshop in partnership with the Ocean Data Foundation, to address how ocean science data can be more effectively shared. The report discussed challenges faced by academic cultures and publishing practices that create incentives for data hoarding and slow data release, thereby reducing data use and data-to-knowledge transfer.

Recommendations

This workshop provided a rare opportunity for this blend of expertise to be gathered in one place to holistically discuss open data sharing from end-to-end and the life cycle of gaining credit for data contributions - identifying and analyzing the challenges and solutions as data flows from academic collection point to data centers and publications, which feeds back into academic credit and funding.

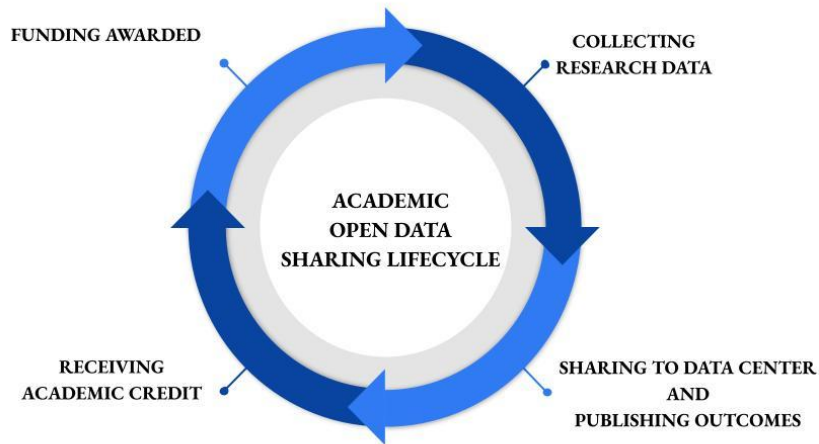


Figure 1: Open data sharing credit life cycle, from academic collection point to data centers and publications, resulting in academic credit and more funding awarded.

The conversation included cross-sectoral ideas on how to collaborate with the goal of improving the data sharing pipeline. Fundamentally, a collective mindset shift will be required, as people are at the heart of this cultural change. Cross-sectoral and sector-specific recommendations and actions are listed below.

Cross-Sectoral Recommendations and Actions

Change the Value Proposition of Data Repositories

Data repositories should be visionary leaders in how to better serve and maximize the use of data, thereby adding to their value proposition and aiding in the scientific literacy of a broader public.

The term “data repositories” inherently implies, to some, that the data is intended to be archived or deposited and not accessed. This notion creates a false impression and a perceived, albeit subtle, barrier to data usability. To overcome this, we suggest the use of the term “data centers” or “data access centers.” Throughout this report, data centers refer to locations where data is archived, deposited, and accessed. Some workshop participants have committed to using this vocabulary going forward.

Persistent Identifiers (PIDs)

Digital Object Identifiers (DOI) are relatively common; however, they are only one option for tracking data use. DOIs are a specific form of PIDs that form part of the traceability solution. Although less prevalent, new frameworks for PIDs are emerging that offer better ways to track the complex life cycle of a dataset and provide a mechanism to give credit to those that contribute to publishing the dataset. Within a well-managed framework that captures elements such as version control, more extensive use of PIDs could allow for:

- traceability of the data, including in publications,
- visibility into changes or updates to the data,
- improved traceability of subsets of the data,
- preservation of connections to the metadata and original dataset,
- identification of all contributors and authors of the dataset,
- sharing of where and how the data is used.

Cross-Sector Training & Funding Opportunities

Resource support for data management, which includes cleaning, curation, analysis, archiving, and dissemination, is crucial. Cross-sector training opportunities and resource and fiscal support for data management throughout a project are necessary to ensure robust, open, and rapid data sharing and access. For example, workshops are periodically offered by the [Marine Environmental Data and Information](#) Network to improve the uptake, knowledge, and use of marine data.

Good Data Management Plans are key to sharing data well. As such, funders and data centers can work together to publish best practices for writing a data management plan in the Ocean Best Practices repository, giving proposers examples and references for what makes a good data management plan.

Conferences & Meetings

During the workshop, it became apparent that participants from each sector involved in data management and collection were siloed. One recommendation was to encourage conversations at conferences as one way to tackle existing challenges and identify solutions together. For example:

- Oceanology International is a forum to engage with the sensor manufacturers and technology developers, in particular, to have discussions on how to hardwire metadata and identification data into sensors - the mechanism by which data is collected;

- UN Decade of Ocean Science for Sustainable Development meetings provide a chance for the Data Coordination Group and the Technology & Innovation Working Group to work together to discuss the connectivity between the start and end of the data pipeline;
- Research Data Alliance conferences provide venues to meet with data sharing centers to promote best practices not just within the domain, but also across other fields of data expertise;
- The International Ocean Data Exchange conference is a place to engage with all four sectors, all of whom should be invited to implement and identify additional recommendations.

Sector-Specific Recommendations and Actions

Publishing

Recommendations for Publishing	Actions for Publishing
Publishing guidelines require that oceanographic data be publicly available prior to, or along with, publishing results and publications.	Provide a timeline for implementing open sharing data formats.
	Provide a list of mandated or recommended data centers. Examples include Nature’s journal Scientific Data and NERC’s Data Catalogue Service .
Empower peer reviewers and editors to enforce Publisher’s rules/guidelines.	Peer reviewers or trained staff check that data is publicly accessible at the time of peer review prior to approving publication.
Recognize and publish datasets as formal publications.	Encourage more “data papers” or establish data journals so that datasets are published and cited.
Provide an avenue for primary authors to acknowledge everyone contributing to the publication, including the data and data management.	PIDs, including but not limited to DOIs, should be used to connect all data contributors to the dataset and publication.

Academia

Recommendations for Academia	Actions for Academia
Provide institutional support (akin to institutional communication support).	Provide institutional computing infrastructure and user support.
	Hire institution-wide experts to help scientists write good data management plans when applying for funding.
Value the sharing of data as a contribution to the greater scientific community.	Consider an individual's data contributions when evaluating promotions or career opportunities.
	Create and share stories about data use and availability.
	Use PR or marketing resources to tell data stories better.
Ensure scientists and staff have the appropriate data curation and stewardship skills.	Encourage oceanographers to take courses and training in data stewardship or management.
	Teach oceanography to data experts, enabling them to solve problems in the field using their expertise.

Data Centers

Recommendations for Data Centers	Actions for Data Centers
Increase data accessibility.	Change perception by changing vocabulary to data centers versus data repositories (mentioned above).
	Use web-based technologies to eliminate the deterrent of cumbersome request forms.
Convey the value of a data center to increase support resources.	Implement web-based technologies to track digital information on usage.

	Gather and release metrics into how data from data centers are accessed, used, by whom, and why.
Ensure traceability throughout the data lifecycle.	Develop standards for the effective implementation of PIDs.
	Host workshops on PIDs and how they are used for data contributors.
Improve the ability of researchers to learn about data center requirements, capabilities, and how to submit quality data in accepted standard formats and access data.	Clearly emphasize that data cannot be archived, downloaded, or used in a data center without appropriate standard formats and metadata.
	Make support tools or code publicly available so that contributors are able to format their metadata and analytics.
Clearly communicate metadata needs and standard formats.	Reject data if metadata is not attached and intact.
Reduce multiple versions of datasets.	Work with researchers to reduce the need to download data to local nodes by providing access to cloud computational services that allow analysis to take place remotely where the data is stored.
Increase confidence in the quality of data from data centers.	Use accreditations or certifications, such as The CoreTrust Seal, to validate the work of data centers.

Funders

Recommendations for Funders	Actions for Funders
Provide resources to promote good Data Management Plans (DMP) when submitting proposals for funding.	Provide templates for submitting a DMP.
	Include examples of good DMP practices when hosting proposal writing workshops or webinars.

Support data management activities post-collection, such as cleaning, curation, analysis, archiving, and dissemination.	Provide funding for some data stewardship activities.
Hold individuals and institutions accountable to deliver data as committed.	Consider data sharing track record when assessing funding for future scientific research.
	Score/Rank institutions based on their data release and sharing and consider posting the scores/rankings publicly.

Industry

Recommendations for Industry/Others	Actions for Industry/Others
Reduce the likelihood that sensor metadata will get lost during moving, manipulating, cleaning, and analyzing the data collected by a specific sensor.	Attribute a PID to sensors, ensuring metadata cannot be lost.
	Hardwire metadata, including calibration, information into sensors.

UN Decade of Ocean Science for Sustainable Development/ Intergovernmental Oceanographic Commission

Recommendations for UNDOS	Actions for UNDOS
Obtain commitment from publishers that by the end of the Decade, all data will be openly available before publication. This commitment is vital to unlocking knowledge about the ocean and would be a revolutionary legacy of the Ocean Decade.	Host a meeting with publishers to agree on this commitment by the end of the UNDOS.
	Release a declaration of the commitment by the end of the UNDOS.

Appendix A: Workshop Programme

Credit Where Credit is Due: Authorship of Open Ocean Data

A workshop hosted by Schmidt Ocean Institute

In partnership with Ditchley

6 - 7 October 2022

October 6, 2022

11:30 am Arrive at Ditchley Park / **Coffee available upon arrival (Saloon)**

12:15 pm Welcome by James Arroyo, Director, Ditchley Foundation & Jyotika Virmani, Executive Director, Schmidt Ocean Institute and Participant Introductions (Library)

1:00 pm Lunch (Dining Room)

2:15 pm SESSION 1: Setting the stage for the challenge (Library)

- *Prof. Ralph Rayner, Editor Journal of Operational Oceanography*
- *Dr. Helen Snaith, Senior Scientific Data Manager, National Oceanography Centre & Head of the Global Data Center, Seabed 2030 Project*
- *Prof. Dr. Martin Visbeck, Head of Research: Physical Oceanography, GEOMAR*
- *Dr. Jyotika Virmani, Executive Director, Schmidt Ocean Institute*

4:00 pm Afternoon Tea (Saloon)

4:30 pm SESSION 2: Brainstorming the future of how supporting institutions can incentivize scientists to share their data (Library)

6:00 pm Short Break

6:15 pm Analysis and ranking of ideas

7:00 pm Pre-Dinner Drinks (Saloon)

7:30 pm Dinner (Dining Room)

9:00 pm Tour of Ditchley House

October 7, 2022

8:00 am Breakfast (Dining Room)

9:00 am SESSION 3A: Exploring the practicalities of implementation of ideas - timeframes, requirements, and solutions (Library)

10:30 am Group photo and coffee break (Saloon)

11:00 am SESSION 3B: Exploring Solutions (Library)

What incentives are needed to make these ideas reality? Who is responsible for providing these incentives? What are the metrics?

12:30 pm Lunch (Dining Room) and time for a nature walk

2:00 pm SESSION 4: Moving forward, Recommendations, Outcomes, Next Steps (Library)

4:00 pm CLOSE OF WORKSHOP

Appendix B: Workshop Participants



Name	Title	Organization	Area
Icarus Allen	Chief Executive	Plymouth Marine Lab	Academia
Corinne Bassin	Data Solutions Architect	Schmidt Ocean Institute	Funding Agency
Louis Demargne	Data and Knowledge Management Officer	UNESCO/IOC	Data
Liz Greenhalgh	Head of Impact and Learning	Ditchley Foundation	
Katrin Linse	Chief Editor of Marine Evolutionary Biology, Biogeography, and Species Diversity / Senior Biodiversity Biologist	Frontiers in Marine Science / British Antarctic Survey	Publishing
Allison Miller	Research Portfolio Senior Manager	Schmidt Ocean Institute	Funding Agency
Graham Moore	Chief Data and Product Officer	HUB Ocean	Data

Ralph Rayner	Editor-in-Chief	Journal of Operational Oceanography	Publishing
Graham Smith	Open Data Programme Manager	Springer Nature	Publishing
Martin Smith	Chief Technology Officer	Ditchley Foundation	Data
Helen Snaith	Head of the Global Data Center / Senior Scientific Data Manager	SeaBed 2030 Project / UK National Oceanography Centre	Data
Fabio Trincardi	Director	Scienze del Sistema Terra e Tecnologie per l'Ambiente, Consiglio Nazionale delle Ricerche	Academia
Jyotika Virmani	Executive Director	Schmidt Ocean Institute	Funding Agency
Martin Visbeck	Head of Research Unit: Physical Oceanography	GEOMAR	Academia