Volume 33 • No. 1 • Winter 2019

The Gulf of Mexico Research Initiative Information and Data Cooperative: Data Transparency and Data Sharing + Classroom Activity

BY SANDRA ELLIS AND KATIE FILLINGHAM

- Data are essential to the scientific process; they enable scientists to examine results of their experiments when exploring new hypotheses. Data sharing and data transparency within the scientific community are relatively new practices that have many potential benefits; as the cost of doing science increases, it can promote continued scientific investigations when funding is tight by reducing duplication of effort. However, standards for requiring data sharing and establishing mechanisms to openly access data sets are still in development.

- When the Gulf of Mexico Research Initiative (GoMRI) was established, the Master Research Agreement required that all data collected through GoMRI funding must be made publicly available, and the program does so through the Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC). GRIIDC is leading the way as a successful model for promoting data sharing, data compliance, and data standardization.

- A classroom activity provides examples of publications with varying levels of data access and encourages students to discuss the benefits and challenges of data transparency and open access to data.

DATA TRANSPARENCY AND DATA SHARING – BACKGROUND

Publishing scientific results in peer-reviewed journals is considered to be the cornerstone of transparency in science. Ideally, results are published with enough detail to allow reproduction, which in theory allows the scientific community to verify results (McNutt et al. 2016). Scientists may then be evaluated based on the number of papers published, the impact factors of journals that they have published papers in, and the number of citations to their published works (Abbott et al. 2010). Despite the fact that it is the norm for scientists to share their work and results through publications, the effort to make the underlying raw data openly available is an emerging trend.

In 2013, the United States Office of Management and Budget (OMB) issued an executive order mandating an open data policy for federal agencies. To comply with this directive, federal grant funding agencies have established and implemented data management and sharing requirements for grantees. Many scientific journal publishers have adopted requirements towards open data availability (Goldstein et al. 2017). These publishers require that the raw data used to generate the results presented in the peer-reviewed manuscript be openly available so that results can be verified.

Data sharing has many documented advantages, such as increasing recognition for scientists and increasing scientific transparency (McNutt et al. 2016; Piwowar et al. 2007; Belter 2014; Costello 2009). To comply with the requirement to share data, scientists may reference their datasets in the methods, acknowledgments, or references sections of peer-reviewed publications. Certain publishers also permit inclusion of datasets as part of the supplementary information in the journal itself (Lawrence et al. 2011). However, there is no standard method available to authors to reference data that they have collected and shared (Lawrence et al. 2011).

While data sharing has documented advantages and federal grant funding agencies and some publishers have embraced it as a requirement, the scientific community has been reluctant to adapt to this change. Researchers express concerns that they will not receive credit for publicly available datasets or that datasets will be used incorrectly (Costello 2009). They also report insufficient time to publish their research before having to make the data publicly available and a lack of funding required to make data electronically available as two major barriers to data sharing (Tenopir et al. 2011).

To overcome reluctance to share data and promote a culture of data sharing, a number of incentives have been proposed. The main recommendation is to make data a citable research object, similar to papers (Lawrence et al. 2011; Costello

2009). In practice, required data sharing by publishers has been found to be an effective approach to increase the availability of underlying data (Vines et al. 2013).

Sharing data and encouraging scientists to use existing data sets can help continue the development of good science even when science funding is tight and the cost of conducting research is constantly increasing.

IMPLEMENTING THE GOMRI DATA SHARING POLICY

One of the founding principles of the Gulf of Mexico Research Initiative (GoMRI) program, as outlined in the Master Research Agreement (MRA), is that all data acquired through GoMRI funding must be archived and made publicly available. GoMRI's commitment to data sharing in a timely manner was at the forefront of data sharing requirements when it came into place in 2011 (Gibeaut 2016). GoMRI quickly realized that there were no existing resources to meet the data sharing needs of both GoMRI administration and researchers, and therefore entered into an agreement with the Harte Research Institute for Gulf of Mexico Studies at Texas A&M University-Corpus Christi to establish the Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC, https://data.gulfresearchinitiative.org/).

Advised by the GoMRI Research Board Data Management Subcommittee, GRIIDC is a team of researchers, data specialists, and computer systems developers who are supporting the development of a data management system. The GRIIDC data management system is a part of the GoMRI Legacy to ensure that all of the information collected by GoMRI scientists will be discoverable and usable by the science community, responders in the event of future oil spills, and the general public long after the program ends in 2020. As of 2018, GRIIDC houses over 2,200 data sets from over 282 research groups and 2,600 scientists and continues to grow every day.

To promote a culture of data sharing within the GoMRI community, a comprehensive approach is used, as outlined in Gibeaut 2016. User-friendly online tools and extensive user training help ease technical barriers that may prevent data from being shared. Disseminating data to national archives, when appropriate, increases research visibility. Publishing digital object identifiers (DOIs) for each dataset provides an internationally recognized tool for citation to further incentivize data submission to GRIIDC. GRIIDC recommends the use of a standard citation, which includes this DOI, if data are used by others in scientific research.

Since there is no standard way within the scientific community to acknowledge, or include an attribution, for data, GoMRI developed a standard attribution statement. In January 2016, GoMRI grantees were asked to use this statement to highlight their publicly available datasets associated with peer-reviewed publications in the acknowledgments section of manuscripts. This standard attribution statement is:

> *Data are publicly available through the Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC) at https://data.gulfresearchinitiative.org (doi:<doi identifier> [, <doi identifer2>, <doi identifier3>, …].*

The purpose of this standard attribution is to make sure that the author receives credit for data in the GRIIDC system and to confirm that authors are fulfilling the GoMRI requirement to make data openly available at the time of publication. Since January 2016, over 98% of 420 published articles that reported results using data acquired with GoMRI funding have included the standard attribution statement that acknowledges publicly available data.

CLASSROOM ACTIVITY RESOURCE

To help illustrate the benefits of open access to data from science publications and the ease through which this information can be accessed through sites like GRIIDC, please consider the following activity (recommended for grade levels 9-12).

Imagine you are a scientist studying the immediate effects of a recent oil spill and evaluating possible long-term impacts. In particular, you are interested in what may happen to the oil in the deep-sea sediments and/or coastal sediments. During review of the existing literature, you find several articles that are openly available online, which is very helpful, but you determine that you require raw data to assist with your task. You are able to find an article published through GoMRI with links to openly available data sets through GRIIDC. It's a good start!

1. Brooks, G.R., R.A. Larson, P.T. Schwing, I. Romero, C. Moore, G.-J. Reichart et al. (**2015.**) Sedimentation Pulse in the NE Gulf of Mexico following the 2010 DWH Blowout. *PLoS ONE*, 10(7): e01323410. doi.org/10.1371/journal.pone.0132341. data.gulfresearchinitiative.org/data/Y1.x031.000:0001, data.gulfresearchinitiative.org/data/Y1.x031.000:0002, data.gulfresearchinitiative.org/data/Y1.x031.000:0003.

In continuing to collect information for your literature review, you find several other publications that are openly available, but they have many different data availabilities:

2. Tarr, M.A., P. Zito, E.B. Overton, G.M. Olson, P.L. Adhikari, and C.M. Reddy. (**2016.**) Weathering of oil spilled in the marine environment. *Oceanography*, 29(3): 126-135. doi.org/10.5670/oceanog.2016.77.

3. Graham, W.M., R.H Condon, R.H. Carmichael, I. D'Ambra, H.K. Patterson, L.J. Linn, and F.J. Hernandez Jr. (**2010.**) Oil Carbon entered the coastal planktonic food web during the Deepwater Horizon oil spill. *Environmental Research Letters*, 5: 045301. doi.org/10.1088/1748-9326/5/4/045301.

4. North, E.W., E.E. Adams, A.E. Thessen, Z. Schlag, R. He, S.A. Socolofsky, S.M. Masutani, and S.D. Peckham. (**2015.**) The influence of droplet size and biodegradation on the transport of subsurface oil droplets during the Deepwater Horizon spill: A model sensitivity study. *Environmental Research Letters*, 10(2). doi.org/10.1088/1748-9326/10/2/024016.

5. Edwards, B.R, C.M. Reddy, R. Camilli, C.A. Carmichael, K. Longnecker, and B.A.S. Van Mooy. (**2011.**) Rapid microbial respiration of oil from the Deepwater Horizon spill in offshore surface waters of the Gulf of Mexico. *Environmental Research Letters*, 6(3). doi.org/10.1088/1748-9326/6/3/035301.

6. Yin, F., J.S. Hayworth, and T.P. Clement. (**2015.**) A tale of two recent spills - comparison of 2014 Galveston Bay and 2010 Deepwater Horizon oil spill residues. *PloS ONE*, 10(4): e0124645. doi.org/10.1371/journal.pone.0124645.

7. Geng, X., Z. Pan, M.C. Boufadel, T. Ozgokmen, K. Lee, and L. Zhao. (**2016.**) Simulation of oil bioremediation in a tidally influenced beach: Spatiotemporal evolution of nutrient and dissolved oxygen. *Journal of Geophysical Research: Oceans*, 10(4): 2385-2404. doi.org/10.1002/2015JC011221.

ACTIVITY AND QUESTIONS FOR STUDENTS

Students can work independently or in pairs. Ask students to review the above papers to find data attributions or acknowledgments. Data attributions may be in the methods, acknowledgments, or references. For each of the publications identified, what is their data access like? (*See right column for answers.*) Then discuss how easy, or difficult, it might be to obtain data for each publication based on the attribution or lack thereof. How does including an attribution affect the ability to find data? What do students think is the best format for referencing data? If there are no data referenced, how do students think that the data could be obtained?

Have the students find and download the datasets, and recreate some of the plots from the first article (#1) with the published data sets. The Excel spreadsheet "Brooks Larson TC 2015-06-09.xlsx" found in GRIIDC dataset Y1.x031.000:0002 (doi: 10.7266/N70K26H7) has data that is used to generate the "Texture" plots found in Figures 2, 3, and 4 in the associated paper. The figures in the paper present the data based on the depth of the core in centimeters, while the raw data provided uses depth in millimeters. Students will have to convert millimeters to centimeters and remove points to recreate the plots. (Note: the sample IDs in the paper do not directly match the sample IDs in the raw data; for example, P-06 in the paper corresponds to PCB-06 in the data, D-08 corresponds to DSH-08, and D-10 corresponds to DSH-10.) Have the students describe what their plot shows.

Discussion: In what other ways might the data provided be used? Are there other scientists from a different field who might be interested in this data and if so, how could they use it?

Summary: As a class, discuss why data sharing is important, why scientists may be reluctant to share data, and approaches that could encourage data sharing. How can the ability to reuse research results be impacted by data availability?

(Answers: 2. Data in GRIIDC referenced but not with direct link; 3. No data reference; 4. Link to the data, but the link is broken; 5. Some data available through NOAA; 6. States that all relevant data are available in the article; 7. Data available upon request by contacting the author.)

ACKNOWLEDGMENTS

In addition to the acknowledgments for the full issue, the authors of this article thank Gregg Brooks for permission to reference the datasets included as a part of the activity in this article. This article was made possible by funding from GoMRI to the Consortium for Ocean Leadership and GRIIDC. The Brooks et al. 2015 datasets listed in this article are publicly available through GRIIDC at: https://gulfresearchinitiative.org (data.gulfresearchinitiative.org/data/Y1.x031.000:0001, data.gulfresearchinitiative.org/data/Y1.x031.000:0002, data.gulfresearchinitiative.org/data/Y1.x031.000:0003).

REFERENCES

Abbott, A., D. Cyranoski, N. Jones, B. Maher, Q. Schiermeier, and R. Van Noorden. (**2010.**) Do metrics matter? *Nature*, 465: 860-862. doi:10.1038/465860a.

Belter, C.W. (**2014.**) Measuring the value of research data: A citation analysis of oceanographic data sets. *PLoS ONE*, 9(3): e92590. doi:10.1371/journal.pone.0092590.

Costello, M.J. (**2009.**) Motivating online publication of data. *BioScience*, 59(5): 418-427. doi:10.1525/bio.2009.59.5.9.

Gibeaut, J. (**2016.**) Enabling data sharing through the Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC). *Oceanography*, 29(3): 33-37. doi:10.5670/oceanog.2016.59.

Goldstein, J.C., M.S. Mayernik, and H.K. Ramapriyan. (**2017.**) Identifiers for earth science data sets: Where we have been and where we need to go. *Data Science Journal*, 16: 23. doi:10.5334/dsj-2017-023.

Lawrence, B., C. Jones, B. Matthews, S. Pepler, and S. Callaghan. (**2011.**) Citation and peer review of data: Moving towards formal data publication. *International Journal of Digital Curation*, 6(2): 4-37. doi:10.2218/ijdc.v6i2.205.

McNutt, M., K. Lehnert, B. Hanson, B.A. Nosek, A.M. Ellison, and J.L. King. (**2016.**) Liberating field science samples and data. *Science*, 351(6277): 1024-1026. doi:10.1126/science.aad7048.

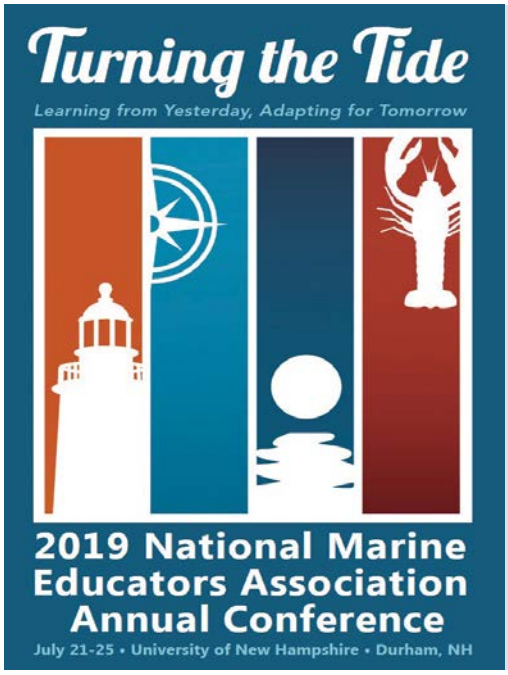
Piwowar, H.A., R.S. Day, and D.B. Fridsma. (**2007.**) Sharing detailed research data is associated with increased citation rate. *PLoS ONE*, 2(3): e308. doi:10.1371/journal.pone.0000308.

Tenopir, C., S. Allard, K. Douglass, A.U. Aydinoglu, L. Wu, E. Read, M. Manoff, and M. Frame. (**2011.**) Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 6(6): e21101. doi:10.1371/journal.pone.0021101.

Vines, T.H., R.L. Andrew, D.G. Bock, M.T. Franklin, K.J. Gilbert, N.C. Kane, J. Moore, B.T. Moyers, S. Renaut, D.J. Rennison, T. Veen, and S. Yeaman. (**2013.**) Mandated data archiving greatly improves access to research data. *The FASEB Journal*, 27(4): 1304-1308. doi:10.1096/fj.12-218164.

SANDRA ELLIS was Program Manager for the Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC) at the Harte Research Institute for Gulf of Mexico Studies at Texas A&M University in Corpus Christi, Texas.

KATIE FILLINGHAM See bio on page 20.



2019 National Marine Educators Association Conference: Save the Date!

JULY 21-25 • UNIVERSITY OF NEW HAMPSHIRE • DURHAM, NH

The National Marine Educators Association holds an annual conference that brings together informal marine educators, classroom teachers, university professors, and researchers. Participants come from all over the world to share knowledge and resources. Each year, a different local chapter has the opportunity to showcase their region. GOMMEA will be hosting the 2019 National Marine Educators Conference along the Seacoast at the University of New Hampshire. Visit the GOMMEA website or the NMEA website for registration and other conference deadlines.