

PRACTICE PAPER

From FAIR Leading Practices to FAIR Implementation and Back: An Inclusive Approach to FAIR at Leiden University Libraries

Kristina Maria Hettne¹, Peter Verhaar¹, Erik Schultes² and Laurents Sesink¹

¹ Centre for Digital Scholarship, Leiden University Libraries, Leiden, NL

² GO FAIR International Support and Coordination Office (GFISCO), Leiden, NL

Corresponding author: Kristina Maria Hettne (k.m.hettne@library.leidenuniv.nl)

Leiden University (LU) adopted a data management regulation in 2016. The regulation embraces the Findable, Accessible, Interoperable and Reusable (FAIR) principles. To implement the regulation a programme was established. The focus of the programme was initially to raise awareness and to set up services to make data Findable and Accessible and to train researchers on data management planning. In 2019, the programme entered its second phase, with an increased focus on Interoperable and Reusable data, and on implementing the machine-actionable aspects of FAIR data. This step is non-trivial, however, because of the fast-developing FAIR research data international research field that requires fast adoption of leading practices by support professionals with the adequate skills. In this paper we describe how LU aims to close the feedback loop between international bottom-up organisations such as GOFAIR, the Research Data Alliance and CODATA on the one hand and university staff engaged in developing and implementing emerging FAIR leading practices on the other. During processes such as these, it is of crucial importance to focus primarily on the needs of researchers. We describe how LU builds up its support for FAIR data before, during and after research through its involvement in leading practices, training and consultancy and end with recommendations for other universities wanting to implement the FAIR principles.

Keywords: FAIR data; Research Data Management; Leading Practice; Implementation

Introduction

In April 2016, Leiden University (LU) launched its research data management (RDM) regulation (Research Data Management Regulations Leiden University 2016). This regulation was formulated to make it easier for researchers to comply with RDM requirements formulated by funders and other external parties to enhance the transparency and the integrity of research. The RDM regulation was inspired by the FAIR Principles (ensuring that data and software are Findable, Accessible, Interoperable, and Reusable by humans and machines), which were initially formulated in 2014 (Jointly designing a data FAIRPORT 2014), published in 2016 (Wilkinson et al. 2016) and now widely adopted by research funding and research performing organisations. The LU RDM regulation applies to three distinct stages of data management: before, during and after research. It states, for example, that a data management plan (DMP) must be drawn up before the start of the data collection phase. The DMP consists more concretely of a further specification of the data management protocols of the faculty or institute for the specific research project in question. During the research project, research data must be preserved securely, meaning that the integrity, availability and – if required – confidentiality of the data, must be guaranteed. After the formal completion of the research project, the data must be managed in such a way that they are findable, accessible, comprehensible and reusable in the long term. This means that data must be stored together with the metadata, documentation and possibly the software required for its reuse. To better ensure the secure and sustainable access to the data, they should preferably be preserved in a repository certified with the CoreTrustSeal (CoreTrustSeal 2019). The regulation stipulates a minimum retention term for research data of ten years.

Following the publication of the LU RDM Regulation, faculties were asked to translate the general principles in this regulation to concrete guidelines, following the needs of individual disciplines. A data management implementation programme was set up to support six faculties, with an expected finish date at the end of 2020. Members of the project team include data management experts from the Centre for Digital Scholarship (CDS) at LU Libraries, policy advisors from Academic affairs, research ICT experts from LU ICT Shared Service Centre and RDM experts from the faculties. The CDS is responsible for information provisioning, advice and training. During the first phase of the program, the CDS started a range of activities to raise awareness of the activities needed to make data FAIR. For example, a Research Data Service Catalogue listing resources essential for research data management at LU was released in 2016 and revised in 2020 (Leiden Research Data Service Catalog 2020). In addition to this, the CDS organises regular workshops (occurring every six weeks) on how to write DMPs. Next to general workshops, the CDS also organises DMP workshops tailored to the needs of specific research institutes. Both types of workshops include a general introduction to the FAIR principles. During the customised workshops, students are also asked to evaluate the 'FAIRness' of a number of studies in their discipline. The participants are encouraged to revise or to update their DMP during their project if necessary. There is no formal assessment of the workshops, but feedback is asked from participants.

The above activities concentrated on making data FAIR for humans as a first step. With FAIR for humans we mean that human researchers can find the data set in a repository, download it, understand it (because it has a README file describing it) and reuse it for their own purposes (because it has a permissive license attached to it). The process of making data FAIR for human researchers demands relatively little effort, but it is still a huge leap forward, compared to the situation in which the data are stored exclusively on the researcher's own hard drive, for instance. The second phase of the data management implementation programme, which started in 2019, had a much stronger focus on Interoperability and Reusability and the machine actionability of data as stated in the FAIR principles. The CDS believes that most of the principles underlying Findability and Accessibility can be implemented on a university-broad level. In contrast, the principles underlying Interoperability and Reusability can be implemented more effectively at faculty-level and institution-level since these more frequently require domain-specific knowledge.

The approach that was adopted at Leiden University was strongly informed by the work of international organisations such as GO FAIR, the Research Data Alliance (RDA) and CODATA, which all formulate and recommend leading practices for the implementation of the FAIR principles. The experiences and lessons learned from the local implementations at the faculty and institutional level can in turn be used to strengthen these international leading practices. In this paper we describe how the CDS tries to close the feedback loop between international leading practices and the LU implementation during this second phase of the data management programme, maintaining a clear focus on the researcher. While the implementation of RDM policies, support and services (see for example Cruz et al. 2019 or Mushi et al. 2020. For a survey of research data services implementation see Tenopir et al. 2014) have already been discussed in a range of other publications, the various ways in which universities can advance the interoperability, reusability and machine-actionability of research data, in agreement with the FAIR principles, has still received little attention, to our knowledge. By describing the approach within LU this paper aims to contribute to this emerging debate.

Approach

The primary focus of the CDS is to deliver the best possible support to all researchers affiliated with Leiden University. We believe that this can best be achieved by making researchers aware of relevant leading practices, by training them in how to make their data FAIR and by providing individual consultancy when needed. To give state-of-the-art support, the CDS needs to keep up with the pace of the fast-developing international field around FAIR research data. Especially the implementation of the FAIR principles for Interoperable and Reusable data requires advanced knowledge that can be found more easily at an international level. Ideally, by actively taking part in current developments related to FAIR data, we can learn from leading practices for implementation solutions and translate these to relevant researcher support while influencing the pace of the availability and the suitability of these solutions to Leiden researchers. In return, we also share our experience with the organisations providing these leading practices. We therefore engage closely with the international bottom-up organizations GO FAIR, RDA and CODATA (**Figure 1**). Below, we describe our activities within three areas: leading practices, training and consultancy. We also discuss how these activities contribute to our researcher support before, during and after research projects.

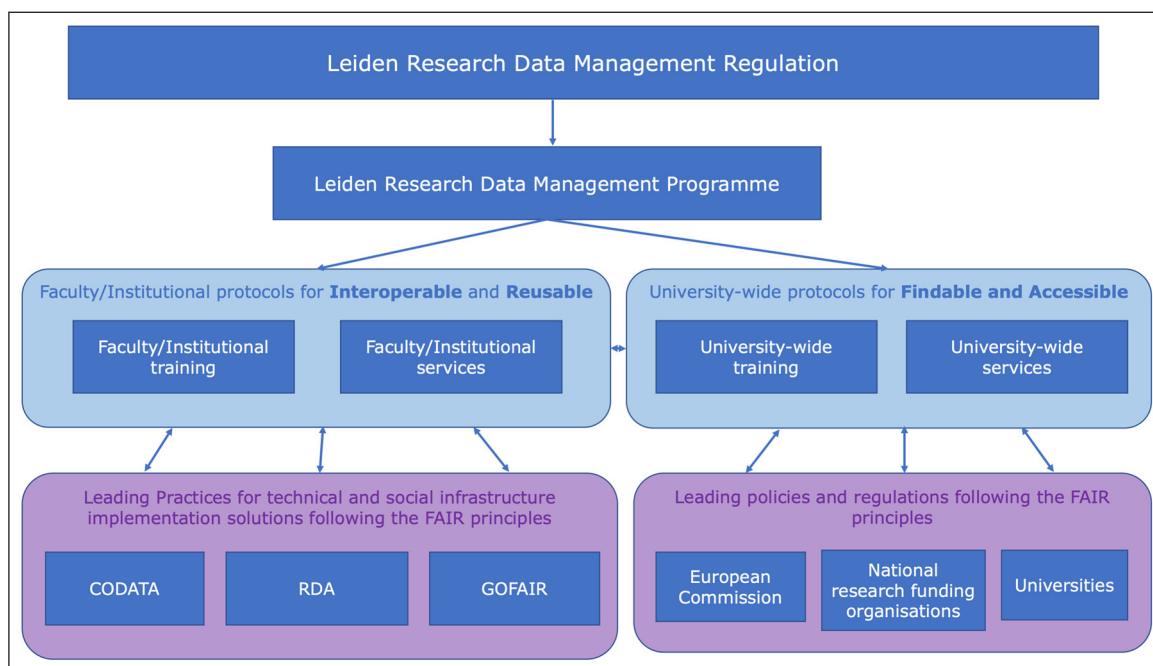


Figure 1: Schema showing the feedback loop between institutional protocols and international leading practices in the context of Leiden University.

Leading Practices

At the end of 2018, The CDS began its active involvement in two international activities to develop leading practices for implementation solutions for FAIR data: the FAIR funders pilot programme (FFPP) (Wittenburg et al. 2019) and the FAIR Implementation Matrix (Sustkova et al. 2020). Both activities were initiated and led collectively by GO FAIR and RDA. The FFPP wanted to make it easier for grantees to produce FAIR data by using tools and services for creating DMPs and FAIR metadata. The FAIR Implementation Matrix was aimed at encouraging communities to record their technical implementation choices for FAIR, and to work towards a degree of convergence around the FAIR leading practices. These two activities resulted in several workshops and preliminary results that were tracked in the Open Science Framework project (FAIR Implementation Matrix, 2019). In 2020, both initiatives entered a phase of rapid development because of the COVID-19 pandemic. It prompted the birth of the Virus Outbreak Data Network (VODAN) as a joint activity of CODATA, RDA, WDS, and GO FAIR (VODAN 2020). The FFPP and FAIR Implementation Matrix became core activities within VODAN and the CDS decided to intensify its involvement and contribute actively to the VODAN network. Led by GO FAIR, a three-point framework for FAIRification (3PFFF 2020) was launched on the 8th of July 2020. One of the key components of the three-point framework for FAIRification are Metadata for Machine (M4M) workshops (Metadata for Machines Workshops 2019). In a M4M workshop data stewards and researchers collaborate to decide on data policy issues and metadata descriptions needed to ensure FAIRness and render their decisions in a machine-actionable form (a metadata schema). The metadata schema produced in the M4M form part of the larger FAIR Implementation Profile (FIP) (Magagna et al. 2020), which in turn guides the configuration of a FAIR Data Point (a metadata repository that provides access to metadata in a FAIR way)(FAIR Data Point Design Specification 2020). The three-point framework for FAIRification was adopted by the Dutch National COVID-19 Research Program from the research funding agency ZonMw, among others. In theory, these activities could have given the CDS a concrete opportunity to explore the added benefit of being actively involved in the development of these leading practices when supporting Leiden researchers applying for funding within this program. In practice, however, we noticed that the CDS was only contacted by researchers at a late stage when the proposal was already almost completed (in most cases just days before the deadline or even the same day). We can only speculate that the quality of these proposals would have improved if we had been involved at an earlier stage and if we could have worked out the details from a FAIR data stewardship perspective. We did note, however, that we could offer valuable advice to researchers about the budget for data stewardship, since we knew, from our involvement in VODAN, what ZonMw was expecting in terms of making data FAIR. We will continue to be involved in the developments of the three-point framework for FAIRification and plan to offer the M4M and FIP

workshops and to set up FAIR Data Points for researchers from any discipline as a way to guide researchers in their decisions how to make their data FAIR for machines. In terms of research support, M4M and FIP workshops should be performed at the beginning of a research project, to make sure that there is still enough freedom to choose the appropriate metadata and data standards.

Training

Researchers need to be trained in FAIR data skills to enable them to communicate effectively with data experts when making their data FAIR. To understand the skills needed, the CDS contributes to the development of a competency framework to describe the skills and knowledge required to do FAIR-related work in a particular discipline. The first workshop (A terminology for FAIR Stewardship Skills Workshop 2019) was organised by CODATA, the Digital Curation Centre, Dutch Techcentre for Life Sciences (DTL)/ELIXIR-NL, FAIRsharing and Royal Holloway. The competency framework is work in progress and a first version was released after the second workshop in October 2019 (FAIR terminology 2019). Some skills are generic, such as data modelling, but to actually model their own data, researchers need to be aware of data formats and metadata standards *in their own field*. As explained previously, the M4M workshops help communities to define/declare their metadata standards while the FAIR Implementation Profile records them. However, these activities are directed at research communities, individual researchers need other training materials to learn about how they can make their data FAIR.

To create discipline-specific and understandable FAIR data resources aimed at single researchers, the CDS participated as collaborators in the international Library Carpentry Top 10 FAIR Data and Software Things sprint in November 2018. During this sprint, training materials were developed for specific academic disciplines including history, biodiversity and archaeology (Erdmann et al. 2019). The CDS lead the sprint on the history resource. The Top 10 FAIR Data Things are intended as living resources that can be edited and expanded by the community. Most of these resources have researchers as their primary audience, but they can also be used as a source by librarians during their efforts to raise awareness of the FAIR principles. In fact, when developing a workshop entitled 'Let your research bloom: practical steps for FAIR data' for the Science PhD Day at the LU, the Top 10 FAIR Data and Software Things resources helped to identify the four first steps to make data more FAIR in a very productive manner. As explained in the Top 10 FAIR Data Things, researchers need to upload their data to a repository (F), decide who has access to the data (A), describe the data using the metadata scheme offered by the repository (I), and choose a license (R). In May 2019, another Library Carpentry international sprint was held, this time in collaboration with Mozilla. The CDS coordinated the sprint on the Top 10 FAIR Things for Astronomy together with the Vrije Universiteit in Amsterdam, the Netherlands and the University of Notre Dame in the United States. The resulting document has been added to the Top 10 FAIR Things resource on Zenodo (Erdmann et al. 2019). The resource on Zenodo has been consulted quite frequently (4315 unique views, data accessed on 28 September 2020) and, since 19 May 2020, it has been recognised by the RDA as an endorsed output. We stimulate Leiden researchers to contribute to these sprints. The involvement of a Leiden researcher in the creation of the Top 10 FAIR Things for Astronomy was a fruitful experience, both for the researcher as for the data stewards in the sprint. In terms of research support, the material is broad and could fit anywhere during the research project, for example right at the start when there is a need to create "FAIR awareness" or at the end since it contains guidelines interoperability and reusability. To meet the need for researchers to work with standards when "wrangling" their own data, the CDS organised a 'Bring Your Own Data FAIRification' pilot workshop for early career researchers at the Leiden Institute for Advanced Computer Science in June 2019. The course was based on information from the first draft of the 'Essential steps of the FAIRification process' book (Schultes et al. 2019). Similar to the Top 10 FAIR Things, it is expected that the book will evolve into a dynamic resource that will be updated by the community. Following the very positive evaluation, the 'Bring Your Own Data' workshop will be followed by similar events at other LU institutes. It complements the M4M and FIP workshops mentioned in the leading practices section by focussing on the data wrangling while the other workshops focus on documenting the standards that need to be used. Given these qualities, M4M and FIP workshops ideally take place before the data collection and a Bring Your Own Data workshop is best organised after data collection.

Consultancy

While it is crucially important to develop good educational resources and to organise interactive training sessions, such initiatives are not always sufficient. Research projects often work with highly complicated data sets, and researchers sometimes lack the skills and the knowledge to manage their data ensuring maximum reuse. Because the questions that researchers may have cannot always be adequately addressed

during training sessions, it is also necessary to be willing to make dedicated appointments with individual researchers and to offer consultancy about specific topics in the field of data management. In the course of the last few years, the CDS has offered advice on a wide range of activities across the entire research life-cycle. Most of the questions that we have received were about database design and about data modelling. A large number of researchers have also asked for advice on how to publish data in compliance with the FAIR principles (Imming et al. 2018). In 2018 and 2019, the CDS has participated actively as a partner in an educational project which concentrated on the FAIRification of an existing born-digital scholarly archive (Verhaar 2019). The project was initiated by a professor of book history at LU who had produced a large number of text documents containing semi-structured research annotations in the course of his academic career. To improve the reusability of the data, the CDS has helped to convert the semi-structured research annotations into a searchable MySQL database, based on a well-considered data model. The entries in this database were connected to entries in Wikipedia, so that the new data set could be integrated more effectively into existing data sets. The project continues to evolve every year since it is embedded in the curriculum for the MA programme on Book and Digital Media studies at the LU. Therefore, students continue to work on the database and related projects.

The CDS has also participated in a project which aimed to expose the legacy data of the LU Centre for Linguistics in a FAIR format. This project was taken up as a case study in the Leiden RDM programme. In 2019, the LUCL sent out a questionnaire and performed interviews with researchers at the centre to investigate what type of legacy data they knew about. 31 questions were asked, concentrating, among other aspects, on the languages that were studied, the data formats and the curation needs. The survey resulted in information about 128 datasets. The metadata about these datasets was entered into an online, searchable database, developed by the CDS. During this process, the CDS also advised on the standards to use in the description of these data sets (Leiden Linguistics Data 2020). The project is still ongoing and the next steps include archiving the metadata in a repository and making the actual datasets FAIR.

The participation in projects such as these helped us to develop a better understanding of the concrete steps that can be taken to make data more FAIR. With the development of the three-point framework for FAIRification, we are beginning to get more tools for our FAIR toolkit to benefit these types of projects. Both the book history project and the LUCL legacy data project could set up a FAIR Data Point, for instance, and a M4M and FIP workshop could be arranged for a project similar to the LUCL project, to stimulate discussions and decisions on metadata and data standards.

Lessons Learned

The CDS uses the leading practices recommended by GO FAIR, RDA and CODATA to support the implementation of FAIR data within LU faculties and institutes. Like the LU RDM regulation, the leading practices are broad, and there is a need for a translation to protocols for faculties and institutes, as well as for training and consultancy to accompany such protocols. We recognise that, as universities, we must take on an active role in the feedback loop going from leading practices to implementation at faculty and institutional level and vice versa. By collaborating closely with GO FAIR, RDA and CODATA, amongst other organisations, we help to co-develop leading practices and implementation guidelines for FAIR, in a manner that is informed by first-hand experiences in the CDS training sessions and by experiences for supporting actual projects to create FAIR data. Such collaborations have the obvious advantage that we can deliver machine-actionable FAIR data support and training to researchers, next to input for institutional protocols that is informed by latest developments. However, this can only be achieved when data stewards have the means and the skills to implement not only the principles of Findability and Accessibility but also the much principles of Interoperability and Reusability, which unfortunately are less straight-forward.

Until very recently, the CDS was partly supported financially by a LU innovation fund and partly by the LU Libraries. After a successful internal evaluation, based on a questionnaire sent out to all researchers and other users of our services, the CDS is now being funded structurally from the LU central budget. The evaluation specifically stated that the involvement in international initiatives as stated above is important for the LU, allowing for quick action in response to new developments. Specific projects have been funded either from LU research or educational grants, or by external grants, but it has always been the goal to have a core, central, structural funding for the team. This stable funding, in combination with the collaborative skills and the expertise of the members of the CDS team, who have been trained in fields such as Information Technology and Computer Science, form the conditions under which the CDS can actively participate in the development of international Leading Practice. The situation in which data stewards are employed by a specific research project only is clearly undesirable, as there will generally be fewer opportunities for them, in that case, to be

involved in similar efforts. The generic tasks of a data steward are to develop RDM infrastructure, policies, and support services. To be able to do this and to further the development of the data stewardship profession itself the data steward needs to balance these activities with innovation. We argue that a data steward should possess a number basic skills in the fields Information Technology and Computer Science, next to innovation skills and collaborative skills. In addition to this, the data steward must be given the opportunity in form of time and funding to take part in leading practices activities to develop the data stewardship profession further. This is important for the less affluent institutes as well, in order to build future-proof FAIR data support. Institutes with smaller budgets can however start on a small scale by being involved in one single working group of one international leading practices organisation, to advance the support in at least one area of FAIR.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

All authors wrote the paper collectively.

References

- 3PFFF.** 2020. Available at <https://www.go-fair.org/2020/07/08/a-three-point-framework-for-fairification/> [Last accessed 2 October 2020].
- A terminology for FAIR Stewardship Skills Workshop.** 2019. Available at <https://terms4fairskills.github.io/Announcement.html> [Last accessed 26 July 2019].
- CoreTrustSeal.** 2019. Available at <https://www.coretrustseal.org> [Last accessed 26 July 2019].
- Cruz, M,** et al. 2019. Policy Needs to Go Hand in Hand with Practice: The Learning and Listening Approach to Data Management. *Data Science Journal*, 18(1): 45. DOI: <https://doi.org/10.5334/dsj-2019-045>
- Erdmann, C,** et al. 2019. Top 10 FAIR Data & Software Things. *Zenodo*. DOI: <https://doi.org/10.5281/zenodo.2555498>
- FAIR Data Point Design Specification.** 2020. Available at <https://github.com/FAIRDataTeam/FAIRDataPoint-Spec> [Last accessed 2 October 2020].
- FAIR Implementation Matrix.** 2019. Available at <https://osf.io/n7uwp/> [Last accessed 2 October 2020].
- FAIR terminology.** 2019. Available at <https://github.com/terms4fairskills/FAIRterminology> [Last accessed October 2020].
- Imming, M,** et al. 2018. FAIR Data Advanced Use Cases: from principles to practice in the Netherlands. *Zenodo*. DOI: <https://doi.org/10.5281/zenodo.1250535>
- Jointly designing a data FAIRPORT.** 2014. Available at <https://www.lorentzcenter.nl/lc/web/2014/602/info.php3?wsid=602> [Last accessed 26 July 2019].
- Leiden Linguistics Data.** 2020. Available at <http://leiland.ullet.net> [Last accessed 2 October 2020].
- Leiden Research Data Service Catalog.** 2020. Available at <https://digitalscholarship.nl/rds/> [Last accessed 2 Oct 2020].
- Magagna, B,** et al. 2020. Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. *OSF Preprints*. DOI: <https://doi.org/10.31219/osf.io/2p85g>
- Metadata for Machines Workshops.** 2019. Available at <https://www.go-fair.org/resources/go-fair-workshop-series/metadata-for-machines-workshops/> [Last accessed 26 July 2019].
- Mushi, GE,** et al. 2020. Identifying and Implementing Relevant Research Data Management Services for the Library at the University of Dodoma, Tanzania. *Data Science Journal*, 19(1): 1. DOI: <https://doi.org/10.5334/dsj-2020-001>
- Research Data Management Regulations Leiden University.** 2016. Available at https://www.bibliotheek.universiteitleiden.nl/binaries/content/assets/ul2ub/research--publish/research-data-management-regulations-leiden-university_def.pdf [Last accessed 26 July 2019].
- Schultes, E,** et al. 2019. Available at <https://osf.io/sjzc8/> [Last accessed 26 July 2019].
- Sustkova, HP,** et al. 2020. FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-Related Resources. *Data Intelligence*, 2(1–2): 158–170. DOI: https://doi.org/10.1162/dint_a_00038
- Tenopir, C,** et al. 2014. Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, 36(2): 84–90. DOI: <https://doi.org/10.1016/j.lisr.2013.11.003>
- Verhaar, P.** 2019. Available at <https://digitalscholarship.leiden.nl/articles/durable-access-to-book-historical-data> [Last accessed 2 October 2020].

- VODAN.** 2020. Available at <https://www.go-fair.org/implementation-networks/overview/vodan/> [Last accessed 2 October 2020].
- Wilkinson, MD,** et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3: 160018. Erratum in: 2019 *Scientific Data*, 6(1): 6. DOI: <https://doi.org/10.1038/sdata.2016.18>
- Wittenburg, P,** et al. 2019. The FAIR Funder pilot programme to make it easy for funders to require and for grantees to produce FAIR Data. *arXiv*. <https://arxiv.org/abs/1902.11162>.


How to cite this article: Hettne, KM, Verhaar, P, Schultes, E and Sesink, L. 2020. From FAIR Leading Practices to FAIR Implementation and Back: An Inclusive Approach to FAIR at Leiden University Libraries. *Data Science Journal*, 19: 40, pp. 1–7. DOI: <https://doi.org/10.5334/dsj-2020-040>

Submitted: 26 July 2019

Accepted: 07 October 2020

Published: 23 October 2020

Copyright: © 2020 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.

OPEN ACCESS 