PhD training on Open Science in French universities

Joachim Schöpfel
University of Lille, GERiiCO laboratory, France
joachim.schopfel@univ-lille.fr

Hélène Prost
CNRS, GERiiCO laboratory, France
helene.prost007@gmail.com

Bernard Jacquemin
University of Lille, GERiiCO laboratory, France
bernard.jacquemin@univ-lille.fr

Eric Kergosien
University of Lille, GERiiCO laboratory, France
eric.kergosien@univ-lille.fr

Abstract

The development of open science requires a cultural change in academic institutions, including the acquisition of new knowledge and skills in fields like open access publishing, research data sharing and citizen science. The paper presents results from a survey on PhD training programs related to open science in ten highly ranked and research-intensive French universities. Based on the discussion of the empirical survey results (content, format, discipline, etc.), the paper establishes a list of some recommendations that may be helpful for the assessment of existing programs and for the development and implementation of new programs.

Keywords: PhD training, graduate education, open science, open access, research data, citizen science, research universities, early career researchers

Introduction

PhD students, as early career researchers, will be “harbingers of change” of the traditional system of academic science and publishing (Nicholas et al. 2018). Along with social media and collaboration, open science is one key factor driving this change. Yet, this change
may take more time than expected, especially regarding practices in the field of open science. What they need is a favourable, rewarding environment, i.e. incentives for open access publishing and data sharing and even before, opportunities to acquire awareness and knowledge about open science and to develop new, open science relevant skills.

Since 2015, the University of Lille has developed PhD training in the field of open science, a doctoral seminar with a focus on research data management in social sciences and humanities (SSH). The Lille model called DRTD (Données de Recherche dans les Thèses de Doctorat) is based on a partnership between the graduate school in SSH, the academic library and the GERiiCO laboratory in information sciences, communication and cultural studies. The project was presented at last year’s ETD conference (Schöpfel et al. 2018). In 2019, we conducted a survey on PhD training programs of the ten most highly ranked French universities. The objective was to identify open science related contents in the graduate schools’ educational programs, to produce empirical evidence on good practice in this field and, as a preliminary work package, to prepare a larger international research project on the impact of open science on PhD dissertations. In the following, we will provide a short literature overview, present the main results of the survey and discuss them in terms of good practice, community, marketing and further perspectives.

**Literature review**

In their advice paper on open science and the role of universities, the League of European Research Universities (LERU) highlights the education and skills training of “all people in the universities” (scholars and scientists, research management staff, data scientists, copyright officers, librarians...) as one of the eight pillars of open science (Ayris et al. 2018). Following LERU, open science skills training is particularly beneficial to doctoral researchers at the beginning of their scientific career.

LERU identifies five essential dimensions for the training in open science: “Clearly, there is an evident need for skills training with regard to scholarly publishing and research data management; those are the areas of Open Science in which universities tend to invest most at the moment. Also, research integrity and ethics courses, and increasingly, citizen science courses, are important”. The LERU report insists that such training should be tailored to the specific needs, resources and requirements of the audience, including a large variety of formats such as in–person or distance, classroom, webinars, blended or not.

Many universities develop and deliver their own training, others work with external providers. Of course, researchers do acquire similar skills also in informal training situations, on–the–job, but regarding skills development by doctoral researchers, LERU states that “Open Science skills training should be firmly embedded (in online progress tools or similar study management and supervisory systems) and should be acknowledged in professional development and career progression”. Also, LERU recommends that universities integrate open science concepts and its practical applications in educational and skills development
programmes, analysing and mapping their needs for Open Science skills training, taking into account the different Open Science dimensions and the varying needs of different audiences, different disciplines, etc.”.

Based on another survey with scientists, a European Commission working group has defined open science skills and expertise needs for researchers so that skilled talent can publish under open access, manage (open) research data, conduct professional research (including research management, legal expertise, ethics and integrity) and engage with citizen science. The final report recommends the introduction of open science education and training tailored for PhD students, the inclusion of open science modules with credits in all European doctoral training programmes by 2020, and the linking of the Innovative Doctoral Training principles1 to open science practices, to encompass an open research environment. “Institutions should offer and promote both traditional and/or online career-level appropriate Open Science training courses for researchers (…) All Open Science skills courses should have career level appropriate accreditation and could also be modularised. In the case of (PhD students), it should be mandatory for universities and research organisations to offer these as part of their training” (O’Carroll et al. 2017).

Several universities launched training programs on open science related topics, in order to raise awareness and/or to teach relevant skills, especially in the field of research data management (Baaske et al. 2018, Wiljes 2018). Whitmire (2015) presents a case study on a “discipline agnostic, credit-bearing course in research data management for graduate students (…) that combined outcomes centred course design and active learning”. The program includes the research lifecycle and data management planning, storage, backup and security, metadata, legal and ethical considerations, data sharing and reuse, and archiving and preservation.

On a more general level, Bogle et al. (2016) highlight the importance of high-quality doctoral education for research-intensive universities. They provide examples of good practice and recommendations on “how universities develop, maintain and evaluate their high quality culture in doctoral education, on how they go about achieving their goal to train doctoral researchers to the highest skill levels to become creative, critical and autonomous intellectual risk takers, and on how they focus on stimulating a rigorous research culture (…)”. Citing this survey and other studies, LERU produced a policy brief on the importance of investment in training of early stage researchers, especially doctoral education, with benefits not only for the scientific communities and structures themselves but above all, for society as a whole (LERU 2016).

In our own study presented at ETD 2018, we compared the format and objectives of the Lille PhD training program in SSH on research data management with similar programs

---

1 IDTP https://euraxess.ec.europa.eu/belgium/jobs-funding/doctoral-training-principles
from eight French universities and eight European and American universities, revealing a large diversity of objectives, content, schedules and formats, some of them tailored for the specific needs of research laboratories and/or scientific communities (Schöpfel et al. 2018). Other surveys have been conducted on academic education and on-job training of academic librarians, with a focus on open access publishing and comparing French and German programs, revealing, exposing not only a dynamic situation but also a lack of conceptualization, coordination and recognized curricula, especially regarding minima and the relation between initial and continuing professional education; also, there is no clear distinction between formal training and other learning situations, e.g. conferences, meetings etc. (Bargheer et al. 2014, Jacquemin et al. 2015, Osswald et al. 2016).

Some preliminary remarks on the French context of PhD education. Each PhD student is registered and affiliated with one disciplinary graduate school (in French: école doctorale) which may be accredited by one or more universities. A Decree of May 25, 2016, laying down the national framework for the delivery of a French doctoral diploma, assigns the responsibility to graduate schools to offer PhD students training that "promotes interdisciplinarity and the acquisition of a broad scientific culture, including knowledge of the international research environment (and to) ensure that each doctoral student receives training in research ethics and scientific integrity". One part of the disciplinary training is organized by research laboratories, such as conference series and seminars. Other training opportunities, especially on cross-disciplinary (transversal) topics like project management, career planning, documentation, ethics and job skills training, are coordinated and organized by campus-wide structures (in French: collèges doctorales), academic libraries and regional scientific and technical information training units (URFIST). The graduate schools define the criteria and validate the individual PhD education; for instance, at Bordeaux, PhD students must validate at least 100 hours, while at Grenoble, they need 120 hours, in three areas (disciplinary, cross-disciplinary, job skills). One part of this training can be done off-campus, with other French institutions or abroad.

Methodology

The survey was conducted with a sample of French universities. Assuming that the leading research universities are the best terrain for the assessment of new and innovative initiatives in doctoral education, our approach lays emphasis on excellence, limiting the survey to a sample of ten leading French universities which are member of the League of European Research Universities (LERU), participant of the French IDEX program, and/or among the top 100 World Universities of the ARWU Ranking ("Shanghai"): Aix-Marseille, Bordeaux 1, Côte d'Azur, Grenoble Alpes, Lyon, Paris Saclay, Paris Sciences Lettres, Sorbonne Universités, Sorbonne Paris Cité, and Strasbourg. They represent about 12% of the French universities and 662,000 students, i.e. more than 20% of the students in French Higher Education (HE)

2 Text on Legifrance https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000032587086
Institutions.

For each university, we assessed the PhD training programs of the graduate schools, research laboratories and campus–wide structures in order to identify open science related contents (via content analysis of titles and abstracts), following the LERU dimensions (open access, research data sharing, ethics and integrity, citizen science). This assessment has been done online, in August 2019, based on the programs and training opportunities published on the web pages of the different structures. The results have been analysed following a short–list of criteria (see annex).

Results

The total number of PhD students in the sample of ten universities is about 34,000 PhD students, registered in 138 graduate schools and affiliated to one of more than 1,500 research laboratories (see figure 1).

<table>
<thead>
<tr>
<th>University</th>
<th>Short name</th>
<th>Students</th>
<th>PhD</th>
<th>Labs</th>
<th>Grad Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aix-Marseille</td>
<td>AMU</td>
<td>63 000</td>
<td>2 927</td>
<td>110</td>
<td>12</td>
</tr>
<tr>
<td>Bordeaux</td>
<td>Bdx</td>
<td>56 000</td>
<td>2 000</td>
<td>88</td>
<td>8</td>
</tr>
<tr>
<td>Côte d’Azur</td>
<td>UCA</td>
<td>24 000</td>
<td>1 700</td>
<td>54</td>
<td>3</td>
</tr>
<tr>
<td>Grenoble Alpes</td>
<td>UGA</td>
<td>82 000</td>
<td>3 500</td>
<td>120</td>
<td>14</td>
</tr>
<tr>
<td>Lyon</td>
<td>Lyon</td>
<td>120 000</td>
<td>5 400</td>
<td>172</td>
<td>17</td>
</tr>
<tr>
<td>Paris Saclay</td>
<td>Saclay</td>
<td>65 000</td>
<td>5 458</td>
<td>300</td>
<td>17</td>
</tr>
<tr>
<td>Paris Sciences Lettres</td>
<td>PSL</td>
<td>17 000</td>
<td>2 310</td>
<td>181</td>
<td>6</td>
</tr>
<tr>
<td>Sorbonne Paris Cité</td>
<td>USPC</td>
<td>120 000</td>
<td>6 000</td>
<td>251</td>
<td>32</td>
</tr>
<tr>
<td>Sorbonne Universités</td>
<td>SUPER</td>
<td>65 000</td>
<td>2 632</td>
<td>201</td>
<td>19</td>
</tr>
<tr>
<td>Strasbourg</td>
<td>Unistra</td>
<td>50 000</td>
<td>2 242</td>
<td>72</td>
<td>10</td>
</tr>
<tr>
<td><strong>total</strong></td>
<td><strong>662 000</strong></td>
<td><strong>34 169</strong></td>
<td><strong>1 549</strong></td>
<td><strong>138</strong></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Global information about the survey sample

Except for two (UCA and PSL), the surveyed universities are large HE institutions with more than 50,000 students. Three universities (Lyon, Saclay and USPC) represent nearly half of the PhD students of the sample. Based on the PhD/students ratio and the number of research laboratories, five universities appear more research–centred than the others (Lyon, Saclay, PSL, USPC and SUPER).

Number of training events

Through Internet search, we found 922 doctoral training events offered by the ten universities to their PhD students, ranging from 45 to 314 events per institution, with a median of 68.

From these events, 65 could be clearly categorized as related to open science (7%). The
differences between the institutions are significant, the percentage of open science (OS) contents varying between 0% and nearly 30%, with five universities offering about 10% or more OS–related activities. In absolute figures, this means that we identified between 0 and 13 OS–related training opportunities per university (see figure 2).

![Figure 2. Number of open science related training events (N=65)](image)

The median number of this category of doctoral training events is between 5 and 6.

39 opportunities could be easily identified as OS related because the relevant topics were explicitly described in the title, e.g. “research integrity in scientific professions” or “open access publishing of research results: gold road and green road” (60%). The other 26 events (40%) – trainings courses, seminars, conferences – were identified through the analysis of the content of the abstracts, lists of sessions etc., e.g. the mention of expertise of the open access publishing landscape as part of a 20h course on “scientific and technical information retrieval”.

**Topics**

The 65 training opportunities were categorized following the LERU and EC description of essential topics (see above), adding a general “open science” category. The most relevant content is open access publishing, followed by ethics and integrity (see figure 3).
Training courses, workshops or other events on open access publishing represent half of all training opportunities (33). One part of these actions puts the focus on the OA deposit and dissemination of the PhD dissertations, via institutional repositories or the French national TEL repository. Other events inform more generally about the OA landscape and/or teach how to publish scientific results in OA, via open repositories like HAL or "gold" journals.

Education and training on ethical challenges and scientific integrity represent the second category, with 19 identified events (29%). Both aspects, ethics and integrity, are often linked and offered together, as part of the same training.

The two other OS relevant topics appear less important. We identified only seven training events on research data management, including data sharing (11%) and only one training on collaborative citizen science.

Five other workshops or training courses provide global information on the new paradigm of open science, covering some or all of the relevant topics mentioned above (8%).

**Objectives**

Whenever possible, we tried to distinguish explicit, stated objectives of training events in terms of skills and knowledge (expertise). In a general way, it is obvious that all actions are designed to raise awareness and to provide knowledge on open science in general and on specific OS relevant topics, contributing to a better understanding and a higher expertise. Some key words of training goals in terms of awareness and knowledge:

- OA publishing: direct scientific communication, open repositories, predatory publishing, OA landscape, French OA infrastructures, OA publishing of dissertations, relevant intellectual property and author's rights, publisher
agreements, open licensing (Creative Commons), digital identity, author identifier (ORCID);
• Research data management: context, challenges, data repositories, data management plans, best practice;
• Ethics and integrity: ethical challenges of OA publishing and open science, research integrity, concepts, laws, ethical awareness of gender issues, post-publication peer review, artificial intelligence, robotics or personalized medicine;
• Citizen science: principles of collaborative projects with civil society;
• Open science: general knowledge and awareness, social and responsible innovation, participative management, open intellectual property.

All these training goals are described as “to learn more about”, “for a better understanding”, “to enhance awareness” etc.

Some training explicitly defines the didactic objectives in terms of acquisition of new skills, as part of the research activity and embedded in good scientific practice. The PhD students are expected to acquire operational and practical knowledge and to learn “how to do” open science. Some examples:

• OA publishing: publishing via the French HAL repository, deposit of dissertations, deposit on institutional repositories, information retrieval, scientific blogging, publishing in OA journals;
• Research data management: data organization, data protection, data documentation (metadata), data publishing, enhancing data reusability, writing a data management plan on the French DMP–OPiDoR platform;
• Ethics and integrity: participation and/or hosting a radio program, whistleblowing;
• Citizen science: ability to prepare a participative (collaborative) research project with partners from the civil society (citizens, associations, NGOs etc.);
• Open science: ability to conduct reproducible research.

Format and duration

The preferred format of the identified training events is a short 0.5–day face-to-face, in-class training course on campus, organized as a workshop, a conference (lecture) or a seminar of two to three hours, sometimes supported by educational resources and materials in the institutional intranet (Moodle…). One training on the deposit of dissertations lasts only one hour while some other programs require one day (= five or six hours).

Only four programs require more time and personal investment, i.e. three to six 0.5–or one–day in-class training units on the campus.

Some universities offer the same training program in French and English, especially on
OA publishing and research ethics; one university has developed a special English training program including a 3-days workshop on open access and open science for a small number of selected high-level international PhD students, with funding from the European H2020 framework program.

Other universities explicitly integrate the French public MOOC platform FUN³ into their PhD education catalogue, in particular because of the FUN online courses on research ethics and integrity developed by various universities, including Bordeaux and Lyon. The recommended MOOCs generally require two to three hours of personal investment per week, during five or six weeks.

Disciplines

For 55 training events, the target audience and disciplinary character could be clearly identified. Most of these events are cross-disciplinary and not limited to a scientific community or domain (see figure 4).

![Figure 4. Disciplinary character of training programs (N=55)](image)

All general introductions into open science issues and most of the training on ethics, integrity and research data management are multidisciplinary and not designed for a specific target group.

The situation is different for the training modules on open access publishing where nearly one third of the training on information retrieval, repositories etc. is tailored specifically for one discipline or a group of disciplines, especially in social sciences and humanities.

³ France Université Numérique [https://www.fun-mooc.fr/](https://www.fun-mooc.fr/)
Staff

Finally, as far as possible, we assessed the composition of the training team, in order to distinguish the professional background and profile of the trainers. Most training is prepared and conducted by academic librarians (78%) (see figure 5).

![Pie chart showing the professional background of the trainers (N=50)](image)

Figure 5. Professional background of the trainers (N=50)

Academic scholars and scientists are in charge of 11 training opportunities, representing a little more than 20% of the assessed offer; some of these events are organized together with librarians. Only in the field of research ethics and integrity do academic scholars appear to be more involved than librarians. On the other hand, we couldn’t identify any training on research data management conducted by (or with) scientists.

Discussion

Limitations of the survey

As mentioned above, the survey is a preliminary part of a larger international research project in the field of open science. This is the main reason for some methodological limitations of the presented results, i.e.:

- The collected data are not exhaustive. The data were collected at a given moment, during a short time period (August 2019). We wanted a representative "instant" photography, not an exhaustive catalogue of all PhD training opportunities organized during an academic year. This implies that the data are not suitable for a comparative assessment ("ranking") of the ten universities;
- The collected data are incomplete. The data collection was limited to online
available and accessible sources, e.g. web pages of graduate schools and campus–wide catalogues of PhD training. Missing abstracts, for instance, limited the possibility to identify training events where open science and related topics are not the main focus (= in the title) yet part of the content (= in the description).

For both reasons, the survey results probably underestimate the real number of relevant training opportunities, even if the overall number of assessed events (more than 900) appears sufficient for a representative description of the relative importance, the main topics and some essential characteristics.

The survey applied a mainly quantitative approach. It will be completed in the future, as part of a larger research project, by a qualitative approach, with semi-directive questionnaires and interviews on didactics, content and evaluation, with scientists, students, librarians and administrative staff from graduate schools, research laboratories, academic libraries and other campus–based structures.

**Good practice**

Following the preliminary empirical results, the literature overview and our own experience with PhD education (Schöpfel et al. 2018), we would recommend six key elements of good practice for the PhD training in open science:

1. **Topics**: the PhD training program should cover the whole range of the essential OS relevant topics, i.e. OA publishing of research results via repositories and journals, research data management and data sharing, research ethics and scientific integrity, and citizen science;
2. **Objectives**: the PhD training program should clearly distinguish between awareness raising, acquisition of knowledge (expertise) and skills development (including transferable skills training), and should offer learning opportunities for all levels and in all topics;
3. **Format**: the PhD training program in open science should be tailored for the specific working and learning conditions of PhD students. While general and cross-disciplinary awareness raising and knowledge acquisition can be organized via MOOCs and similar learning opportunities, accompanied by short face-to-face sessions for debate and feedback, skills development should be organized on-the-job, near-to-the-research-bench (i.e. laboratory);
4. **Staff**: the PhD training should be organized by mixed teams, with scholars, scientists, technical and administrative staff, and academic librarians;
5. **Networking**: the local PhD training should be open and connected to other education programs, from other universities, from research organizations or from other (national etc.) structures, with an explicit labelling and integration into the local PhD education;
6. Status: following the EC working group report, the PhD training on open science should be mandatory for universities, and one part of this training (not only knowledge about ethics and integrity, like in France) should be mandatory for the PhD students, e.g. awareness of OA publishing options, writing of data management plans, archiving and sharing of research data.

In addition, especially the highly ranked and research-intensive universities should apply at least some of the international Innovative Doctoral Training Principles (IDTP) within the framework of open science, considering that good training practice requires more than good staff, conditions and materials, in particular, for instance, attractive institutional environments including OS infrastructures and policies, international networking, and quality assurance.

Community

A large part of research practice is specific to disciplines and research fields, scientific communities and institutions, infrastructures, instruments and tools. Beyond some general assumptions and guiding principles, the application of open science should (must) be disciplinary and community-based. Therefore, the PhD training in open science should be more than general and cross-disciplinary (transversal) awareness raising. PhD students should learn about open science and acquire related skills in their usual environment, with members of their community, with their specific research projects and topics, tools, infrastructures, challenges, funding opportunities, scientific partners etc. If open science is to be taken seriously, they should learn (about) it as a bottom-up way of doing and thinking research, not as a top-down ideology.

The composition of the training staff was mentioned above. The required connection with the research community is another argument in favour of mixed teams and academic trainers.

Marketing

One major challenge of the survey was the (lack of) visibility of the PhD training in open science on the web. Some universities including Aix-Marseille, Grenoble, Bordeaux and Paris-Saclay, make their training offer available via the networked portal ADUM hosted by the University of Montpellier\(^4\). Others publish their own catalogues at least for the “transversal” education; in particular the disciplinary training is often available on the graduate schools’ and/or research laboratories’ websites.

However, the visibility issue is more than a simple methodological problem and should (also) be considered as a problem of marketing and communication. The LERU advices paper highlights that institutional policy development, governance and leadership at university level

\(^4\) Accès Doctorat Unique et Mutualisé [https://www.adum.fr](https://www.adum.fr)
PhD training is a major and significant part of academic life. It is a showcase of research and education excellence and it contributes to the preparation of the future of academic scholarship. Because of its strategic importance, PhD training should become one of the advocacy programs recommended by the LERU to “identify the benefits of Open Science approaches” on the campus. For this reason, graduate schools or other campus-wide structures should pay attention to the explicit labelling and description of open science relevant learning opportunities and contents, for instance via a specific category of “open science training” which could include other than the essential four topics, such as for instance project management or related legal issues (licensing, author’s rights...).

In other words, graduate schools and universities should (re)consider their catalogues, databases and web pages on PhD training as a tool for the communication and marketing of open science and ensure an easy and large visibility for this part of the PhD education.

Conclusion

The LERU states that “to embrace Open Science, universities and researchers need to embrace cultural change in the way they work, plan and operate” (Ayris et al. 2018). The development of open science requires a progressive cultural and technological modification of the ecosystem of research, including new incentives (peer review, reward system...), new infrastructures (data repositories, publishing platforms, research information systems...), new attitudes (sharing, reuse...), new behaviours and skills (publishing, describing, archiving...).

Training early career researchers and in particular, PhD students, is one major way to contribute to this change. The assessment of PhD training programs in highly ranked French universities provides meaningful insight into some features of graduate education. The objective was to establish empirical elements that may be helpful for the evaluation of other existing programs (benchmarking) and for the preparation and implementation of new programs.

However, in order to assess the impact and the outcome of PhD training on open science, more research is required, especially on the content, on didactics, evaluation, follow-up, incentives and feedback of seminaries, workshops, MOOCs and other learning opportunities. This will be the purpose of a future research project.

Acknowledgment

The study is preliminary work and part of a larger research project on the development of PhD dissertations in the new environment of open science (xDiss – Special Dissertations:
The Impact of Open Science on Doctoral Dissertations) for which funding has been requested from the French and German Research Agencies (2019 Franco–German call in humanities and social sciences, ANR/DFG). Project partners are the Universities of Oldenburg, Saarbrücken, Pau and Lille and the CNAM.

References


**Annex**

Five topics
- Open access publishing
- Research data management
- Citizen science
- Ethics/integrity
- Others (related to or mentioning open science)

Short-list of assessment criteria
- Number of programs with topic in title
- Number of programs with topic in content/abstract
- Objectives (in terms of knowledge and skills)
- Format and didactics (virtual, innovative features…)
- Duration (in days)
- Discipline (multidisciplinary, disciplinary)
- Training staff (scholars/scientists, librarians, others)