Preparing Data Managers to Support Open Ocean Science: Required Competencies, Assessed Gaps, and the Role of Experiential Learning

Lee Wilson
ACENET
Atlantic Computational Excellence Network
Halifax, Canada
Email: lee.wilson@ace-net.ca

Adrienne Colborne, Michael Smit
School of Information Management
Dalhousie University
Halifax, Canada
Email: Adrienne.Colborne@dal.ca, Mike.Smit@dal.ca

***Abstract*—Ocean science is experiencing an explosion of data as researchers employ a widening variety of sensors, operating at higher fidelity and frequency, to inform our understanding of the global ocean. This is further complicated by the increasing integration of open science data from other disciplines to analyze complex systems, like climate change, animal migration, and sea/air interaction. This shift has been unplanned, chaotic, and emergent, and has placed the onus on researchers to stay current with best practices for managing, analyzing, and sharing data. Ocean scientists who do not have the technical skill to manage this data are turning to technologists, on the assumption they have the expertise required to help. To test this assumption, we examined an experiential learning program that placed technologists at ocean data centres in Canada, conducting interviews with students and employers to identify the competencies they believed were required to manage ocean data, which were missing in students' education up to that point, and which students gained during the work term placement.**

***Keywords*-big data education; ocean data management; data policy; experiential learning; open science**

I. Introduction

Ocean science, like many other disciplines, is transitioning from data scarcity [1] to a data deluge, with data now being captured at an unprecedented volume, variety, and velocity [2][3][4]. Physical sciences focusing on the marine environment, including marine biology, climatology, geology, and ecology, are increasingly dominated by satellites and remote sensing tools on a global scale[1] that dwarf traditional sampling methods [1][5]. In these fields, a perfect storm of data heterogeneity, a lack of standardized data management procedures, and rapid technological change has created a uniquely challenging environment for researchers [3][6][7][8]. In addition, institutions and granting agencies around the world are increasingly recognizing the importance of data management in scientific research, with requirements for data management planning quickly becoming the norm [6][9][10]. This simultaneous rise of Big Data coupled with a growing move to open research data in support of open science have created a need for technologists with research data management skills.

The adherence to, and application of, modern research data management practices and principles offer a means for scientists to contend with 21st century data challenges while also ensuring that data is discoverable and accessible for re-use, beyond the study for which it was originally collected [11]. However, data management and data sharing are plagued by barriers, both technical and cultural [12][13]. While the broad concept of research data management has begun to gain traction predominantly within the information sciences, there has been little work done on the more technical aspects of data management (e.g., database architecture, data ingestion and cleaning, and the application of technical metadata standards to ensure interoperability), and even less that focuses on data management in the oceans sector. While a few programs currently exist, fundamental pedagogical questions around ocean data management education remain relatively unexplored. In short, are graduates of technology programs suitably equipped to support managing and sharing big, open science data?

Drawing on interviews conducted with ocean data management practitioners and co-op students who work within several Canadian ocean data centers, this paper examines three key questions: 1) what skills are needed to prepare students for a career in ocean data management, 2) are students graduating from technology programs prepared with these skills, and 3) in what ways does experiential learning add value when cultivating these skills? While these interviews are specifically about ocean data in a Canadian context, we anticipate the insight from this unique study will be of interest to other disciplines in other countries.

We will begin by briefly introducing the Canadian ocean data context and the experiential learning program that is the subject of our study, then review the literature on challenges to managing and sharing data. We introduce our methods in Section III, and present the results in Section IV. We discuss the implications and analysis of these results in Section V, before concluding in Section VI.

[1]See the Global Earth Observation System of Systems (GEOSS; *https://www.earthobservations.org/geoss.php*) and the US Integrated Ocean Observing System (IOOS; *https://ioos.noaa.gov/*).

II. BACKGROUND & RELATED WORK

A. Ocean Data Management Community of Practice co-op program

The Community of Practice for Ocean Data Management (ODM CoP) connects Canadian ocean data centres to share and mobilize expertise and best practices, promote cooperation and alignment, and develop a shared vision for ocean data management in Canada. Conceived by its members at a 2014 Data Management Workshop, the community is composed of organizations from government, academic, and NGO sectors and is working to create an integrated ocean observation system for the open sharing of ocean observation data [15][16]. This Community of Practice is supported by the Marine Environmental Observation Prediction and Response (MEOPAR) Network, a Canadian Networks of Centres of Excellence (NCE) network that facilitates partnerships among academia, including natural, human health, and social scientists; government; and the private sector, including NGOs and not-for-profit organizations.

Initiated in the summer of 2015, the ODM CoP Summer Student Co-op Program seeks to support, inform, and further the ODM CoP through the creation and maintenance of data-related tools and infrastructure; the identification and establishment of best practices for ocean data management; and collaborative identification of potential challenges and barriers to effective data management. Six Canadian ocean data centres with varying regional focuses participated in the co-op. Students were recruited from Canadian universities, with a search focused on computer science and engineering co-op offices. Students completing their third work-term were preferred. The authors of this paper provide support and facilitation for the ODM CoP and the summer student program.

B. Literature on the Challenges of Open Data Management & Sharing

Strong data management and data sharing practices (i.e., "open" data) are increasingly being identified as essential ingredients for conducting good science [9][14]. Yet, despite many scientists agreeing that research data management is important, a prominent barrier is a cultural unwillingness to share research data, a key component in managing data and making it discoverable [4][7][13][14][17]. Research suggests that the reasoning behind scientists' unwillingness to share data are multifarious, ranging from concerns over sharing potentially sensitive data (i.e., data that contains personally identifiable information (PII), Traditional Knowledge, or data related to commercially valuable or endangered species) and the need for quality assurance of research data [1][14][4][18] to a sense of competition among scientists working in similar fields [7][14][19][20][21].

Another prominent barrier highlighted in the literature is a dearth of skills in research data management (RDM). RDM is often not a part of scientific curricula [22]. While there is increasing agreement that education in data management best practices and principles are important for researchers, it is unclear who should be responsible for the training and how it can be incorporated into existing programs of study. For example, some suggest that research librarians, particularly those trained in research data management, should teach these classes [23][24][25], others suggest that archivists will be the most qualified instructors [2][4], while another group argues that these classes should be part of the curriculum and taught by professors [10][25][26][27][28]. The latter are once again split by discipline–should these skills be taught to science students or expanded to students in all research areas, such as the social sciences? Poole [4] is in favour of the latter; he also argues that the sooner this training begins, the better. While some argue that training should occur during graduate studies, as this is the time that most students are performing research on a professional level [22][23][25][26][28], others suggest that an undergraduate foundation in these methods will be more beneficial [10][24][27]. There is a consensus; however, that this ought to be accomplished before students enter a professional environment. This is why some argue that such classes should be mandatory and included in the curriculum, rather than optional [25].

The requirements of what is taught are more generally agreed upon. Best practices must be taught (though care must be taken to make this instruction generalizable, as research practices vary widely across disciplines and programs), and more importantly must be explained in context of issues of data security and storage [4][10][19][24][29][30]). Real life examples are encouraged, especially with students who have little research experience[27][29]. There are several types of classes for RDM. Classes may be taught face-to-face in the classroom, as a trial class was taught at Syracuse University, or in a workshop or seminar [10][27]. Some examples of classes include the "New England Collaborative Data Management Curriculum," the "DataONE Data Management Education Modules" and "MANTRA" [28].

In addition to the general points made above, ocean data management presents a unique challenge. Ocean research deals with a high level of data heterogeneity, and as thinking shifts to a "whole-ocean" approach, the field has become increasingly multidisciplinary [1][6][12][21][31][32][33]). This problem is further compounded by a lack of standardized terminology and shared metadata standards, making it difficult to bring research together in any meaningful way [1][6][12][14][32]. Several ocean-specific training programs currently exist. The International Oceanographic Data and Information Exchange (IODE) has created the OceanTeacher Global Academy. This program "aims to develop a global training centre network and utilize this network to increase national capacity in coastal and marine knowledge and management" [34]. This program was implemented in 2014, and has several Research Training Centres across the globe and is the current standard for teaching oceans research data management [34].

III. Methods

The study population was composed of participants in the ODM CoP Student Co-op Program in 2015-2016, including both student employees and employers from the ODM CoP. This is the first academic year the program was active. This research was national in scope, as employers are based in various areas in Canada, and students were sought from 15 Canadian universities.

Having both student and employer perspectives allowed for a broader understanding of the effectiveness of the co-op program, as well as the current issues concerning the ocean data management field. Participation for employers was limited only to those who worked directly with summer co-op students (e.g., in a supervisory role). In total, 10 participants, five students and five employers, were drawn from the six Canadian ocean data centers that participated in the co-op. This distribution offers a good representation of our identified study population. The educational backgrounds of the students were varied, ranging from second year Computer Science, Mechanical Engineering, and Informatics to a post-doctoral student with a Ph.D. in Computer Science. The employers all had multiple years of experience in the field, and were technologists with varied educational backgrounds.

Participants were recruited via email, and were not provided compensation for participating in this study. The study was completed after the end of the student work terms, and the students were no longer employed by the CoP. As one of the authors had been involved in the recruitment and compensation of the employees, another author conducted all recruitment and interviews to avoid the potential for perceived conflict of interest or concern about dual roles. This study received ethics approval from the Research Ethics Board of Dalhousie University.

Participants were asked to participate in a semi-structured interview that consisted of questions related to: 1) participants' organizational (for employers) and educational (for students) backgrounds; 2) their experiences with the co-op program, including primary tasks or duties; and, 3) their opinion on what skills are most beneficial for working within an ocean data context. The semi-structured interview instruments used in this study were developed in consultation with Gajda and Jewiss' (2004) research on program evaluation [35]. The interviews were audio recorded, transcribed, and then coded using established qualitative data analysis techniques [36][37][38][39]. To protect the identities of participants, organization names have not been included and participants are identified by a number and whether they were a student or an employer.

IV. Results

Qualitative coding of interview sections revealed several prominent themes, including: the tasks students performed in the fulfilment of their summer co-op term; the education and skills needed to work in the ocean data industry, as well as perceived gaps in current educational offerings; and the value of the co-op experience to both students and employers. These will be presented in the following sections and then discussed in more detail in Section V.

To provide additional context, students and employers were asked to describe the roles and responsibilities of students in fulfilment of their summer co-op term. Table I details the specific projects undertaken by students.

A. Skills needed in ODM

One of the main questions driving this research centered on understanding the educational requirements for ODM, a topic that has not been adequately addressed in the existing literature. To that end, participants were asked a series of questions related to the types of skills needed within the ocean data management field, as well as the perceived gaps in current education programs. Table II summarizes the main themes identified by participants about the skills and experience needed in the ocean data management field.

B. Gaps in technology education

Table III shows the perceived educational gaps, organized according to respondent type (i.e., was the theme expressed by employers, students, or both). While there was some overlap between the two categories, it should be noted that the skills and experience described in Table II are considered important for the data management field, but are not concepts that participants felt must be covered in existing educational programs. It is understood that technology education programs aim to produce generalists who can readily learn additional skills in a particular application context. The most important components outlined in Table II will be expanded on in the Discussion section.

C. Benefits of experiential learning program

During interviews participants were also asked to discuss the overall value of the co-op experience. Students were asked to identify which skills, if any, they learned through their co-op placement and employers were asked to comment on the benefits accrued from hosting a student. Table IV summarizes learned skills that were specific to the ODM field, Table V summarizes general technical skills learned, and Table VI describes the stated overall value of the co-op from both student and employer perspectives.

V. Discussion

A. Skills needed in ODM

Examining the tasks co-op students undertook during placement provides insight into what is expected in this

TABLE I: Overview of tasks undertaken by trainees.

Built a Graphical User Interface (GUI) for a data visualizer, loaded data into the platform, and created several visualizations.
Conducted a survey within his/her regional university community to discover the types of data people had, how it was stored, and whether it contained sufficient metadata and contextual descriptors to be loaded into an ocean data database.
Worked on an initiative to create a directory for researchers working with gliders, concentrating UI/UX design and building relationships with stakeholders. Another student at the same data center built databases and wrote server-side code to load datasets into databases.
Participated in the creation of a dashboard for seismic data, doing interviews to determine user needs and creating mockups of the dashboard using Balsamic. They also engaged in public outreach activities for their organization.
Worked with datasets to make them compatible with ERDDAP, worked with a D3 library for data visualization, performed quality control procedures on datasets from researchers, and created a tag metadata parser. They also helped with an international acoustic telemetry workshop, worked with iPython (now Jupyter) notebooks, and designed a fully interactive map using Leaflet.
Worked with researcher data, performed data integration procedures, created web-based applications to improve data visualization, and set up data security for sensitive datasets.

rapidly expanding field. Our participants performed a variety of tasks over the course of the four-month co-op. Students helped to create and augment existing tools, software, and services offered by the data centres. This included web-based data visualization and distribution platforms, backend database infrastructure, scripts for data loading and quality control and assurance (e.g., automated metadata parsers), and protocols for data security. Even within a short work-term, students took on substantial projects.

> [The student] was hired to build a data visualization [platform]. We had an old data viewer and ... s/he completely changed the way it looked, s/he changed the user interface, s/he added ... about 50 more datasets. S/he significantly grew the amount of data that became available through it. (Participant #1 [employer])

Also included was important, but at times tedious, tasks, such as processing and loading complex, highly variable research data:

> I'd say, the biggest thing that I worked on would have been the preprocessing of data, and so by that I mean the [ocean data centre] would be given a lot of data from researchers and this data wasn't necessarily in a structured, standard format that we could then upload to a database. The forms [researchers] would hand in ... didn't follow the guidelines all the time, and so you'd need to go through what they actually submitted and make sure you could upload that ... to the database. (Participant #9 [student])

In addition to the technical elements of data management, students were also given an opportunity to attend industry workshops and conferences, participate in internal meetings, and engage with directly with researchers. As noted in the literature, interacting with researchers and navigating the political challenges associated with data sharing is a vitally important aspect of data management [4][7][13][14][17][40][24]. One student discussed the considerations that went into consultation with researchers:

> How do we communicate with people and make them understand that we're not trying to get anything from them? [This is] trying to be a collaborative project, so trying to deal with a little bit of the political side as well: making sure you're not asking the wrong person, asking the right person at the right time, making sure the right people are asked first, and [that] the stakeholders are considered throughout the process. (Participant #8 [student])

Employers were cognizant of the training component of the work term. Students were encouraged to engage in self-directed learning by creating projects that aligned with their personal and/or academic interests.

B. Gaps in technology education

Our interview data also provides insight into perceived gaps in current educational offerings within the represented programs of study. Both students and employers stated that increased training in managing geospatial data would have been beneficial during their co-op placement. As one employer suggested, it is difficult to find students who

TABLE II: Summary of skills and experience needed in ocean data management.

Theme	Description
Basic understanding of the science	Ocean data managers do not require expert knowledge of the data; however, a basic understanding of the science was deemed useful. More importantly, data managers should have some contact with the scientists responsible for the data collection and be able to communicate well with researchers in a variety of fields.
Data-specific training	Experience working with and refining large datasets programmatically; cleaning "messy" data for use in a database structure; knowledge of using a scripting language to load datasets into a database.
Database management training	An operational understanding of how to build, manage, manipulate, and maintain a database.
General principles and best practices for data management	Core concepts and best practices related to navigating the research data life cycle; understanding the basic principles of data management, including "why" data management is important.
Knowledge of industry-specific tools	Prior experience with industry-specific tools (e.g., ERDDAP) was described as useful, if not strictly necessary.
Standard operating procedures for data management	Knowledge of standard operating procedures for data management (e.g., the workflows detailing the processes from data collection, inception, processing and loading, to dissemination to the end-user).
User experience/user interface design training	Able to interact with and translate end user needs into a final product.

possess both coding proficiencies as well as knowledge of Geographic Information Systems (GIS):

> [GIS] was one of the things s/he had to dive in and study in order to understand how to deal with geospatial data because even though s/he was a computer engineer, s/he had never used maps or anything like that before. And coding is a totally different thing when trying to display geospatial data than it is when writing FORTRAN programs. The thing that I almost always find in new co-ops, when they come in, is that … if they know how to code, they don't know geospatial data, and if I had a GIS person, then their coding skills are not as great. So I'm not sure if I could hire a person who is actually a good coder and who knows programming, who knows CS and engineering, and then they could pretty much pick up on the geospatial stuff … but I almost never find the two go together. (Participant #1 [employer])

Employers and students also expressed the need for increased practical experience working with data, managing databases, and designing intuitive user-interfaces. Operational knowledge of databases, including database creation and maintenance, was felt to be lacking. Students described feeling unprepared to work with database applications in a practical setting and stated that they would have benefitted from projects based more on experiential, rather than theoretical, learning. Employers stated that it could take one to two months to teach students the fundamentals necessary to manage ocean data. The perceived lack of data training was expressed in similar terms: a general unpreparedness to work with large datasets in an operational setting. As one employer described:

> you'll catch someone trying to open a half a gigabyte file in Excel and wondering why that didn't work. So the idea about operational data exploration: the idea that you can get handed 500 megabytes of log files and you don't instantly start scanning them by eye, the idea that you could write things to do the hard parts of scanning things for you, or that you could analyze a file without opening it in a text editor. (Participant #4 [employer])

Participants also discussed the challenge of designing functional, intuitive user interfaces in the creation of tools to visualize ocean data. It was suggested that increased training in user interface design and experience working with end-users, including researchers, would have facilitated this

TABLE III: Perceived gaps in education prior to start of work term.

Participant type	Theme
Employer	• Operational knowledge of a scripting language (e.g., Python) • Standard operating procedures for data management
Student	• Exposure to basic scientific concepts • Experience working directly with researchers • "Soft skills" training on working with clients/end-users
Both	• Experience with geospatial data • Data-specific training • Database management training • User experience/user interface design training • General principles and best practices for data management

task. The formation of strong relationships with researchers based on mutual trust was highlighted by participants and the literature as being an essential component of the "soft" or non-technical side of data management. Ocean data managers must be able to work closely with scientists to better understand their needs, as well as the often complex scientific data.

This skills gap can be bridged, and participants had suggestions on how to cultivate the skills needed in this area. These included the introduction of more data intensive assignments in upper-year courses, working with local school boards to promote data literacy at an earlier age [41], teaching high-level scripting languages in addition to the standard programming languages (e.g., Python as well as JavaScript), and more advanced database courses. Many aspects of this theme relate to an overarching concept of application versus theory-based approaches to learning that permeated many of the interviews.

C. Benefits of experiential learning program

The skills learned during placement results (see Tables IV & V) suggest that the hands-on style learning afforded through a co-op environment may provide training in several areas that would be difficult to achieve through the more traditional classroom experience. In particular, student participants discussed the importance of a co-op in understanding the unique context of the ODM field, including the political challenges and data complexity inherent in the ocean research sector. As noted in the literature, the ocean research sector is fragmented and non-uniform, resulting in a variety of competing interests and technologies. Navigating the political side of sharing ocean data poses challenges that may not be covered in traditional Computer and Information Science programs. As one student noted, it is important to "have people with political skills in ocean data management because there's so many different partners … you always need to establish a good relationship and try to get their data. Because we're not producing data ourselves, we need to go get this data from all different members" (Participant #7 [student]). Another participant highlighted the importance of being able to work with diverse groups and promote collaboration, as they worked to "understand what everyone needs and wants and make sure that it's very clear what I'm asking of them" (Participant #8 [student]).

Students also described an added benefit from exposure to ocean-specific tools and technologies, for example, autonomous marine gliders, sensor buoy arrays, and domain-specific software such as ERDDAP[2]. This provided increased understanding of the field itself, as well as hands-on experience with the ocean data lifecycle. Students experienced first-hand how data moves from collection, clean-up, ingestion, storage, and, in many cases, discovery and reuse by other researchers in the oceans field (see Table VI).

VI. CONCLUSION

As a typically-sized qualitative study, the results described here are not generalizable. They do, however, offer insight that should be of interest to educators, scientists, and policy-makers who are interested in achieving the goals of open science with Big Data. In our interviews, students and employers were able to identify a set of core skills for ocean data management, and identified key areas where they did not perceive their education at met their needs. Through the work term, students discovered the value of the skills they had learned in class and the nature of the ocean data management field, including its unique tools and data lifecycle. By participating in experiential learning, the students gained experience with databases and datasets, an understanding of

[2]See https://coastwatch.pfeg.noaa.gov/erddap/index.html

TABLE IV: Ocean data management skills learned during the work term.

Skills	Theme(s)
Experience with Ocean Data Management Equipment	• Operational knowledge of a scripting language (e.g., Python) • Standard operating procedures for data management
Conducting data interviews	• Effective interviewing skills: Learning the best ways to conduct effective data interviews • Evaluating data needs: Using interviews to evaluate user's data needs
QA/QC scripting	• Importance of quality checking: Accuracy is crucial to ocean data. However, data is often transmitted in a raw, unprocessed format and additional tools (e.g., metadata sheets) are needed to prepare the data for ingest and reuse
Working with researchers, working in a marine science context	• Diversity of experience: Being in an environment with people of various backgrounds • Value of communication: Understanding the need for ocean researchers and ocean data managers to communicate their motivations and goals
Learning about the oceans field, including negotiating the interpersonal side of ocean data management	• Importance of relationships: With diverse stakeholders it is vital to ensure that everyone's needs are met. • Strengthening relationships means that people are more willing to participate in data efforts
Contending with the variety/complexity of data in ODM	• Ocean data is highly variable, which can increase the complexity of processing or working with the data • Oceans "field" involves multiple disciplines, often with competing standards and different data needs
Data Lifecycle	• Students were involved at many stages of the data lifecycle. This helped them to gain insight into the lifecycle and understand the processes that move data along the stages, as well as the importance of maintaining quality and metadata at each stage.

the political dimension of ocean data management, and more technical experience. Employers discussed the educational gaps they noted in the co-op students, and described the varied tasks and learning opportunities provided by the co-op. They also mentioned that the co-op was a valuable experience for the organization as a whole, and they found that it was a good chance to train ocean data managers.

The educational gaps identified by participants could be addressed through increased opportunities for experiential learning (e.g., co-op terms that embed students in real-world work environments), better integration of data management into technical curricula, and by providing workshops and other training opportunities outside the classroom. It is important to note that ocean data management has field-specific challenges and requisite skills, such as coping with highly variable, multi-faceted data and collaborating in an interdisciplinary field composed of a diverse stakeholder group, sometimes with competing interests. While the former could be taught through ocean data-specific course offerings, cultivating an understanding about the policy and political side of data management is difficult to achieve in a traditional classroom setting. Partnerships between academia and the ocean data community could be used to generate classroom assignments or term projects that reflect the actual needs and challenges of working in the ocean data field, in addition to exposing students to an interdisciplinary environment with many kinds of data, developing interpersonal skills, and learning about the data lifecycle in an ocean context.

TABLE V: General technical skills learned during the co-op term.

Skill	Theme(s)
Operational experience with databases/servers and manipulating data programmatically	• Processing data: Working through data as it comes in to get it ready for the database, considering sensitivity and compatibility
Project management skills	• Working on bridging the gap among various ocean stakeholders
Scripting languages (especially Python vs. Java learned in CS)	• Python experience: Learning how to use this language, becoming more familiar, and using Python software packages such as iPython (now Jupyter) notebooks • Using different platforms: Learning new software platforms and gaining increased experience with software taught in educational background
Data visualization and analysis	• Data visualization projects: Working on data visualization to develop web applications and using D3 to visualize it
Troubleshooting systems	• Learning how to troubleshoot: Learning the process for how to troubleshoot problems, and gaining confidence in trying new solutions • Cleaning data: Learning the process of preparing data for analysis

ACKNOWLEDGMENT

This study was funded by the MEOPAR NCE.

REFERENCES

[1] M. J. Costello and E. Vanden Berghe, "'Ocean biodiversity informatics': a new era in marine biology research and management," *Marine Ecology Progress Series*, 2006.

[2] D. Akmon, A. Zimmerman, M. Daniels, and M. Hedstrom, "The application of archival concepts to a data-intensive environment: working with scientists to understand data management and preservation needs," *Archival Science*, vol. 11, no. 3/4, pp. 329–348, Nov. 2011.

[3] S. Nativi, P. Mazzetti, M. Santoro, F. Papeschi, M. Craglia, and O. Ochiai, "Big Data challenges in building the Global Earth Observation System of Systems," *Environmental Modelling & Software*, vol. 68, pp. 1–26, Jun. 2015.

[4] A. H. Poole, "How has your science data grown? Digital curation and the human factor: a critical literature review," *Archival Science*, vol. 15, no. 2, pp. 101–139, 2015.

[5] Pouliquen, S., Hankin, S., Keeley, R., Blower, J., Donlon, C., Kozyr, A., & Guralnick, R. (2010), "The Development of the Data System and Growth in Data Sharing" (pp. 7–19), European Space Agency. https://doi.org/10.5270/OceanObs09.pp.30

[6] K. S. Baker, S. J. Jackson, and J. R. Wanetick, "Strategies Supporting Heterogeneous Data and Interdisciplinary Collaboration: Towards an Ocean Informatics Environment," in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 2005, p. 219b–219b.

[7] T. L. Kolb *et al.*, "How to Manage Data to Enhance Their Potential for Synthesis, Preservation, Sharing, and Reuse—A Great Lakes Case Study," *Fisheries*, vol. 38, no. 2, pp. 52–64, Feb. 2013.

[8] W. K. Michener and M. B. Jones, "Ecoinformatics: supporting ecology as a data-intensive science," *Trends in Ecology & Evolution*, vol. 27, no. 2, pp. 85–93, Feb. 2012.

[9] J. Vertesi and P. Dourish, "The value of data: considering the context of production in data economies," in *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, 2011, pp. 533–542.

[10] Y. Shorish, "Data Information Literacy and Undergraduates: A Critical Competency," *College and Undergraduate Libraries*, vol. 22, no. 1, pp. 97–106, 2015.

[11] Shearer, K. (2015), "Comprehensive Brief on Research Data Management Policies". Retrieved from: https://portagenetwork.ca/wp-content/uploads/2016/03/Comprehensive-Brief-on-Research-Data-Management-Policies-2015.pdf

[12] S. Hankin *et al.*, "Data Management for the Ocean Sciences - Perspectives for the Next Decade," 2010, pp. 570–579.

[13] J. E. Kratz and C. Strasser, "Researcher Perspectives on Publication and Peer Review of Data," *PLOS ONE*, vol. 10, no. 2, p. e0117619, Feb. 2015.

[14] J. Gallagher, J. Orcutt, P. Simpson, D. Wright, J. Pearlman, and L. Raymond, "Facilitating open exchange of data and information," *Earth Science Informatics*, vol. 8, no. 4, pp. 721–739, Dec. 2015.

[15] L. Wilson, M. Smit, and D. W. Wallace, "Results of a stakeholder engagement session on integrated ocean data management in Canada," *Data Science Journal*, under review.

[16] L. Wilson, M. Smit, and D. W. Wallace, "Results of an expert forum on ocean data management in Canada," *Canadian Meteorological and Oceanographic Society Bulletin*, vol. 44, no. 4, 2016.

[17] S. Mutula, "Managing Research Data in the Information Society," *African Journal of Library, Archives & Information Science*, pp. 119–121, Oct-2014.

[18] J. C. Wallis, C. L. Borgman, M. S. Mayernik, A. Pepe, N. Ramanathan, and M. Hansen, "Know Thy Sensor: Trust, Data Quality, and Data Integrity in Scientific Digital Libraries," in *Research and Advanced Technology for Digital Libraries*, L. Kovcs, N. Fuhr, and C. Meghini, Eds. Springer Berlin Heidelberg, 2007, pp. 380–391.

[19] C. Awre *et al.*, "Research Data Management as a 'wicked problem,'" *Library Review*, vol. 64, no. 4/5, pp. 356–371, 2015.

[20] L. Hsu, R. L. Martin, B. McElroy, K. Litwin-Miller, and W. Kim, "Data management, sharing, and reuse in experimental geomorphology: Challenges, strategies, and scientific opportunities," *Geomorphology*, vol. 244, pp. 180–189, Sep. 2015.

[21] S. J. Jackson and K. S. Baker, "Ecological Design, Collaborative Care, and Ocean Informatics," *PDC*, pp. 64–67, Jan. 2004.

[22] J. Carlson, M. S. Nelson, L. R. . Johnston, and A. Koshoffer, "Developing Data Literacy Programs: Working with Faculty, Graduate Students and Undergraduates," *Bulletin of the Association for Information Science & Technology*, vol. 41, no. 6, pp. 14–17, Sep. 2015.

[23] A. Simmons *et al.*, "Observation and integrated Earth-system science: A roadmap for 2016–2025," *Advances in Space Research*.

[24] C. A. Strasser and S. E. Hampton, "The fractured lab notebook: undergraduates and ecological data management training in the United States," *Ecosphere*, vol. 3, no. 12, p. UNSP 116, Dec. 2012.

[25] R. Darby *et al.*, "Enabling scientific data sharing and re-use," in *2012 IEEE 8th International Conference on E-Science*, 2012, pp. 1–8.

[26] D. H. . Charbonneau, "Strategies for Data Management Engagement," *Medical Reference Services Quarterly*, vol. 32, no. 3, pp. 365–374, Jul. 2013.

[27] L. Corti and V. Van den Eynden, "Learning to manage and share data: jump-starting the research methods curriculum," *Int. J. Soc. Res. Methodol.*, vol. 18, no. 5, pp. 545–559, Sep. 2015.

[28] C. Eaker, "Educating Researchers for Effective Data Management," *Bulletin of the Association for Information Science & Technology*, vol. 40, no. 3, pp. 45–46, Mar. 2014.

[29] J. Frugoli, A. M. Etgen, and M. Kuhar, "Developing and Communicating Responsible Data Management Policies to Trainees and Colleagues," *Science and Engineering Ethics*, vol. 16, no. 4, pp. 753–762, 2010.

[30] A. Via *et al.*, "Best practices in bioinformatics training for life scientists," *Briefings in Bioinformatics*, vol. 14, no. 5, pp. 528–537, Sep. 2013.

[31] K. S. Baker and C. L. Chandler, "Enabling long-term oceanographic research: Changing data practices, information management strategies and informatics," *Deep Sea Research (Part II, Topical Studies in Oceanography)*, vol. 55, no. 18–19, pp. 2132–2142, 2008.

[32] J. de La Beaujardire *et al.*, "Ocean and coastal data management," *OceanObs 2009*, 2009.

[33] F. Rack, P. Milne, P. Fippinger, and R. Jahnke, "Emerging Needs and Existing Links in Distributed Ocean Science Research Data Sets," in *Proceedings of the 2005 IEEE International Symposium on Mass Storage Systems and Technology*, Washington, DC, USA, 2005, pp. 9–13.

[34] "International Oceanographic Data and Information Exchange (IODE) - OceanTeacher Global Academy." Available: http://www.iode.org/index.php?option=com_content&view=article&id=431&Itemid=100177. Accessed: 27-Jan-2016.

[35] R. Gajda and J. Jewiss, "Thinking About How to Evaluate Your Program? These Strategies Will Get You Started," *Thinking*, vol. 9, no. 8, 2004.

[36] C. Anderson, "Presenting and Evaluating Qualitative Research," *Am J Pharm Educ*, vol. 74, no. 8, Oct. 2010.

[37] Bradley, E.H., Curry, L.A., & Devers, K. J. (2007). Qualitative data analysis for health services research: developing taxonomy, themes, and theory. Health Services Research, 42(2), 1758-72.

[38] G. W. Ryan and H. R. Bernard, "Techniques to Identify Themes," *Field Methods*, vol. 15, no. 1, pp. 85–109, Feb. 2003.

[39] D. Turner, "Qualitative Interview Design: A Practical Guide for Novice Investigators," *The Qualitative Report*, vol. 15, no. 3, pp. 754–760, May 2010.

[40] R. R. Christian, P. M. DiGiacomo, T. C. Malone, and L. Talaue-McManus, "Opportunities and challenges of establishing coastal observing systems," *Estuaries and Coasts: J ERF*, vol. 29, no. 5, pp. 871–875, Oct. 2006.

[41] C. Ridsdale, J. Rothwell, M. Smit, H. A. Hassan, M. Bliemel, D. Irvine, D. Kelly, S. Matwin, and B. Wuetherick, "Strategies and best practices for data literacy education: Knowledge synthesis report", tech. rep., Dalhousie University, 2015.

TABLE VI: Overall value of the co-op experience

EMPLOYERS	
Named Values	**Description**
Student projects facilitated the creation of new tools, software, services, etc.	• Students writing new tools: Writing new database tools or changing the user interface of existing tools to improve usability
Students benefit from hands-on work experience	• Students making an impact: Able to have some permanent code as part of ERDDAP • Students getting opportunity: Having the freedom to work on tools and learn how to use them
Co-op program a vehicle to connect Canadian ocean data centres	• The benefits of being connected: Helps ocean data relationships, helps on many levels (not just financial) for everyone to be working towards similar goals • The limits of this particular program: Wasn't a lot done overall to connect the different co-op students, needed to expand beyond project boundaries
Increased capacity within organization (due to increased human resources)	• Planning for co-op students work: Allocating time for particular projects during co-op students • Students taking over projects: Freeing up time for researchers, allowing new projects to be implemented • Ideas for improvements: Allowing for longer co-ops to let them have time to get over learning curve about ocean data and become more useful
Co-ops as a means to train ocean data managers	• Opportunity to train through co-op: Allowing students with Computer Science backgrounds to have experience with the ocean data lifecycle • "I find that it is much easier to teach computer scientists enough biology to get them to write a tool than it is to teach biologists enough CS to write a tool." (Participant #4 [employer])

STUDENTS	
Named Values	**Description**
Application of classroom learning to a real-work setting	• Value of electives: Co-op students benefitting from electives such as informatics and human-computer interaction • Computer language background: Having experience with Python and Javascript was a useful foundation for co-op projects
Benefits with field experts, embedded within an ocean data centre	• Learning about the field: Speaking to people in the field and learning about their work gave a better idea of the ocean field in general • Working in a multidisciplinary environment
Increased confidence to take on new and unfamiliar challenges (coping with uncertainty, time limits)	• Coping with the learning curve: The learning curve was steep, but the process of getting over the curve was a valuable challenge • Complex problems in a short timeframe: Confident about being handed difficult problems with a time limit and being able to solve them
Some freedom to choose work direction	• Reasoning behind diverse tasks: The ocean data field is limited by budget, which meant that multitasking was important • Co-op students choosing their work: Students were allowed freedom to choose work that fit with their interests