The International Journal of Digital Curation
Volume 7, Issue 2 | 2012

The Data Management Skills Support Initiative: Synthesising Postgraduate Training in Research Data Management

Laura Molloy and Kellie Snow,

HATII,

University of Glasgow

Abstract

This paper will describe the efforts and findings of the JISC Data Management Skills Support Initiative ('DaMSSI'). DaMSSI was co-funded by the JISC Managing Research Data programme and the Research Information Network (RIN), in partnership with the Digital Curation Centre, to review, synthesise and augment the training offerings of the JISC Research Data Management Training Materials ('RDMTrain') projects.

DaMSSI tested the effectiveness of the Society of College, National and University Libraries' *Seven Pillars of Information Literacy* model (SCONUL, [2011]), and Vitae's *Researcher Development Framework* ('Vitae RDF')[1] for consistently describing research data management ('RDM') skills and skills development paths in UK HEI postgraduate courses.

With the collaboration of the RDMTrain projects, we mapped individual course modules to these two models and identified basic generic data management skills alongside discipline-specific requirements. A synthesis of the training outputs of the projects was then carried out, which further investigated the generic versus discipline-specific considerations and other successful approaches to training that had been identified as a result of the projects' work. In addition we produced a series of career profiles to help illustrate the fact that data management is an essential component – in obvious and not-so-obvious ways – of a wide range of professions.

We found that both models had potential for consistently and coherently describing data management skills training and embedding this within broader institutional postgraduate curricula. However, we feel that additional discipline-specific references to data management skills could also be beneficial for effective use of these models. Our synthesis work identified that the majority of core skills were generic across disciplines at the postgraduate level, with the discipline-specific approach showing its value in engaging the audience and providing context for the generic principles.

Findings were fed back to SCONUL and Vitae to help in the refinement of their respective models, and we are working with a number of other projects, such as the DCC and the EC-funded Digital Curator Vocational Education Europe (DigCurV[2]) initiative, to investigate ways to take forward the training profiling work we have begun.

---

[1] Vitae Researcher Development Framework: www.vitae.ac.uk/rdf

[2] DigCurV: http://www.digcur-education.org/

Background to DaMSSI

In recent years, significant effort has gone into defining data management roles and responsibilities for those involved in the production of digital research data. The National Science Foundation's 2005 report *Long-lived digital data collections: enabling research and education in the 21st century* (NSF, 2005) suggested a number of responsibilities that data authors should recognise, but despite these recommendations being around for some time, there is still little evidence that data management skills are being embedded within UK postgraduate courses. Feedback from attendees at events such as the JISC Innovation Forum 2008 data management skills and capacity session (Davidson & Jacobs, 2008), indicates that while some UK university departments are delivering training to their postgraduates, much more needs to be done to embed data management training into all postgraduate programmes; a sound understanding of research data management is an important enabler to postgraduate participation in the open data debate and informs awareness of requirements to ensure high-quality data. There is also evidence that researchers in UK HEIs are likely to respond favourably to data management support with a focus relevant to their discipline (Ward et al., 2010).

The identification of basic data management skills for various roles has been further investigated by a number of ongoing working groups, such as DigCCurrII[3] and the International Digital curation Education Action (IDEA) Working Group (Hank & Davidson, 2009). Some progress has been made by these groups, but the range of skills identified varies widely – ranging from very technical aspects to more traditional library and information sciences skills. To enable UK HEIs to effectively embed data management training into their postgraduate courses, agreement is needed on what constitutes a basic set of postgraduate data management skills; a means of consistently describing these skills; and a framework that supports the progression of skills development over time.

DaMSSI worked with an understanding of research data as the information and materials which form 'the evidence necessary to produce and evaluate research results, and to reconstruct the events and processes leading to them'[4]. We are using the term 'research data management' to incorporate the concepts of digital preservation and curation of such data produced both during research processes in higher education institutions and in the course of graduates' professional practice.

DaMSSI aimed to facilitate the use of two higher education skills development frameworks to help researchers and their institutions to effectively plan data management skills development and training. The first of these is Vitae's Researcher Development Framework and its supporting Researcher Development Statement (RDS). The latter expanded and replaced the 2001 Joint Statement of the UK Research Councils' Training Requirements for Research Students (the 'Joint Skills Statement' or JSS[5]), which outlined the skills and attributes Council-funded doctoral researchers

---

[3] DigCCurr II: http://www.ils.unc.edu/digccurr/aboutII.html

[4] A statement based on a RIN definition offered in their valuable report, *Stewardship of Digital Research Data: Principles and Guidelines* (2007) available at http://www.rin.ac.uk/our-work/data-management-and-curation/stewardship-digital-research-data-principles-and-guidelines.

[5] Joint Skills Statement: http://www.vitae.ac.uk/policy-practice/1690/Joint-Skills-Statement.html

were expected to develop during their studies. The Vitae RDF and RDS are endorsed by over thirty organisations, including Research Councils UK and the Russell Group universities, as well as library and information groups, such as Research Libraries UK and CILIP[6]. The Vitae RDF is intended to be used as:

> "a tool for planning, promoting and supporting the personal, professional and career development of researchers. It describes knowledge, skills, behaviours and personal qualities acquired by researchers and encourages researchers to aspire to excellence through development to a high level." (RIN, 2009)

The Vitae RDF offers great potential for describing the basic data management skills required at each stage of a researchers' career and for securing agreement on basic skill sets. However, while data management skills are implied throughout several sections of the Vitae RDF, they are not explicitly defined.

The second model examined by DaMSSI was the SCONUL Seven Pillars Model (SCONUL, 2011). Originally developed in 1999 and adopted by HE librarians and teachers around the world, the model helps to define a pathway from basic library and IT skills through to complete information literacy, and describes progressive stages ranging from the novice to the expert. While the model has proved valuable as a planning tool, it may benefit from additional data management detail.

DaMSSI sought to investigate whether these two models – both widely accepted within their own spheres – could be effectively used by institutions to describe and embed postgraduate data management training. To that end, DaMSSI staff mapped the skills addressed by the training outputs of each of the RDMTrain projects to both of the models. This illustrated where the training outputs were confluent and where their areas of coverage diverged. This also provided a starting point for the projects to describe the results of their work in a way that would interoperate with training providers at each of their institutions. The timing of the Initiative's work coincided with a revision and refinement period for both models, with the Seven Pillars being adapted as a research 'lens' to outline information literacy specifically for HE researchers, and the RDF being analysed by the RIN Information-Handling Working Group[7] in order to feed back recommendations for more explicit data management skills, and to assist in the development of an information literacy 'lens' for the Framework. DaMSSI used this opportunity to communicate its findings from the mapping work back to SCONUL and Vitae, to inform the development of each of the models alongside the work of the RIN Information-Handling Working Group.

DaMSSI therefore had several overriding objectives:

- Support the RDMTrain projects in engaging with the Seven Pillars and Vitae RDF, giving hands-on support in mapping their training materials to these models;

---

[6] A full list of organisations who have endorsed the RDS is available at:
http://www.vitae.ac.uk/researchers/278641/RDS-endorsements.html

[7] Information about the working group is available at: http://www.rin.ac.uk/mind-skills-gap

- Develop recommendations for Vitae and SCONUL on how models could be refined to include data management skills, particularly the development of an information-handling 'lens' for the Vitae RDF;

- Produce a comparison and synthesis of the training approaches adopted by the projects, reflecting similarities and disciplinary differences;

- Produce a series of career profiles highlighting how data management is relevant to a wide range of careers both within and beyond HE;

- Establish contact with professional bodies and LIS course providers, advising them of the significance of data management skills and how they might use the models to support researchers and professionals, as well as to develop their own knowledge and skills sets.

Mapping Data Management Training to the Models

To improve the provision of research data management practice at postgraduate level, JISC funded the five projects of the RDMTrain strand, with the aim of creating a body of discipline-focussed postgraduate training units that could be reused by other institutions to stimulate curriculum change and create a greater awareness of the need for research data management skills training. DaMSSI worked with and supported the RDMTrain projects in a reciprocal relationship.

With the cooperation of the projects, DaMSSI tested the effectiveness of the Seven Pillars of Information Literacy model, and Vitae's Researcher Development Framework for consistently describing data management skills and skills development paths in UK HEI postgraduate courses.

The Initiative began by conducting initial site visits to each RDMTrain project and mapping individual course modules to the two frameworks. Following feedback on these early mappings from the projects, DaMSSI then produced a larger mapping of all the projects' modules to the RIN Information-handling Working Group's combined RDF/Seven Pillars framework (Goldstein, 2010), from which basic generic data management skills were identified alongside discipline-specific requirements. Along with highlighting issues about the value of the Vitae RDF and Seven Pillars models themselves, the mapping of the projects' course outputs to the models suggested that there was consistency in the data management skills required across the disciplines, despite variety in the arrangement of course modules among the projects. Discipline-specific variations through examples and case studies constituted the main ways courses were further customised.

The mapping and feedback from the projects allowed DaMSSI to identify the extent to which the two models appeared useful for describing and supporting data management training. Overall, we found that the models were potentially useful for consistently describing learning outcomes of courses, as well as for embedding courses within institutions' existing research methods courses. At the same time there was a need for a clearer focus on information-handling and data management within the Vitae RDF if it was to be used to support training, for which the Seven Pillars could prove useful if it was revised to reflect the wider data management lifecycle. The terminology used for both models also required further refinement to more

accurately describe data management skills in a way that made sense to research audiences.

These findings were fed back to SCONUL and Vitae to help in the further refinement of their models. SCONUL reported that the feedback was useful and included changes in their revisions as a result, including a bigger emphasis on data collection alongside published information gathering. Overall, Vitae also welcomed the feedback provided, and their ongoing work on a new information literacy lens for the Vitae RDF (assisted by the RIN Information-handling Group) should address many of the issues identified by DaMSSI. DaMSSI used its findings to contribute to this lens.

Synthesis of Research Data Management Training Materials

Once the mappings were completed, DaMSSI then embarked on subsequent synthesis work to explore the findings from the mappings further. The Initiative collated and compared final course materials, feedback from students on courses delivered, and each project's own findings, conclusions and recommendations to see if there was agreement with what the mapping had suggested. Overall, the synthesis work confirmed the conclusions drawn from the mappings. The following list is a summary of the findings from the exercise, with regards to generic and discipline-specific considerations, course delivery approach and wider skills development:

- Generic RDM principles applied across all disciplines, but discipline-specific definitions, examples and exercises were seen as beneficial by the projects and course participants;

- Data management plans (DMPs) required discipline-specific interpretation to be understood by students, and projects which placed a focus on developing individual plans as part of skills development received good feedback;

- Many researchers don't understand the specialist language associated with RDM and information handling;

- Courses that successfully balanced the need for discipline-specific detail with relatively brief and concise training showed better delegate retention;

- Feedback suggested that training should come at an early stage in studies and should ideally be integrated into existing institutional research methods courses;

- Course attendees liked the face-to-face delivery method, as they were able to ask questions and share experiences with others;

- Without understanding the potential benefits of RDM, students are still reluctant to put real effort into data management unless required to do so, and low attendance at some courses suggested a lack of understanding of the importance of the topic.

The findings were then refined into a set of project recommendations.

Data Management Skills for Professionals

Alongside the mapping and synthesis work, DaMSSI supported the production of a number of guidance documents to raise awareness of the importance of data management beyond the postgraduate and academic environment. The Initiative produced a series of career profiles to help illustrate the fact that data management is an essential component – in obvious and not-so-obvious ways – of a wide range of professions. Accordingly, students in a wide range of disciplines need to acquire and hone their data management skills. The profiles of a conservator, social science researcher, archaeologist, clinical psychologist and data manager were produced to link in with the disciplines covered by the RDMTrain projects. Each career profile provides a description of the profession, an outline of key roles and responsibilities, an explanation of how RDM skills fit into the day-to-day responsibilities of the profession, and a list of further reading, relevant professional bodies and standards. The profiles were inspired by the RIN's previous biocurator career profile (RIN, 2010), and are openly available from the DCC[8] and RIN[9] webpages for DaMSSI.

When considering the suite of five profiles, the Initiative wanted to ensure that these could be re-used by higher education institutions to help prospective students understand the possible range of activities that they might be expected to undertake upon entering their chosen profession – both with regards to data management and more broadly. The profiles have the added purpose of helping to highlight the potential role of professional bodies in promoting and supporting data management skills development amongst professionals in their fields. Professional bodies will be encouraged to make use of the profiles to promote and endorse the value of acquiring data management skills to their members and also as a means of promoting their own training and dissemination activities.

The mapping activities of DaMSSI ran in tandem with the work of the RIN Information-Handling Working Group on improving the information-handling focus of the RDF. A taxonomy of information literacy for the RDF was drawn up by the group (RIN, 2011), and a guide entitled 'The Informed Researcher'[10], to be published in October 2011, was commissioned to complement the RDF information-handling lens. DaMSSI used its findings to contribute to this guidance, comparing the taxonomy to the combined mapping and noting areas for additions from a data management perspective.

As the RDMTrain data management training was mapped to the Seven Pillars, the Initiative was also able to identify a number of skills which may be valuable to information professionals with regards to data management, particularly those involved in academic institutions and supporting students and researchers. The skills areas are:

---

[8] DCC skills frameworks: http://www.dcc.ac.uk/training/data-management-courses-and-training/skills-frameworks

[9] RIN data management, information literacy and DaMSSI: http://www.rin.ac.uk/data-management-skills

[10] This guidance will form part of Vitae's series of researcher support booklets; it will be issued jointly by Vitae, RIN and SCONUL.

- Knowledge of what constitutes research data across a variety of disciplines;

- Understanding of data created by the researcher and how this differs to published data in terms of gathering and evaluating;

- Understanding of how to organise research data, including structuring file names, organising folders, and adding metadata;

- Knowledge of data sharing options, including which licence is right for the researcher's needs (Ball, 2011);

- Awareness of data preservation and curation options.

Most of these new skills derive from those traditionally being taught to information specialists. With minor updating of course content, new graduates of information science programmes could be very well-placed placed to provide much-needed data management and curation advice, support and training within potential employers' organisations. The recommendations are now being communicated to CILIP and other relevant LIS contacts.

Recommendations and Future Development

Based on the findings from the mapping and synthesis work, DaMSSI has drawn together the following list of recommendations for future providers of data management training:

- Work closely with disciplinary experts to ensure that terminology used within courses is accurate and clear, including agreeing a basic definition of core concepts such as what 'data' can be within the discipline;

- Keep overviews and central descriptions of topic areas basic and generic, introducing the topic at a digestible level and allowing for easier integration into existing larger research methods courses;

- Interlace generic with discipline-specific examples, references and case studies wherever possible, highlighting relevance, engaging the audience and putting basic principles into context;

- Translate jargon for the audience, explaining principles and issues in language researchers/students can understand;

- Offer access to customised DMP guidance for the discipline so students can produce plans specifically relevant to them;

- Use trainers with extensive knowledge of the discipline, who can provide the context and interlaced examples that engage students and make the topic seem relevant to them;

- Offer training in the basic principles of data management at an early stage in postgraduate studies, allowing students to begin their project using best practice;

- Be concise, with basic modules short enough to maintain interest and be integrated into larger research skills courses;

- Deliver face-to-face training, as attendees find the opportunity to exchange experiences and thoughts with others invaluable. However students also want access to online training materials for ongoing reference and for those unable to attend courses in person;

- Stress the potential benefits associated with good data management practice, such as helping researchers to secure funding and meeting legal requirements;

- Work with professional bodies and funders to endorse and promote good data management practice, helping students and researchers to have support and potential reward for their efforts from leaders and funders within their discipline.

The work begun by DaMSSI is now being taken forward by the RIN and DCC. The DCC are keen to develop DaMSSI's work with the models by mapping course details to their own curation lifecycle model, as well as using the Vitae RDF and Seven Pillars to draw out longer term data management skills development for specific disciplines. The RIN Information-Handling Working Group is interested in analysing current UK LIS courses against the skills identified by DaMSSI and the RDMTrain projects, and in providing further recommendations on convergence of these with graduate LIS courses. There have been expressions of interest in extending the suite of career profiles by the DCC, members of the RIN Information-Handling Working Group, professional bodies and some international partners. The EU-funded DigCurV project will also reference the findings of DaMSSI in their training profile work as part of the design and development of a digital curation training curriculum.

Acknowledgements

References

Ball, A. (2011). *How to license research data.* DCC How-to Guides. Edinburgh: Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/resources/how-guides.

Davidson, J. & Jacobs, N. (2008). *Capacity and skills issues.* Report of the JISC Innovation Forum 2008. Retrieved from http://jif08.jiscinvolve.org/wp/theme-2-the-challenges-of-research-data/sub-page-2/

Goldstein, S. (2010). *Data management, information literacy and DaMSSI.* Retrieved from http://www.rin.ac.uk/our-work/researcher-development-and-skills/data-management-and-information-literacy

Hank, C. & Davidson, J. (2009). International Data curation Education Action (IDEA) Working Group: A report from the second workshop of the IDEA. *D-Lib Magazine 15(3/4).* doi:10.1045/march2009-hank

National Science Foundation. (2005). *Long-lived digital data collections: enabling research and education in the 21st century.* Retrieved from http://www.nsf.gov/pubs/2005/nsb0540/nsb0540_5.pdf

Research Information Network. (2009). *Joint response to Vitae consultation on draft Researcher Development Framework.* Retrieved from http://www.rin.ac.uk/resources/consultation-responses/joint-response-vitae-consultation-draft-researcher-development-fram

Research Information Network. (2010). *Managing research data: A guide to biocuration.* Retrieved from http://www.rin.ac.uk/resources/factsheets/managing-research-data-guide-biocuration

Research Information Network. (2011). *The Researcher Development Framework and the Seven Pillars of Information Literacy.* Retrieved from http://www.rin.ac.uk/our-work/researcher-development-and-skills/information-handling-training-researchers/rdf-and-pillars

Society of College, National and University Libraries. (2011). *The seven pillars of information literacy.* Retrieved from http://www.sconul.ac.uk/groups/information_literacy/seven_pillars.html

Ward, C., Freiman, L. Jones, Snow, K. & Molloy, L. (2010). *Incremental scoping study and implementation plan.* Retrieved from http://www.lib.cam.ac.uk/preservation/incremental/documents/Incremental_Scoping_Report_170910.pdf