

## Training early career researchers

**Authors:** Paul Ayris - Pro-Vice-Provost (UCL Library Services), Co-Chair of the LERU INFO Community (League of European Research Universities) & Adviser to the LIBER Board (Association of European Research Libraries) & Ignasi Labastida (Head of the Research Unit at the Learning and Research Resources Centre [CRAI] of the University of Barcelona)

**Email:** [p.ayris@ucl.ac.uk](mailto:p.ayris@ucl.ac.uk) / [ilabastida@ub.edu](mailto:ilabastida@ub.edu)

### 18.1. RESEARCH AND RESEARCH-BASED EDUCATION

Research – blue sky and applied – is fundamental to the mission of research-intensive universities. As such, it is enunciated in the Mission Statements of such institutions. The University of Barcelona for example, a research-intensive university in the Catalan region, states that ‘The University of Barcelona is a public institution committed to the environment, whose mission is to provide a quality public service of higher education primarily through the study, teaching, research and effective management of knowledge transfer.’<sup>1</sup> This is its Mission. Research also features prominently in its Vision, ‘Barcelona University must be a university that offers comprehensive training, ongoing and critical evaluation at the highest level, and research which is both advanced and efficient.’<sup>2</sup>

The importance of research in a university is also captured in the Mission Statements of university organisations. LERU, the League of European Research Universities, advocates:<sup>3</sup>

- education through an awareness of the frontiers of human understanding;
- the creation of new knowledge through basic research, which is the ultimate source of innovation in society;
- and the promotion of research across a broad front in partnership with industry and society at large.

Learning through research and enquiry is a fundamental feature of study in a research-intensive university. Universities, with a strong tradition of producing world-class research, wish to demonstrate that excellence not only in their research outputs but also in the learning experience of their students, both undergraduate and postgraduate.

University College London (UCL), a research-intensive university in the UK, has developed a model for research-based education via its Connected Curriculum initiative.<sup>4</sup> This is made up of six inter-connected strands of activity:

1. Students connect with researchers and with the institution’s research;
2. A through line of research activity is built into each programme;

3. Students make connections across subjects and out to the world;
4. Students connect academic learning with workplace learning;
5. Students learn to produce outputs – assessments directed at an audience;
6. Students connect with each other, across phases and with alumni.

Research data can also be seen as a learning object. In a digital environment, research outputs cannot be restricted/limited to traditional written works such as journal articles or monographs. Nowadays, research outputs consist of a mixture of objects, amongst which can be found written works and data. One of the building blocks for these publications is research data. Via digital networks, it is possible to share both publications and the underlying data to anyone who can access them. The emergence of research data as a major source of information is now becoming apparent. To take advantage of this revolution researchers, especially early career researchers, need to be trained in best practice in research data management. This Case Study offer one example of how this can be done.

### 18.2. EARLY CAREER RESEARCHERS

In 2001, a US study sponsored by the Pew Charitable Trusts found that:<sup>5</sup>

Students in 11 arts and sciences disciplines from 27 institutions and 1 cross-institutional program [...] were surveyed. Responses were received from 4,114 students, a response rate of 42.3%. Results suggest that the training doctoral students receive is not what they want, nor does it prepare them for the jobs they take. Many students do not understand what doctoral study entails, how the process works, and how to navigate it effectively. There is a mismatch among the purpose of doctoral education, the aspirations of the students, and the realities of their careers within and outside academia.

In 2017, the situation is better. The UCL Doctoral School, in its Code of Practice, stresses:<sup>6</sup>

UCL offers a programme for the development of generic research and personal transferable skills to help you develop the skills necessary not only for successful completion of your degree but also to equip you for later life and for the workplace ... The specific menu of courses and other training opportunities should be discussed between you and your Supervisors using the skills self assessment section of UCL’s Research Student Log. The self-assessment process is based on a national framework, the Researcher Development Framework.

It follows that the need for skills development has been identified and courses/materials put in place. One of those training needs concerns research data management.

<sup>1</sup> ‘La Universitat de Barcelona és una institució de dret públic compromesa amb l’entorn, la missió de la qual és prestar el servei públic (de qualitat) de l’ensenyament superior principalment per mitjà de l’estudi, la docència, la recerca i una gestió eficaç de la transferència del coneixement.’ University of Barcelona: <http://www.ub.edu/pladirector/en/missio.html>; last accessed 29/1/17.

<sup>2</sup> ‘La Universitat de Barcelona ha de ser una universitat que inclogui una formació integral, continuada i crítica del més alt nivell, i una recerca avançada i eficient.’ University of Barcelona: <http://www.ub.edu/pladirector/en/missio.html>; last accessed 29/1/17.

<sup>3</sup> LERU: <http://www.leru.org/index.php/public/about-leru/mission/>; last accessed 7/1/17.

<sup>4</sup> UCL: <https://www.ucl.ac.uk/teaching-learning/education-initiatives/connected-curriculum>; last accessed 7/1/17.

<sup>5</sup> Golde, Chris M., Dore, Timothy M., *At Cross Purposes: What the Experiences of Today’s Doctoral Students Reveal about Doctoral Education*, Wisconsin University, Madison, 2001; available at <http://files.eric.ed.gov/fulltext/ED450628.pdf>; last accessed 7/1/17.

<sup>6</sup> UCL: <http://www.grad.ucl.ac.uk/codes/Graduate-Research-Degrees-Code-of-Practice-1617.pdf>; last accessed 7/1/17.

### 18.3. THE LERU DOCTORAL SUMMER SCHOOL AS A MODEL OF BEST PRACTICE

LERU itself has produced a LERU Roadmap for Research Data.<sup>7</sup> Chapter 6 looks at Roles, Responsibilities and Skills and identifies a need for training for early career researchers, for academics and for support staff. The training needs and routes for skills development are clearly identified in Figure 18.1 below. The separate categories are not mutually exclusive. All stakeholders – student/PhD + Senior Researcher + Librarian + Data Scientist need to work together to share knowledge and Best Practice. Nonetheless, the categorisation in Figure 18.1 does attempt to codify the learning needs of each stakeholder group and how these needs can realistically be met. It accepts that there is a graduated series of learning needs, starting with postgraduate/PhD students, and which increase in complexity as early career researchers become Senior Researchers. In this model, Librarians have a new role to play in the research space. They need to acquire new skills and to impart that knowledge to the groups that they train. This partnership is crucial in embedding RDM skills into the research landscape. Finally, there is the emerging new career of Data Scientist, and this is discussed more fully in the section below on the European Open Science Cloud.

WHO	Postgrad/PhD	Senior Researcher	Librarian	Data Scientist
WHEN	Early stages of postgraduate study	As needed, or at beginning of research project/proposal state	CPD for subject librarians/during library education	Discipline-specific academic courses (doctoral)/CPD
WHAT	Basics of data management practice, FAIR <sup>8</sup> principles, data citation, data evaluation. Competence in legal and ethical issues	Training on discipline-specific data management practices; an understanding of the FAIR principles; how to write a data management plan (tailored as necessary to funder requirements), data reuse skills. Competence in legal and ethical issues	Data curation. An understanding of the FAIR principles. Some disciplinary-specific e-research methods (TDM)/data collection skills, IT skills. Competence in legal and ethical issues	Discipline-specific skills for data management/exploitation/interoperability. An understanding of the FAIR principles. Competence in legal and ethical issues
HOW	Credited models	Practical training	Accredited CPD/ Professional courses	Professional (academic) courses and accredited CPD

Figure 18.1: Training needs and routes for skills development

Having identified the training needs, how can those needs be met? The LERU Roadmap suggests that, for most categories of user, what is required are credited models and/or professional courses. LERU universities have taken this to the next stage by devising a format for a formal Summer School to train PhD students new to research and to RDM. The first meeting was held in Leiden in Summer 2016.<sup>9</sup> This is a taster for future activity, which is currently being discussed in the LERU network.

The Programme for the Summer School<sup>10</sup> had as its ambition the creation of the ‘new generation of data scientists’. Each of the 21 LERU member universities<sup>11</sup> was invited to send one or more members of their doctoral programme to attend the week, the intention being that having received training in Leiden they

<sup>7</sup> LERU: [http://www.leru.org/files/publications/AP14\\_LERU\\_Roadmap\\_for\\_Research\\_data\\_final.pdf](http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf); last accessed 7/1/17.

<sup>8</sup> Force11: <https://www.force11.org/group/fairgroup/fairprinciples>; last accessed 29/1/17.

<sup>9</sup> LERU Doctoral Summer School (2016): <http://www.dtls.nl/fair-data/fair-data-training/leru-summer-school/>; last accessed 7/1/17; and for general background LERU Doctoral Summer School (2016): <http://www.dtls.nl/first-kind-leru-doctoral-summer-school-data-stewardship/>; last accessed 7/1/17.

<sup>10</sup> LERU Doctoral Summer School (2016): <http://www.dtls.nl/fair-data/fair-data-training/leru-summer-school/programme/>; last accessed 29/1/17.

<sup>11</sup> LERU: <http://www.leru.org>; last accessed 29/1/17.

could then return to their organisations and cascade that knowledge around local doctoral candidates. 38 doctoral students attended the event from 21 universities and associated hospitals.<sup>12</sup>

The format of the Programme was to aim for a mixture of keynote speakers on specific topics, speakers to lead in particular thematic areas and student presentations/discussions. The Summer School highlighted a number of issues, which are likely to form the core of RDM training activity going forward. Some of the more prominent are listed here:

- The importance of research data being FAIR (Findable, Accessible, Interoperable and Reusable)<sup>13</sup>
- The importance of data management plans in providing a framework for the creation, storage, and sharing of research data<sup>14</sup>
- Licensing issues and an explanation of the meaning of the Creative Commons suite of licences and its use in research data<sup>15</sup>
- Big Science is Open Science<sup>16</sup>
- The future infrastructure for Open Science<sup>17</sup>

### 18.4. TOP-LEVEL ISSUES CONCERNING RESEARCH DATA FOR THE LERU SUMMER SCHOOL

FAIR data is one of the building blocks of the new information age. If research data is findable, accessible, interoperable and reusable, it increases in value as a tool for supporting innovation and new discoveries. Effective licensing of research data, when needed, increases their usefulness and makes it clear what the terms of re-use are. One of the drawbacks of the early development of Open Access is that many of the published research outputs tagged as Open Access outputs have no accompanying licence. This makes it difficult to understand exactly what the rules for reuse are in every case. Moreover, the lack of a licence has to be interpreted as all rights reserved in accordance with copyright law. Can an Open Access publication, with no accompanying licence, be re-used for commercial advantage?

Not all research data is big data. Many collections of data form part of a long tail of data creation, where research data has been created/collected to support the publication of a particular article, or a lecture to taught-course students. The term ‘big data’ is sometimes overused and brings with it legal issues such as privacy into the discussion. Nonetheless, the best future for research data, whether big or small, is that it is open where that is legally possible. Finally, to deliver and perform Open Science, infrastructure is needed – not simply technical platforms but also training and skills development programmes to create the ‘new generation of data scientists’. All this and more was discussed in a focussed and intensive week in the LERU Doctoral Summer School.

A particularly important part of the model for the Summer School was the balance between formal presentations and the opportunity for students themselves to present case studies using their own research data, and to interact with speakers.<sup>18</sup> ‘With data sharing, new scientific discoveries can be made’ was one

<sup>12</sup> LERU Doctoral Summer School (2016): <http://www.dtls.nl/fair-data/fair-data-training/leru-summer-school/students/>; last accessed 29/1/17.

<sup>13</sup> FAIR: <https://www.force11.org/group/fairgroup/fairprinciples>; last accessed 7/1/17

<sup>14</sup> Digital Curation Centre: <http://www.dcc.ac.uk/resources/data-management-plans>; last accessed 7/1/17.

<sup>15</sup> Creative Commons: <https://creativecommons.org/licenses/>; last accessed 7/1/17.

<sup>16</sup> For a description of Big Data in the science ecosystem, see Royal Society, *Science as an Open Enterprise* (London: Royal Society, 2012), available at <https://royalsociety.org/~media/policy/projects/sape/2012-06-20-saoe.pdf>; last accessed 7/1/17.

<sup>17</sup> European Commission: <http://ec.europa.eu/research/openscience/index.cfm>; last accessed 7/1/17.

<sup>18</sup> Some of this is captured on the summary video at LERU: <https://www.youtube.com/watch?v=sSQPY5Mc5Rs>; last accessed 7/1/17; and in the tweets recorded at LERU Summer School (2016): <https://twitter.com/hashtag/lerusummerschool2016?src=hash>; last accessed 7/1/17.

of the recorded tweets. Other participants felt that the Summer School was a valuable mirror to reflect how science is done in the twenty-first century. The participants expressed real enjoyment at being able to participate in the event. In fact, they wished they had had more time to discuss the new information that they were learning each day. With feedback like this, the objective of the Summer School to provide solid training in data stewardship for the next generation of future leaders does not seem to have been unrealistic.

### 18.5. THE FUTURE PATTERN OF SKILLS DEVELOPMENT: THE EUROPEAN OPEN SCIENCE CLOUD?

In July 2016, the European Commission published the Report of its High Level Expert Group on the European Open Science Cloud:<sup>19</sup>

... The European Open Science Cloud (EOSC) aims to accelerate and support the current transition to more effective Open Science and Open Innovation in the Digital Single Market. It should enable trusted access to services, systems and the re-use of shared scientific data across disciplinary, social and geographical borders. The term cloud is understood by the High level Expert Group (HLEG) as a metaphor to help convey both seamlessness and the idea of a commons based on scientific data. This report approaches the EOSC as a federated environment for scientific data sharing and re-use, based on existing and emerging elements in the Member States, with lightweight international guidance and governance and a large degree of freedom regarding practical implementation. The EOSC is indeed a European infrastructure, but it should be globally interoperable and accessible. It includes the required human expertise, resources, standards, best practices as well as the underpinning technical infrastructures. An important aspect of the EOSC is systematic and professional data management and long-term stewardship of scientific data assets and services in Europe and globally. However, data stewardship is not a goal in itself and the final realm of the EOSC is the frontier of science and innovation in Europe.

Important in this summary of activity, for present purposes, is the recognition of the importance of skills development. The LERU Roadmap itself identified the category of Data Scientist as the summation of skills development in terms of research stakeholders. This identification is further developed in the EOSC Report by emphasising the absolute importance of developing the role of data stewards to deliver the vision of a global commons of scientific data. The Report suggests:

A first cohort of core data experts should be trained immediately to translate the needs for data driven science into technical specifications to be discussed with hard-core data scientists and engineers. This new class of core data experts will also help translate back to the hardcore scientists the technical opportunities and limitations.<sup>20</sup>

Elsewhere, the Report puts figures to the training requirement:

The number of people with these skills needed to effectively operate the EOSC is, we estimate, likely exceeding half a million within a decade. As we further argue below, we believe that the implementation of the EOSC needs to include instruments to help train, retain and recognise

this expertise, in order to support the 1.7 million scientists and over 70 million people working in innovation. The success of the EOSC depends upon it.<sup>21</sup> These are significant numbers. It will take a significant investment in European teaching infrastructures to develop the curricula, agree success criteria for measuring successful delivery and finance this huge training undertaking. Commissioner Moedas (Research, Science & Innovation), however, has highlighted the need for skills development and has said, 'Such recommendations deserve detailed consideration by the scientific community and other stakeholders.'<sup>22</sup>

Research performing organisations need to start somewhere, as the LERU Roadmap makes clear. In this context the model of the LERU Doctoral Summer School seems a measured, successful and immediate response that such bodies need to make to manage the training needs implicit in the data deluge.

### 18.6. CONCLUSION

The purpose of this Case Study has been to look at the role of research performing organisations in skills development for early career researchers, and to set those needs in the context of the growing importance of research data and the emerging role of the data steward. In 2001, a North American survey decried the value and use of much doctoral training activity. In 2017, examples from Western Europe show that levels of provision, and the understanding of the need, have improved. The LERU Roadmap for Research Data (2013)<sup>23</sup> identified clear needs for research data training and the LERU Doctoral Summer School in Leiden (2016) provides a Best Example Case Study of how that training can be delivered in practice. The scale of the need, however, is illustrated by the Report of the High Level Expert Group on the European Open Science Cloud – half a million skilled data stewards needed in the next 10 years. If true, there is no room for complacency and every need for action. The LERU Summer School provides an excellent model for the training seminars needed to deliver generic skills and subject-specific insights into the emerging activity of data stewardship.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

<sup>19</sup> European Commission: <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>, p. 6; last accessed 7/1/17.

<sup>20</sup> Ibid., p. 16.

<sup>21</sup> Ibid., p. 12.

<sup>22</sup> Ibid., p. 4.

<sup>23</sup> LERU: [http://www.leru.org/files/publications/AP14\\_LERU\\_Roadmap\\_for\\_Research\\_data\\_final.pdf](http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf); last accessed 7/1/17.