

When lions play leapfrog! Adopting the CODATA-RDA Data Science Summer School curriculum for introductory data science training



The challenge

Postgraduates across Africa do not have access to the foundational data science and open research training that would allow them to become part of a growing community of researchers, who, if not fully skilled, would at least understand the rules of entering the arena of modern, open science. Such training material is available, but a curriculum has not yet been decentralised and localised for Africa because it is not clearly understood where to place the hubs of knowledge from which the training could be disseminated. There is general consensus that introductory level data science training is essential for all research disciplines – to the extent that some of us see it as an element of the digital literacies. South Africa did not, at the time of us adopting the curriculum, see any institution taking on a mandate to provide such training. From our Research Data Management (RDA) implementation experience, we also knew that it would be some time before generic training - which would put any post-graduate on a reliable path to understanding the data science ecosystem - would be developed in South Africa. We understood that gaining a holistic view of what is needed as foundation training requires a number of stakeholders to collaborate. The challenge was too many issues to address simultaneously if we wanted to start from scratch.

The RDA outputs adopted

The International Data Week in Botswana (November 2018) led to the identification of the CODATA-RDA Summer School curriculum as a possible way to address our need. The University of Pretoria Department of Information Science took the lead in providing a home for the initiative, and our selected partners assisted us in reviewing the curriculum for relevance. We found an excellent fit in the CODATA-RDA Summer School curriculum for data science. We needed to localise the content, but in essence we were ready to go.



Summer School participants gaining hands-on experience in data science practices.

Find out more at:
www.rd-alliance.org/recommendations-outputs

Visit rd-alliance.org or write us at enquiries@rd-alliance.org

Benefits of adoption and impact

The most important benefit is that we were able to leapfrog from an existing, tried and tested initiative – which saved us considerable time. In addition, the knowledge that the training is also being rolled out to an international community gave us the assurance that we were on the right track. The alumni network linked to the initiative is another very important benefit. Our candidates were immediately pulled into a professional network that would, under different circumstances, take many years to develop. Similarly, the exposure to peers from different disciplines has created shared jargon and experience that we firmly believe will show impact in future.

The adoption process

We decided to run the first Summer School as a trial or pilot effort, and that we would then take a firm decision about the responsibility of taking on such an initiative after the completion of the school. A Steering committee, which included two CODATA-RDA co-chairs, was constituted in March 2019. The target date for our first South Africa Summer School was set for January 2020, to coincide with our Summer holidays, and to be completed just before the start of the new academic year. We had very good support and guidance from the appointed co-chairs, and it worked in our favour that one of our partners is a Summer School Trieste alumnus. We took firm decisions regarding the use of local trainers, who to approach for funding, and what aspects of the content to localise. We had really commendable assistance from those who had developed the content to the extent that the majority of

trainers met with our trainers online and answered several questions regarding the reasoning behind including selected content. In the absence of a suitable application system we gained assistance from Brazil and were able to adapt the system that was used for the Costa Rica School. It was a tight squeeze to get all the administration completed before the summer holidays started in early December, but it worked out well and we were ready and able to put all participants to work in January.

Because of COVID-19 limitations and an extended academic year, the 2021 Summer School will take place after March and much of the content will be provided online. To counter some of the expected constraints the time period will be extended allowing for some asynchronous learning.



Summer School participants gaining hands-on experience in data science practices.

Find out more at:
www.rd-alliance.org/recommendations-outputs

Visit rd-alliance.org or write us at enquiries@rd-alliance.org

Lessons learned

There were several lessons taken from this experience. Most of these inspired us to attempt to bring more partners on board and to commit to arranging further Summer Schools, using the University of Pretoria Department Information Science as the coordinating body for at least the next three years. The feedback we received from both participants (our 'young lions') and trainers was very positive. We think that the 'hand-holding' and advice from the co-chairs is essential for a first instance of such a Summer School. The same is true for the guidance we received from more experienced trainers. The knowledge sharing and knowledge transfer is indicative of a healthy learning ecosystem. We think that it was a wise decision to make use of local trainers and to localise some of the content. As we all know, leapfrog is not only about leveraging that which exists, but also about doing your bit to ensure that the system makes progress. There is much gratitude that it was possible to gain access to the training material and the planning related to this Summer School initiative. We have saved effort and much time.

About the Organisation

University of Pretoria (UP)'s Department of Information Science in collaboration with the Data Intensive Research Initiative of South Africa (DIRISA), South African Centre for Digital Language Resources (SADiLaR) and the Network of Data and Information Curation Communities (NeDICC) and Data Carpentry SA initiated the first school in South Africa with the help of three co-chairs respectively from the Institute of Science Innovation and Society, University of Oxford, the Royal Holloway, University of London and Tofail University, Morocco. As of 2021 the University of Cape Town is also represented on our Steering Committee.

Summer School Steering committee members also teaching at the Summer School, from left to right: Juan Steyn (SADiLaR), Prof Marlene Holmner (UP), Dr Louise Bezuidenhout (University of Oxford, UK), Dr Bianca Peterson (NWU), Sara El-Jadid (Tofail University, Morocco) and Dr Martie van Deventer (UP). Johann van Wyk (UP Library Services / NeDICC) was not present when the photo was taken and Prof Hugh Shannahan (Royal Holloway University of London, UK) participated via Skype.



Find out more at:
www.rd-alliance.org/recommendations-outputs

Visit rd-alliance.org or write us at enquiries@rd-alliance.org