

RDM TRAINING FOR PHD STUDENTS

Bernard Jacquemin¹, Joachim Schöpfel¹, Eric Kergosien¹, Cécile Malleret²

¹University of Lille, GERiiCO laboratory, 59653 Villeneuve d'Ascq, France
bernard.jaquemin@univ-lille.fr, joachim.schopfel@univ-lille.fr, eric.kergosien@univ-lille.fr

²University of Lille, Academic library, 59653 Villeneuve d'Ascq, France
cecile.malleret@univ-lille.fr

Abstract: Objective: Since 2015, we have trained more than 50 PhD students from several disciplines in social sciences and humanities in research data management (RDM). From this experience, we propose a detailed assessment of the outcomes and shortfalls of the program. The objective is to reveal program specificities, key factors of success and barriers, in order to improve the contribution of such program to open scholarship and to optimize the preparation of young scientists for the new world of open science.

Method: The assessment of the training program is based on direct feedback and on an online survey with participants and includes a comparison with similar programs in other academic institutions.

Findings: Students are generally satisfied with the personalized and pragmatic approach of the Lille training program. They ask for more “real life examples” from research projects and for more training on metadata formats and on XML. The online survey (in progress) will provide additional suggestions and critics to improve the program content and schedule. The comparison with other RDM training programs assesses 10-12 criteria, including the duration and schedule of the program and its objective and content. Such evaluation may be helpful to identify gaps, key factors of success and best practice.

Keywords: Research data management, PhD students, training program, data management plan, open science, open scholarship.

INTRODUCTION

Recently, the French Minister for Higher Education and Research, Mrs Frédérique Vidal, made a speech at the Sorbonne University on open science, calling in particular for training in research data management (RDM) in graduate schools in order to improve the data literacy and teach young scientists good practice in RDM. Three years ago, the University of Lille implemented the first French RDM training program for PhD students in social sciences and humanities. This program is run by the Lille laboratory of information and communication sciences, GERiiCO, together with the university library and the graduate school in social sciences and humanities. Open science means to facilitate the availability of scientific results to the research community and to society, in particular to publish scientific papers via open repositories or open access journals and to share research data whenever possible. Part of the open science, research data management, ie data curation, preservation and sharing become good scientific practice, obligatory academic skills along with statistics, ethics or information retrieval.

Since 2015, we have trained more than 50 PhD students from psychology, linguistics, history, information sciences, sociology and other disciplines, from all universities in the Northern region of France (Lille, Valenciennes, Arras, Boulogne, Amiens...). From this experience, we propose a detailed assessment of the outcomes and shortfalls of the program. The objective is to reveal program specificities, key factors of success and barriers, in order to improve the contribution of

such program to Open Scholarship and to optimize the preparation of young scientists for the new world of Open Science.

METHODOLOGY

The assessment of the training program is based on direct feedback from the students, on the evaluation of the DMPs, and on an online survey with participants. Additionally we compared our program with similar initiatives in other academic institutions, as for instance at the Universities of Helsinki, Berlin, Cambridge and Oregon.

FINDINGS

The training program started in 2015 with three conferences on the challenge of research data management, on legal aspects and on the potential of content mining for research data in dissertations, and some preliminary results have been presented in 2016 at 19th International Symposium on Electronic Theses and Dissertations in Lille (France) and in 2017 at the 10th Conference on Grey Literature and Repositories in Prague (Czech Republic).

The Lille program provides an introduction in RDM with specific workshops on data curation and structuring, and helps the students to write their own data management plan (DMP), for their doctoral research project. It includes ethics, legal issues and data security. The program is designed for an entire semester, with seven sessions during six months and homework. The program didactics combines lecturing, interactive learning, and personalized follow-up. The training takes place in a computer room of the university library and is assisted by the French online platform DMP-OPIDoR, for the creation of the individual DMP. Each DMP is written with the European Research Framework Horizon 2020 FAIR DMP template, shared on the DMP-OPIDoR platform and commented by the training team. The result, ie the final version of the DMP, is evaluated. The program is open to PhD students from all disciplines in the field of arts, social sciences and humanities.

The feedback from the students shows that they are generally satisfied with our personalized and pragmatic approach. They ask for more “real life examples” from research projects, for more training on metadata formats and on XML. On the other side, they are less interested in general information on open science and open access, and they don’t want any specific “PhD data plan template”. Also, they ask for a better coordination with other, local or national activities in the field of data management and open access.

The online survey (in progress) will add complementary information to this feedback, including (hopefully) suggestions on how to improve the program content.

The comparison with other RDM training programs will assess 10-12 criteria, including the duration and schedule of the program, its objective and content, the governance, the evaluation and didactics, etc. The landscape is changing rapidly, and such evaluation may be helpful to identify gaps, key factors of success and best practice.

CONCLUSIONS

The Lille RDM training program for PhD students has been developed as part of an institutional initiative to improve the data literacy of PhD students in arts, social sciences and humanities. The objective is to help them for their PhD research project and moreover, to teach them good practices for their future academic career. We do not focus on “open data marketing” but try to

contribute to an academic data culture in social sciences and humanities, as one dimension of open science.

Our model has some particular and original features, such as the integration in the graduate school program, the focus on data management plans and the partnership between library and research. Of course, one size does not fit all. Each academic institution must develop its own data program, in the particular local context. However, we are convinced that the success of our training program may incite other colleagues to develop their own initiative, and that it can provide some inspiration on how to do it.

REFERENCES (SELECTED)

Baaske, F., R. Gerlach, and B. König-Ries (2018). Teaching and training concepts on research data management at the Friedrich Schiller University Jena. In *CODATA 2018*, March 19, 2018, Göttingen.

Prost, H., Malleret, C., Schöpfel, J., 2015. Hidden treasures. Opening data in PhD dissertations in social sciences and humanities. *Journal of Librarianship and Scholarly Communication* 3 (2), eP1230+. <http://dx.doi.org/10.7710/2162-3309.1230>

Schöpfel, J., Chaudiron, S., Jacquemin, B., Prost, H., Severo, M., Thiault, F., 2014. Open access to research data in electronic theses and dissertations: An overview. *Library Hi Tech* 32 (4), 612-627. <http://dx.doi.org/10.1108/LHT-06-2014-0058>

Schöpfel, J., Prost, H., Malleret, C., 2017. Research and development in the field of research data and dissertations. The D4Humanities project at the University of Lille (France). In: *10th Conference on Grey Literature and Repositories*, National Library of Technology (NTK), 19 October 2017, Prague, Czech Republic. http://repozitar.techlib.cz/record/1218/files/idr-1218_1.pdf

Whitmire, A. L. (2015). Implementing a Graduate-Level research data management course: Approach, outcomes, and lessons learned. *Journal of Librarianship and Scholarly Communication* 3 (2), eP1246+. <http://doi.org/10.7710/2162-3309.1246>

Wiljes, C. (2018). Teaching research data management skills to students. In *CODATA 2018*, March 19, 2018, Göttingen.