



Implementing Data Management Workflows in Research Groups Through Integrated Library Consultancy

ESSAY

JOSHUA BORYCZ 

**Author affiliations can be found in the back matter of this article*

]u[ubiquity press

ABSTRACT

Comprehensive research data management is fundamental to ensuring reproducible, open scientific research. However, sufficient research data assistance is often not offered at universities, and when offered, typically only provides basic services that are viewed as optional. Integrating information specialists into research groups provides a potentially promising means of improving data management by providing personalized *data management workflows*. Workflows are comprehensive, executable guides that require planning, implementation, feedback, and adaptation. Comprehensive data management workflows should include a file organization scheme, the creation of data management roles for members, a data storage/sharing guide, and training and evaluation. Librarians, who regularly interact with faculty and students and are familiar with data management tools, are uniquely situated to assist with the creation and assessment of these workflows.

CORRESPONDING AUTHOR:

Joshua Borycz

Vanderbilt University - US

joshua.d.borycz@vanderbilt.edu

KEYWORDS:

open data; consultant; incentives; science; global challenges; library; data management; coding; git; training

TO CITE THIS ARTICLE:

Borycz, J. 2021. Implementing Data Management Workflows in Research Groups Through Integrated Library Consultancy. *Data Science Journal*, 20: 9, pp. 1–9. DOI: <https://doi.org/10.5334/dsj-2021-009>

Organizing and sharing research data is a fundamental part of the research process (Mons, 2020) that has been shown to increase citation rates for publications (Dorch, Drachen, & Ellegaard, 2015; Henneken & Accomazzi, 2011; Piwowar, Day, & Fridsma, 2007), improve research reproducibility (Hothorn & Leisch, 2011; Popkin, 2019), and even save lives (Pisani et al., 2016). To improve the data sharing practices of researchers, the barriers to scientific data sharing must be reduced. Surveys of researchers have revealed that lack of knowledge and institutional resources often prevent them from sharing data (Curty, Crowston, Specht, Grant, & Dalton, 2017; J. Kim, Schuler, & Pechenina, 2018). One interesting series of surveys by Tenopir et al. demonstrated that scientists have positive views on data sharing in general, but do not wish to change their own practices (Tenopir et al., 2011; Tenopir, Christian, Allard, & Borycz, 2018; Tenopir et al., 2015). This is partly because their practices appear adequate and organizing and sharing data seems like a waste of time, but one of the most striking discoveries was that researchers did not know where to go for help when they wanted to find or share their data. They almost always sought help from a member of their group or a colleague, but rarely sought help from data and information specialists like librarians. Encouraging closer interactions between librarians and research groups could be one way to shift the scientific culture to a more open data environment.

Many initiatives that incorporate librarians into the research process are already in place at universities. Most of the top 100 universities in the world already have research support services involving their libraries (Si, Zeng, Guo, & Zhuang, 2019). However, in most cases only basic data management services are provided, such as Data Management Plan assistance and instruction on best practices (Cox, Kennan, Lyon, & Pinfield, 2017; Cox, Kennan, Lyon, Pinfield, & Scaffi, 2019; Kollen et al., 2017). A small proportion of universities offer more advanced instruction on data curation, processing, and analysis (Kollen et al., 2017), and most universities do not have standardized data management policies (Cox et al., 2019). Close interaction with librarians has been shown to improve research output and data sharing (Rethlefsen, Farrell, Osterhaus Trzasko, & Brigham, 2015). One promising method in research support services is the *consultative leadership model*, which encourages researchers to consult librarians at each stage of the research data lifecycle (Yu, Deuble, & Morgan, 2017). Lab-integration takes this model a step further by encouraging researchers to consider librarians as part of their lab groups (Carroll, Eskridge, & Chang, 2020). Scientific reliability and openness can thus be improved by encouraging a new type of interaction between existing departments in academia.

Describing the successful implementation of research data management plans within academic research is an important step for determining the unique challenges and lessons that other information professionals might use to change practices in their institutions. There are several examples of successful data management case studies at universities in the literature (Brocke & Lippe, 2015; Burnette, Williams, & Imker, 2016; Petters, Brooks, Smith, & Haas, 2019; Soehner, Steeves, & Ward, 2010). Generally case studies focus on large, university wide projects (Soehner et al., 2010). Smaller, more focused studies can demonstrate the importance of tailoring plans to the needs of the primary investigator (PI) and group members. For example, Petters et al. created and taught a data management curriculum targeted towards an ecological research group at Virginia Tech (Petters et al., 2019). Burnette et al. used interviews to evaluate the plan implementation process after providing some initial research consultations (Burnette et al., 2016), which provided insight into the management challenges that arise when introducing data management workflows into the research process.

Many data management case studies involve close interaction with the research group for only a portion of the research data life cycle. Surveys indicate that most libraries do offer personal, in-depth interactions in the form of research consultations (Si et al., 2019), but most of these interactions occur in the form of workshops, general data management instruction, or tool/code training (Murray et al., 2019). Research data services are typically available to help with the early stages of a project in the form of data management plans or data management training, or near the end of project when data preservation and dissemination become issues (Murray et al., 2019; Petters et al., 2019). More holistic approaches that incorporate information specialists in the planning, implementation, uptake, and preservation of data are rare (Curdt, 2019; Curdt & Hoffmeister, 2015; Jones, 2013).

Resource allocation is an important issue for research data services. Research grants and universities often do not allocate funds for data management support and there are typically only a few data or liaison librarians available to cover the needs of hundreds or thousands of researchers (Mons, 2020). The Delft University of Technology addresses this by providing one Data Steward for each academic department that is available to answer questions for faculty (TUDelft, 2020). Virginia Tech provides Data and Informatics Consultants that collaborate with other university units to provide consistent and widespread services (Ogier, Brown, Petters, Hilal, & Porter, 2018). One way to amplify the impact of research consultations is to record and track progress and methodology in an open repository (e.g., Open Science Framework) for other faculty members and librarians to access (COS, 2020). Furthermore, if libraries focus on providing generalizable, non-technical data support (e.g., file naming conventions, data role assignments) then more librarians could be made available to provide data services as little technical expertise would be needed.

This essay, informed by interactions with science and engineering faculty, describes the philosophy behind the design of research data services offered by the Vanderbilt Science and Engineering Library. Librarians will be integrated into research groups to engage with the PI and students regularly and learn their daily research routines. Working closely with the PI, the librarian will then create comprehensive *data management workflows* that include file organization schemes, the creation of data management roles for members, data storage/sharing guides, and training. The procedures and tools that are developed for each group will be stored on the Open Science Framework to provide guidance for other researchers and librarians. Assessment of the workflows will also be carried out by the librarian to make adaptations and improvements to the process.

2 BECOMING PART OF THE GROUP

2.1 INITIAL INTERACTION

Trust between librarians and faculty must be established before faculty will consider integrating librarians into the research process. In previous work, interactions like this have been initiated by advertising consultancy services to faculty and students using online applications and appointments (Carroll et al., 2020; Si et al., 2019). However, it is generally a good idea to take an active approach by contacting faculty that have already shown an interest in improving their data management practices or those that have a good professional relationship with the specific librarians that will be working with their groups. In some cases librarians may feel reluctant to participate in the implementation of a data management workflow out of fear that faculty members will view this as an imposition or that they do not have the subject specific knowledge necessary to assist with data organization. However, faculty largely have positive views of data sharing and are amenable to working with librarians as long as they maintain control of their data (Lage, Losoff, & Maness, 2011). Furthermore, most of the knowledge necessary to implement data management workflows is group specific rather than field specific, which means that field specific expertise is often not necessary (Murray et al., 2019).

Most faculty members do not have substantive data organization rules for their research groups (Akers & Doty, 2013; Tenopir et al., 2020). Introducing basic practices that are tuned to the needs of the specific research group, and working with the group long enough to address issues that arise should have a more lasting impact as it substantially reduces the cost, time, and knowledge barriers that prevent most research data from being organized properly. Regular meetings with the PI, attendance of group meetings, and availability for questions from students are all important in the initial phase of the workflow design.

3 DATA MANAGEMENT WORKFLOWS

The primary difference between a data management plan and a workflow is actionability. A workflow closes the gap between data management theory and actual practice by providing group specific, concrete steps that researchers can take to improve their data management practices. Teaching data management principles and best practices is not enough to change the behavior of individual researchers. Research is notorious for addressing very specific, complex, and difficult problems. Data management workflows should follow suit. They should be specific enough that each group member knows exactly what they need to do with their data at every step. Data should be fully incorporated into the daily tasks of each group member.

The librarian can provide general knowledge about data organization and tools, but to create a workflow the PI and students need to provide insight into the daily routine of the group. The most important aspect of this interaction is understanding how to balance the introduction of new procedures, tools, and compliance requirements with the knowledge level of group members as well as the level of work needed to implement the proposed ideas. If the new procedures require learning how to use multiple complex tools, substantially change existing group workflows, require hours of extra work to fulfill, and/or have no accountability mechanism, then these changes will prove unsustainable.

The goal of a data management workflow is to make research data understandable and procedures reproducible for people unfamiliar with the work. Accomplishing this goal will require group members to have clear and simple guides and tasks, accountability for their work, and a way to share progress and feedback. Hence, the four fundamental parts of an efficient data management workflow are:

1. **File organization and naming scheme** to ensure that folders and files are consistent enough to easily navigate and understand their contents.
2. **Code/data cleaning and upload procedures** to allow other group members to understand and check the work of their colleagues.
3. **Group data management roles** to ensure file security and compliance with the new procedures.
4. **Research group wiki** for group members to share important lab documents like lab notebooks, project updates, and published research data.

3.1 FILE ORGANIZATION AND NAMING SCHEME

Organizing and naming files and folders properly is fundamental to reproducible scientific research. Yet this topic is not always sufficiently addressed by research data services offered at universities, which tend to focus on providing technological assistance (Kollen et al., 2017; Murray et al., 2019). Organizations like the Data Observation Network for Earth (DataONE) offer guidelines and lessons on file organization that are very useful (Strasser, Cook, Michener, & Budden, 2011), but general guidelines are often ambiguous and those from different organizations and fields may conflict. Consistency is key when designing a data management workflow, so file organization principles should be flexible enough that they can be adapted to a specific group upon discussion with the PI and students. The guidelines should then be clearly described and checked for consistency regularly. Five general guidelines for designing a file organization scheme that should work for most research groups are provided here:

1. **A central location for file storage** should be used by all group members. The security needs for this storage location should be discussed with the PI. This can be a local server or an online repository.
2. **A project centered folder hierarchy** should be used to allow group members to find files that are connected by a specific project goal in one location. Each project folder should contain a consistent set of folders that hold files related to specific aspects of the research life cycle (e.g., Literature, Analysis, Funding, Publications, Training) (Brocke & Lippe, 2015; Noble, 2009).
3. **Only clean code/data/manuscripts** should be stored in the central project folders for the research group. These folders should only contain work that you would be happy for someone else to critique.
4. **Tag based file names** that give information about the contents and the location of each file should be used. Tags should be recorded and described in a README file and used consistently (Borer, Seabloom, Jones, & Schildhauer, 2009; MIT, 2018). Tags should be separated using delimiters that are consistent between projects. Two common variable naming conventions are *snake case*, which uses underscores as delimiters and *camel case*, which uses capital letters to separate words (Divine, 2018).
5. **README files** containing the author, path, name, edit dates, and description of the important files should be created for every project. There are many templates available that can help with README design (MIT, 2018).

The exact details of the hierarchy, files names, README file contents, and code cleaning rules should be discussed with the PI and students. Once these details are decided, the rules should be recorded and stored in a central location that all group members can access. These guidelines should then be followed for all files placed in the project folders. Making adaptations to the file organization will likely be necessary. Upon making these changes all of the files describing the file organization scheme should be updated immediately, followed by all of the existing files that do not comply with these changes.

3.2 CODE/DATA CLEANING AND UPLOAD PROCEDURES

Guidelines for cleaning code should also be provided for all group members. Teaching and enforcing basic coding standards has clear benefits to the quality of research output (Li & Prasad, 2005; Popic, Velikic, Jaroslav, Spasic, & Vulic, 2018; Wilson et al., 2014). Coding procedures can be fairly general but should be described clearly and used consistently by all group members. It is a good idea to use existing code and data cleaning best practices procedures as a starting point (NOAA, 2007; Ruder & Ge, 2014). The procedures do not necessarily have to be code specific, but can be if the PI and/or students think it is necessary. Some guidelines for code and data cleaning are listed here:

1. **Include a detailed header** in all code files that lists the author, description, date of last edit, and dependencies.
2. **Code should be divided into modules** that accomplish specific purposes to make the code easy to follow.
3. **Comments should be included** that describe each module and each variable.
4. **Descriptive variable names** that describe the variable contents should be used. Logical variables, constants, functions, and iterative variables can be distinguished with specific naming conventions.
5. **Scripts, data, and output** files should each have their own folders and be named in accordance with the file naming conventions of the lab.
6. **README files** that list the code files, purpose, author, edit dates, and changes to the code should be included with the code.
7. **One table per data file** makes all data visible in the file list.
8. **Versions of code/data** should be tracked using a tool like git or within README files and version tags in files names.

Some of these steps may not be necessary for every research group. Research groups that write very complex, language specific codes may require a much more detailed set of coding standards, but most research groups should benefit from consistently enforced, basic coding practices (Popic et al., 2018; Wilson et al., 2016). Once the cleaning and upload procedures are decided. A detailed description of the minimum requirements for code to be stored on the group's server should be provided to the group by the librarian. Any code existing in this storage location should be updated based on these standards. Librarians should also provide training in the use of any tools that will be used for code/data versioning or cleaning, such as GitHub or OpenRefine (GitHub, 2020; OpenRefine, 2020). The Software Carpentries website offers an excellent introduction to these tools that might be useful (Software Carpentry, 2020). This training should include a precise series of steps on how these new tools will be incorporated into the group workflow. It is vital for librarians to remain available after the initial training to answer questions for students unfamiliar with these tools.

3.3 DATA MANAGEMENT ROLES

The primary step that makes a data management workflow actionable is assigning data management roles to group members. Descriptions of the daily and weekly tasks for each role should be stored in a central location so each member understands their duties. These roles should be designed so as not to place an undue burden on any single group member. Multiple checks should be put in place to make sure that all group members are performing the tasks assigned to them. A few key roles that should be considered are provided here:

1. **Data role manager** will track the role assignments in the group, onboard new group members, offboard members that will soon leave the group, and register issues of non-compliance with data management guidelines.

2. **Project manager** will be responsible for tracking the progress of a project and ensuring compliance with data management and coding procedures. The project manager is responsible for addressing feedback from the file reviewer.
3. **Project member** will be responsible for writing and cleaning code, updating READMEs, and providing updates to the project manager on his or her portion of a project.
4. **File reviewer** will check the code, data, naming conventions, and file hierarchy for a project on which he or she is not directly involved and provide feedback to the project manager.

The librarian and PI should be responsible for deciding on and writing the specific tasks for each of the data management roles in the group. It is best to have the daily and weekly tasks for each role clearly described and have procedures in place for times when roles need to be changed. Meetings between project managers, project members, and file reviewers should occur regularly. Procedures for offboarding group members should be in place to ensure that the exiting member's files are clean and organized before leaving. Onboarding training procedures should be clearly outlined as well. If the file organization and cleaning guidelines are followed strictly then the onboarding process for research projects should be quite simple.

3.4 FACILITATION AND TRAINING

To facilitate the use of the data management workflows designed for the group, a group wiki should be used for the storage of important lab documents like the data role assignment list, lab notebooks, project updates, data management procedures, pre-prints, and published data. This can be done locally or in an online repository like the Open Science Framework, depending on the security needs of the group. The librarian will be responsible for the creation of documents outlining the file organization rules, code/data cleaning procedures, onboarding and offboarding procedures, and tasks for the data management roles. The tools created by the librarian will make it simple for each of the data management roles to perform their tasks. These tools could include a data role tracking spreadsheet, README templates, code checklists for file reviewers, versioning procedures, and data publication guidelines. Each of these tools should be presented to the PI for approval. All of these tools should be organized, described, and stored in an open repository to allow other faculty and librarians to reproduce and adapt these procedures for their needs. The Open Science Framework is an excellent option for sharing these procedures as many research groups have already used it to manage their files and data (Roesch et al., 2016; Wycle et al., 2017).

Training and monitoring the success of the data management workflow represents a substantive time commitment from the librarian. However, most of the time commitment occurs in the early stages of the project when the workflow documents are created. Once the workflow guidelines and tools are in place, the librarian will need to meet with the group members to provide training and assistance. The training should include descriptions of the file organization, data/code cleaning procedures, data management roles, and how to navigate the documentation describing these aspects of the new workflow. To do so the librarian should attend group meetings and meet with students to answer questions. After the training the librarian will need to be available for questions and to address issues with the workflow, but should not have to commit a great deal of time to assisting the group. After six months to a year an assessment of the new data management workflow should be performed to evaluate the successes and failures of the program.

4 CONCLUSIONS

The importance of personal interaction in the implementation of data management workflows is extremely important. Creating workflows with library support should reduce barriers preventing PIs from taking full advantage of their data. Once in place these workflows should help faculty save a great deal of time, publish more papers, and increase the reproducibility and citation rates of their publications (Piwowar et al., 2007). Effective data management workflows could give early adopters a competitive edge that would motivate other faculty to follow suit, which could change the practices of entire departments and universities. This requires a large time commitment from university libraries, but this is one of the primary reasons why the data consultation model between librarians and research faculty is potentially so powerful. Libraries are permanent fixtures

at all universities, containing professionals whose primary job is to make it easier for faculty and students to find and use information. Librarians do not cost extra money, evaluate students, or compete with faculty for publications or resources. This means that librarians could easily be viewed as research group members whose primary focus is making research more efficient.

From the perspective of an academic library, another advantage of this consultation model is that librarians can gain in-depth understandings of and build meaningful relationships with research groups over longer periods of time, while also fostering researchers' competencies in data management practices. This is in direct contrast to the data management workshop model, which despite its popularity among academic libraries, frequently encounters issues related to both effectiveness and sustainability (Reed, 2013). Practices like these have been discussed widely by members of professional data and information science communities (Borycz & Carroll, 2018, 2020; Tenopir, Sandusky, Allard, & Birch, 2014), but funding often goes to the creation of new organizations rather than the training and development of people in long-standing professions at research institutions. In my view, refocusing an existing system is far faster, cheaper, and easier than creating a new one to address an old problem.

It may not seem like changing the nature of a few professional relationships could change the scientific community as a whole, but human behavior can be altered drastically just by making an existing task slightly easier. It is clear that open and organized data will be necessary to resolve many of the existing issues that plague the sciences. All that stands in the way are misaligned personal incentives (Curty et al., 2017; Y. Kim & Yoon, 2017; Tenopir et al., 2011, 2018, 2015). Changing these incentives will require a great deal of trust in the data management and sharing practices of colleagues. Librarians could help to build this trust as they are established members of the academic community with no competing research interests and expertise in information management. Providing quality, consistent, actionable data management support at universities is well within the scope of the mission of libraries, and once the value of this service is demonstrated researchers will be eager to participate.

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR AFFILIATION

Joshua Borycz  orcid.org/0000-0002-1505-148X
Vanderbilt University, US

REFERENCES

- Akers, KG and Doty, J. 2013. Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, 8(2): 5–26. DOI: <https://doi.org/10.2218/ijdc.v8i2.263>
- Borer, ET, Seabloom, EW, Jones, MB and Schildhauer, M. 2009. Some Simple Guidelines for Effective Data Management. *Bulletin of the Ecological Society of America*, 90(2): 205–214. DOI: <https://doi.org/10.1890/0012-9623-90.2.205>
- Borycz, J and Carroll, B. 2018. Managing Digital Research Objects in an Expanding Science Ecosystem: 2017 Conference Summary Digital Objects – The Core of Our Complex Data Market. *Data Science Journal*. DOI: <https://doi.org/10.5334/dsj-2018-016>
- Borycz, J and Carroll, B. 2020. Implementing FAIR data for people and machines: Impacts and implications – results of a research data community workshop. *Information Services & Use*, 1(Preprint): 1–15. DOI: <https://doi.org/10.3233/ISU-200083>
- Brocke, J and Lippe, S. 2015. Managing collaborative research projects: A synthesis of project management literature and directives for future research. *International Journal of Project Management*, 33(5): 1022–1039. DOI: <https://doi.org/10.1016/j.ijproman.2015.02.001>
- Burnette, M, Williams, S and Imker, H. 2016. From Plan to Action: Successful Data Management Plan Implementation in a Multidisciplinary Project. *Journal of EScience Librarianship*, 5(1): e1101. DOI: <https://doi.org/10.7191/jeslib.2016.1101>
- Carroll, AJ, Eskridge, HN and Chang, BP. 2020. Lab-integrated librarians: A model for research engagement. *College and Research Libraries*, 81(1): 8–26. DOI: <https://doi.org/10.5860/crl.81.1.8>
- COS. 2020. The Open Science Framework. Retrieved January 11, 2021, from <https://www.cos.io/products/osf>.

- Cox, AM, Kennan, MA, Lyon, L and Pinfield, S.** 2017. Developments in research data management in academic libraries: Towards an understanding of research data service maturity. *Journal of the Association for Information Science and Technology*, 68(9): 2182–2200. DOI: <https://doi.org/10.1002/asi.23781>
- Cox, AM, Kennan, MA, Lyon, L, Pinfield, S and Sbaffi, L.** 2019. Maturing research data services and the transformation of academic libraries. *Journal of Documentation*, 75(6): 1432–1462. DOI: <https://doi.org/10.1108/JD-12-2018-0211>
- Curdt, C.** 2019. Supporting the Interdisciplinary, Long-Term Research Project ‘Patterns in Soil-Vegetation-Atmosphere-Systems’ by Data Management Services. *Data Science Journal*, 18(1): 1–9. DOI: <https://doi.org/10.5334/dsj-2019-005>
- Curdt, C and Hoffmeister, D.** 2015. Research data management services for a multidisciplinary, collaborative research project: Design and implementation of the TR32DB project database. *Program*, 49(4): 494–512. DOI: <https://doi.org/10.1108/PROG-02-2015-0016>
- Curty, RG, Crowston, K, Specht, A, Grant, BW and Dalton, ED.** 2017. Attitudes and norms affecting scientists’ data reuse. *PLoS ONE*, 12(12): 1–22. DOI: <https://doi.org/10.1371/journal.pone.0189288>
- Divine, P.** 2018. Case Styles: Camel, Pascal, Snake, and Kebab Case. Retrieved December 15, 2020, from BetterProgramming website: <https://medium.com/better-programming/string-case-styles-camel-pascal-snake-and-kebab-case-981407998841>.
- Dorch, SBF, Drachen, TM and Ellegaard, O.** 2015. The data sharing advantage in astrophysics. *Proceedings of the International Astronomical Union*, 11(A29A): 172–175. Retrieved from <http://arxiv.org/abs/1511.02512>. DOI: <https://doi.org/10.1017/S1743921316002696>
- GitHub.** 2020. GitHub Desktop. Retrieved July 29, 2020, from <https://desktop.github.com/>.
- Henneken, EA and Accomazzi, A.** 2011. *Linking to Data – Effect on Citation Rates in Astronomy*. Retrieved from <http://arxiv.org/abs/1111.3618>.
- Hothorn, T and Leisch, F.** 2011. Case studies in reproducibility. *Briefings in Bioinformatics*, 12(3): 288–300. DOI: <https://doi.org/10.1093/bib/bbq084>
- Jones, S.** 2013. The range and components of RDM infrastructure and services. In: Whyte, A, Pryor, G and Jones, S (eds.), *Delivering Research Data Management Services: Fundamentals of Good Practice*. pp. 89–114. DOI: <https://doi.org/10.29085/9781783300242.006>
- Kim, J, Schuler, ER and Pechenina, A.** 2018. Predictors of Data Sharing and Reuse Behavior in Academic Communities. In: Alemneh, DG, Allen, J and Hawamdeh, S (eds.), *Knowledge Discovery and Data Design Innovation*, 14: 1–25. DOI: https://doi.org/10.1142/9789813234482_0001
- Kim, Y and Yoon, A.** 2017. Scientists’ Data Reuse Behaviors: A Multilevel Analysis. *Journal of the Association for Information Science and Technology*, 68(12): 2709–2719. DOI: <https://doi.org/10.1002/asi.23892>
- Kollen, C, Kouper, I, Ishida, M, Williams, S, Fear, K, Kouper, I, Williams, SC, et al.** 2017. Research Data Services Maturity in Academic Libraries. In: *Curating Your Reserch Data: Practical Strategies for Your Digital Repository*, 1st ed. pp. 153–170. Retrieved from <http://hdl.handle.net/10150/622168>.
- Lage, K, Losoff, B and Maness, J.** 2011. Receptivity to library involvement in scientific data curation: A case study at the University of Colorado Boulder. *Portal*, 11(4): 915–937. DOI: <https://doi.org/10.1353/pla.2011.0049>
- Li, X, and Prasad, C.** 2005. Effectively teaching coding standards in programming. *Proceedings of the 6th Conference on Information Technology Education, SIGITE 2005*, 239–244. DOI: <https://doi.org/10.1145/1095714.1095770>
- MIT.** 2018. Organize your files. Retrieved July 29, 2020, from <https://libraries.mit.edu/data-management/store/organize/>.
- Mons, B.** 2020, February 1. Invest 5% of research funds in ensuring data are reusable. *Nature*, 578: 491. DOI: <https://doi.org/10.1038/d41586-020-00505-7>
- Murray, M, O’Donnell, M, Laufersweiler, M, Novak, J, Rozum, B and Thompson, S.** 2019. A survey of the state of research data services in 35 U.S. academic libraries, or “Wow, what a sweeping question.” *Research Ideas and Outcomes*, 5: 48809. DOI: <https://doi.org/10.3897/rio.5.e48809.figure1>
- NOAA.** 2007. *NOAA National Weather Service NWS/OHD General Software Coding Standards and Guidelines General Software Development Standards and Guidelines Version 3.5*.
- Noble, WS.** 2009. A Quick Guide to Organizing Computational Biology Projects. *PLoS Computational Biology*, 5(7): e1000424. DOI: <https://doi.org/10.1371/journal.pcbi.1000424>
- Ogier, A, Brown, A, Petters, J, Hilal, A and Porter, N.** 2018. Enhancing collaboration across the research ecosystem: Using libraries as hubs for discipline-specific data experts. *ACM International Conference Proceeding Series*, 1–6. DOI: <https://doi.org/10.1145/3219104.3219126>
- OpenRefine.** 2020. OpenRefine. Retrieved January 11, 2021, from <https://openrefine.org/>.
- Petters, JL, Brooks, GC, Smith, JA and Haas, CA.** 2019. The Impact of Targeted Data Management Training for Field Research Projects – A Case Study. *Data Science Journal*, 18(1): 1–7. DOI: <https://doi.org/10.5334/dsj-2019-043>

- Pisani, E, Aaby, P, Breugelmans, JG, Carr, D, Groves, T, Helinski, M, Guerin, PJ, et al. 2016. Beyond open data: realising the health benefits of sharing data. *BMJ (Clinical Research Ed.)*, 355: i5295. DOI: <https://doi.org/10.1136/bmj.i5295>
- Piwowar, HA, Day, RS and Fridsma, DB. 2007. Sharing detailed research data is associated with increased citation rate. *PLoS One*, 2(3): e308–e308. DOI: <https://doi.org/10.1371/journal.pone.0000308>
- Popic, S, Velikic, G, Jaroslav, H, Spasic, Z and Vulic, M. 2018. The Benefits of the Coding Standards Enforcement and its Influence on the Developers' Coding Behaviour: A Case Study on Two Small Projects. *2018 26th Telecommunications Forum, TELFOR 2018 – Proceedings*. DOI: <https://doi.org/10.1109/TELFOR.2018.8612149>
- Popkin, G. 2019, May 1. Data sharing and how it can benefit your scientific career. *Nature*, 569: 445–447. DOI: <https://doi.org/10.1038/d41586-019-01506-x>
- Reed, M. 2013. Workshops Don't Work. Retrieved July 30, 2020, from Inside Higher Ed website: <https://www.insidehighered.com/blogs/confessions-community-college-dean/workshops-don't-work>
- Rethlefsen, ML, Farrell, AM, Osterhaus Trzasko, LC and Brigham, TJ. 2015. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. *Journal of Clinical Epidemiology*, 68(6): 617–626. DOI: <https://doi.org/10.1016/j.jclinepi.2014.11.025>
- Roesch, EB, Scrivener, CL, Ras, I, Malik, A, Lindner, M, Dubuc, TT, Figueiredo, N, et al. 2016. *EBRLab*. DOI: <https://doi.org/10.17605/OSF.IO/EDE88>
- Ruder, D and Ge, J. 2014. *All-In-One Code Framework Coding Standards*.
- Si, L, Zeng, Y, Guo, S and Zhuang, X. 2019. Investigation and analysis of research support services in academic libraries. *Electronic Library*, 37(2): 281–301. DOI: <https://doi.org/10.1108/EL-06-2018-0125>
- Soehner, C, Steeves, C and Ward, J. 2010. E-Science and Data Support Services: A Study of ARL Member Institutions. Retrieved from Association of Research Libraries; Web site: www.arl.org.
- Software Carpentry. 2020. Our Lessons. Retrieved January 11, 2021, from <https://software-carpentry.org/lessons/>.
- Strasser, C, Cook, R, Michener, W and Budden, A. 2011. *Primer on Data Management: What you always wanted to know*. Retrieved from www.dataone.org.
- Tenopir, C, Allard, S, Douglass, K, Aydinoglu, AU, Wu, L, Read, E, Frame, M, et al. 2011. Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE*, 6(6): e21101. DOI: <https://doi.org/10.1371/journal.pone.0021101>
- Tenopir, C, Christian, L, Allard, S and Borycz, J. 2018. Research Data Sharing: Practices and Attitudes of Geophysicists. *Earth and Space Science*, 5(12): 891–902. DOI: <https://doi.org/10.1029/2018EA000461>
- Tenopir, C, Dalton, ED, Allard, S, Frame, M, Pjesivac, I, Birch, B, Dorsett, K, et al. 2015. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLOS ONE*, 10(8): e0134826. DOI: <https://doi.org/10.1371/journal.pone.0134826>
- Tenopir, C, Rice, NM, Allard, S, Baird, L, Borycz, J, Christian, L, Sandusky, RJ, et al. 2020. Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLOS ONE*, 15(3): e0229003. DOI: <https://doi.org/10.1371/journal.pone.0229003>
- Tenopir, C, Sandusky, RJ, Allard, S and Birch, B. 2014. Research data management services in academic research libraries and perceptions of librarians. *Library and Information Science Research*, 36(2): 84–90. DOI: <https://doi.org/10.1016/j.lisr.2013.11.003>
- TU Delft. 2020. Data Stewardship. Retrieved January 11, 2021, from <https://www.tudelft.nl/en/library/current-topics/research-data-management/r/support/data-stewardship/>.
- Wilson, G, Aruliah, DA, Brown, CT, Chue Hong, NP, Davis, M, Guy, RT, Wilson, P, et al. 2014. Best Practices for Scientific Computing. *PLoS Biology*, 12(1): e1001745. DOI: <https://doi.org/10.1371/journal.pbio.1001745>
- Wilson, G, Bryan, J, Cranston, K, Kitzes, J, Nederbragt, L and Teal, TK. 2016. *Good Enough Practices in Scientific Computing*. Retrieved from <http://software-carpentry.org/>. DOI: <https://doi.org/10.1371/journal.pcbi.1005510>
- Wyble, B, Glenn, A, Hess, M, Maceyko, A, Callahan-Flintoft, C, Swan, G, O'Donnell, RE, et al. 2017. *Wyble Lab*. Retrieved from <https://osf.io/ur3a4/>.
- Yu, F, Deuble, R and Morgan, H. 2017. Designing research data management services based on the research lifecycle – a consultative leadership approach. *Journal of the Australian Library and Information Association*, 66(3): 287–298. DOI: <https://doi.org/10.1080/24750158.2017.1364835>

TO CITE THIS ARTICLE:

Borycz, J. 2021. Implementing Data Management Workflows in Research Groups Through Integrated Library Consultancy. *Data Science Journal*, 20: 9, pp. 1–9. DOI: <https://doi.org/10.5334/dsj-2021-009>

Submitted: 19 October 2020

Accepted: 14 January 2021

Published: 17 February 2021

COPYRIGHT:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.