Final Report

Project Information			
Project Identifier	*To be completed by JISC*		
Project Title	DaMSSI (Data Management Skills Support Initiative)		
Project Hashtag	#DaMSSI		
Start Date	01/11/2010	**End Date**	30/08/2011
Lead Institution	DCC at HATII, University of Glasgow		
Project Director	Joy Davidson, DCC		
Project Manager	Joy Davidson		
Contact email	Joy.Davidson@glasgow.ac.uk		
Partner Institutions	Research Information Network; Digital Curation Centre; University of Newcastle; CILIP Information Literacy Group; University of Glasgow; Edge Hill University; SCONUL Working Group on Information Literacy; Vitae; Loughborough University; Higher Education Academy; British Library; University of Southampton; Research Libraries UK; DPC; BAILER; Glasgow Caledonian University; UKCGE		
Project Web URL	http://www.rin.ac.uk/data-management-skills and http://www.dcc.ac.uk/training/data-management-courses-and-training/skills-frameworks		
Programme Name	Grant 04/10: Managing Research Data programme: Promoting discipline-focused research data management skills		
Programme Manager	Dr Simon Hodson		

Document Information			
Author(s)	Kellie Snow; Laura Molloy		
Project Role(s)	Research Officers		
Date	27/06/2011	**Filename**	Draft
URL			
Access	This report is for general dissemination		

Document History		
Version	**Date**	**Comments**
0.1	27/06/2011	Initial draft for review
0.2 – 0.3	24/06/2011	Review (Laura Molloy, Joy Davidson)
0.4	05/07/2011	Revised draft for review by JISC
0.5 - 0.8	23/8/11	Review (Kellie Snow, Laura Molloy)
0.9	14/9/11	Review (Joy Davidson, Stéphane Goldstein)
0.10	30/9/11	New expanded draft (Kellie Snow, Laura Molloy)
0.11	18/10/11	Review JISC, accepted.

Table of Contents

NB : This table of contents 'auto-populates' - to update the table of contents – place cursor in the table of contents, right-click your mouse, click 'update field', select appropriate option

1 ACKNOWLEDGEMENTS 3

2 PROJECT SUMMARY 3

2.1 PROJECT OUTPUTS AND OUTCOMES 4
2.2 HOW DID YOU GO ABOUT ACHIEVING YOUR OUTPUTS / OUTCOMES? 5
2.3 WHAT DID YOU LEARN? 7
2.4 IMMEDIATE IMPACT 13
2.5 FUTURE IMPACT 14

3 CONCLUSIONS 14

4 RECOMMENDATIONS 15

5 IMPLICATIONS FOR THE FUTURE 17

6 REFERENCES 17

1 Acknowledgements

DaMSSI was jointly funded by the JISC Managing Research Data Programme and the Research Information Network (RIN), to support the JISC Research Data Management training materials (RDMTrain) projects. It was run by the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow as a core partner of the Digital Curation Centre (DCC).

We would like to thank all the RDMTrain projects for their assistance, support and co-operation. The projects include:
MANTRA, University of Edinburgh;
DATUM for Health, University of Northumbria;
CAiRO, University of Bristol;
DMTPsych, University of York;
DataTrain, Cambridge University Library.

The DaMSSI team would like to extend their thanks to the DCC and members of the RIN Information-Handling Working Group, including representatives from Vitae and SCONUL, for their feedback, support and continuing promotion of DaMSSI's work. Special thanks go to Stéphane Goldstein, RIN for his guidance and advice throughout the project.

The DaMSSI team are also very grateful for the contributions of a number of professionals who generously gave up their time to be interviewed for the DaMSSI career profiles.

Finally we would like to thank the JISC Programme Manager, Dr Simon Hodson, for his support throughout DaMSSI.

2 Project Summary

Working as a support project for the JISC Research Data Management Training Materials (RDMTrain) projects[1], DaMSSI tested the effectiveness of the Society of College, National and University Libraries' (SCONUL) *Seven Pillars* information literacy model[2], and Vitae's *Researcher Development Framework* (RDF)[3] for consistently describing data management skills and skills development paths in UK HEI postgraduate courses. With the collaboration of the 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 investigated further 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. Accordingly students in a wide range of disciplines will need to acquire and hone their data management skills. The profiles will be of value for course providers but also may be a potential marketing tool for engaging with professional bodies to secure their endorsement for promoting and supporting data management skills development amongst professionals in their fields.

We found that both models had potential for consistently and coherently describing data management skills training and embedding this within broader institutional postgraduate curriculums. 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.

[1] http://www.jisc.ac.uk/whatwedo/programmes/mrd/RDMTrain.aspx
[2] http://www.sconul.ac.uk/groups/information_literacy/seven_pillars.html
[3] http://www.vitae.ac.uk/policy-practice/234381/RDF-overview.html

Findings were fed back to SCONUL and Vitae to help in the refinement of their respective models, and we worked alongside the RIN Information Handling Working Group to assist in the production of guidance for an information-handling version or 'lens' for the RDF. We are also working with a number of other projects such as the DCC and Digital Curator Vocational Education Europe (DigCurV[4]) to investigate ways to take forward the training profiling work we have begun.

Main Body of Report

2.1 Project Outputs and Outcomes

Output / Outcome Type *(e.g. report, publication, software, knowledge built)*	**Brief Description and URLs (where applicable)**
Mapping RDMTrain course outputs to Seven Pillars and RDF	Following a series of site visits with each of the JISC RDMTrain projects, initial individual mappings were drafted by the DaMSSI team and sent to projects for comment and feedback. These were then brought together into a combined RDF and Seven Pillars mapping[5], which identified generic skills and discipline-specific differences.
Recommendations to Vitae and SCONUL	Feedback on models was gathered during the site visits and communicated to SCONUL and Vitae via email and short descriptions. SCONUL used this feedback to inform the ongoing development of a new version of the Seven Pillars research lens. Vitae accepted our recommendations and found them confluent with their plans to produce an information-handling lens, which they will ask DaMSSI staff to review. The SCONUL and Vitae responses are discussed more fully in section 2.3 below.
Contributions to development of information-handling 'lens' for RDF	Contributed to the development of an information literacy taxonomy for RDF, drafted by the RIN. Contributed to the information-handling lens for the RDF, produced by Vitae with input from the RIN Information-Handling Working Group.
Career profiles	Five profiles were produced highlighting the roles and responsibilities of various professions and how data management skills were relevant: - conservator - social science researcher - clinical psychologist - archaeologist - data manager These will be made available in the first instance at http://www.rin.ac.uk/data-management-skills and http://www.dcc.ac.uk/training/data-management-courses-and-training/skills-frameworks, and shared with relevant professional bodies.
Comparison and synthesis of the training approaches adopted by the JISC Managing Research Data Training Projects	At the close of the production phase of the RDMTrain projects, DaMSSI was in a position to review and synthesise the approaches taken by the training projects. Findings from this exercise are reported in section 2.3.

[4] http://www.digcur-education.org/
[5] Available at http://www.rin.ac.uk/data-management-skills

2.2 How did you go about achieving your outputs / outcomes?

One of the major aims of DaMSSI was to test the effectiveness of the RDF and Seven Pillars models for describing data management skills and skills development paths in UK HEI postgraduate courses. The DaMSSI team planned to achieve this aim through the following activities:

- Support the RDMTrain projects in engaging with the Seven Pillars and RDF, giving hands-on support in mapping their training materials to these models;
- Produce a comparison and synthesis of the training approaches adopted by the projects, reflecting similarities and disciplinary differences;
- Develop recommendations for Vitae and SCONUL on how models can be refined to include data management skills, particularly the development of an information-handling 'lens' for the RDF;
- Produce case studies and career profiles, offering guidance to other institutions wishing to use the models and 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.

Although all of the above objectives have been addressed by the project, there were some changes to the focus on some areas. As we began work on the career profiles it quickly became clear from the feedback we were receiving from the projects and interviewees that these documents had the potential to be extremely useful not only for HEIs delivering training, but also by other researchers in the data management field and by professional bodies who organised professional development training courses for specific professions. As a result, more time was dedicated to the production of these profiles and to using them to make contact with various professional bodies. At the same time, the proposed case studies work was refined into a more structured synthesis exercise as it was difficult to document the projects' use of the models in future roll-out and embedding of the training products unless DaMSSI ran well beyond the projects' completion.

Some of the work on developing an information-handling lens for the RDF was taken forward through activities coordinated by the RIN Information-Handling Working Group. The DaMSSI team concentrated on contributing to these activities wherever possible rather than duplicating effort. Joint activities included the review of the RIN WG information-handling taxonomy and cooperation on the production of 'The Informed Researcher' booklet. This cooperation ensured that the findings from DaMSSI were incorporated into the overall messages of the lens and support materials.

Methodology

Engagement with, mapping to and evaluation of models

The mappings began with site visits to each of the projects in February 2011, where we explained the purpose of DaMSSI and discussed the projects' initial thoughts on the usability and potential value of the SCONUL Seven Pillars and Vitae RDF models. After each visit we then produced initial mappings of the projects' outputs (as they stood at that point) to each of the models, and sent the mappings to the projects for feedback. Any issues we noticed were communicated to the projects to encourage dialogue. A combined mapping was then produced,[6] which drew from a joint RDF/Seven Pillars table produced by the RIN prior to the start of DaMSSI work. Here the individual mappings were synthesised and common abilities and areas of understanding were listed.

Recommendations to Vitae and SCONUL

In order to assist SCONUL in the development of their revised research lens of the Seven Pillars, and to inform Vitae on ways the RDF could be developed, timely feedback on the models from the projects was communicated to each organisation. We provided feedback and suggestions to SCONUL shortly after the site visits to coincide with their revisions timeframe, and sent Vitae a short report on our findings once the combined mapping had been analysed. The latter highlighted strengths and weaknesses, and listed general recommendations on how both models could be refined. Details of

[6] *Ibid.*

these findings, recommendations and the responses from SCONUL and Vitae are contained in section 2.3.

Contributions to RIN Information-Handling Working Group

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[7], and a guide entitled 'The Informed Researcher'[8], 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.

Career profiles

DaMSSI produced a total of five career profiles:

- conservator (for the CAiRO project);
- social science researcher (for the MANTRA and DataTrain projects);
- clinical psychologist (for the DATUM, DMTPsych and MANTRA projects);
- archaeologist (for the DataTrain project);
- data manager (in cooperation with the Scottish Centre for Social Research).

Each career profile provides a description of the profession, an outline of key roles and responsibilities, an explanation of how data management skills fit into the day-to-day responsibilities of the profession and finally a list of further reading and relevant professional bodies and standards. The profiles will be openly available from the DCC[9] and RIN[10] webpages for DaMSSI, and will be shared with relevant professional bodies. The approach to producing the career profiles began through consultation with each of the projects, asking them to suggest possible professions which would be relevant to graduates in their programmes and to provide, where possible, contact details of individuals within these professions with whom we could speak. A series of short interviews were then conducted with approximately three individuals per profile, along with some general desktop research, to learn more about the professions. The profiles were then drafted, and RIN worked on the design and format inspired by their previous biocurator career profile[11].

For the first four career profiles, we deliberately embedded a description of data management responsibilities within the description of the tasks and skills required by each profession. In many of the professions we investigated, data management activities were not explicit. However, upon completion of the interviews it became clear that effective data management is an essential and intrinsic aspect of the day-to-day activities in a diverse range of professions. We balanced this approach with the development of a fifth profile explicitly describing the role of a data manager. As a developing profession, this role is much debated and sometimes poorly understood. After interviewing and receiving feedback from various individuals currently in this role, as well as analysing the skills listed in a number of current data manager job descriptions gathered from recruitment agencies and from data centre staff, we built a profile of a data manager incorporating elements of the roles of 'data scientist', 'data manager' and 'data librarian' as defined by Swan and Brown (2008)[12]. Specifically, our definition allows for the inclusion of most aspects of the Swan/Brown definitions, but excludes the 'responsibility for computing facilities' from their 'data manager' definition, and 'people originating from the library community' from their 'data librarian' definition. These decisions were made as a result of the information gained in the initial interviews and were refined by feedback on the draft profile from a further set of individuals currently working in the role.

[7] www.rin.ac.uk/rdf-7pillars

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

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

[10] http://www.rin.ac.uk/data-management-skills

[11] http://www.rin.ac.uk/resources/factsheets/managing-research-data-guide-biocuration

[12] Swan, Brown (2008). *The Skills, Role and Career Structures of Data Scientists and Curators: An assessment of current practice and future needs: report to the JISC.* September 2008. Available at http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dataskillscareersfinalreport.pdf

When considering the suite of five profiles, we 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. We also see the profiles being of use to the professional audience. Once published the profiles will be sent to relevant professional bodies and, through follow-on activity by the DCC, 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. We hope the professional bodies will use the profiles as a way of reiterating their commitment to developing their own training or to endorsing and/or accrediting externally-developed data management training courses for their members. The DCC and the RIN Information-Handling Working Group are well-placed to take this work forward and to maintain communications with these professional bodies, and the DCC plan to produce further career profiles that, alongside those from DaMSSI, can be promoted to various disciplines and professional bodies. International contacts have also shown interest in expanding the series for their own contexts.

Synthesis of projects' training outputs and approaches

As the production phase of the RDMTrain projects drew to a close, DaMSSI had the opportunity to review the training outputs with the projects, and to analyse the approaches taken by the projects as a group, allowing us to identify where approaches were confluent and where they diverged. Each project in the strand gave us sight of their final report to JISC, which provided us with information on the scope, scale and focus of their training offering. We gathered information on what the projects felt had worked well, the feedback they had obtained from delegates, the usefulness of the RDF and Seven Pillars models during both the production process and the subsequent embedding of the training products. We were also interested in determining whether the projects had found generic guidance and materials to be sufficient for their training materials or if discipline-specific elements needed to be produced.

Engagement with professional bodies and LIS course providers

As stated above, engagement with professional bodies has focused around promoting the career profiles and using them as a way to connect with professional bodies to highlight the fact that data management is core to the professions they support and that provision of data management skills development courses should be included as part of ongoing professional development. During the project we realised that while there may be some areas of overlap between the skills being taught to researchers and those being taught to information science students, the reality is that there may be gaps emerging as well. Improved communication between the various data curation roles involved in the data curation lifecycle will be vital to ensure that the skills being taught to each group are complementary. As part of the mapping work, DaMSSI developed a list of recommendations for providers of postgraduate research data management training to illustrate the range of skills that may be needed by information professionals in order to support researchers in the future. The recommendations are listed in section 2.3 and are being communicated to CILIP and other relevant LIS contacts.

2.3 What did you learn?

Much was learnt during the course of the project: about the value of the models; the success and pitfalls of various training approaches of the RDMTrain strand; and how data management training might further become embedded into institutions and career development paths.

Comparison and synthesis of RDMTrain project approaches and outputs

A key objective of DaMSSI was to compare and synthesise the training outputs of the RDMTrain projects, to identify generic similarities and disciplinary differences, and to draw recommendations from this that could be taken forward by JISC, RIN, DCC and institutions embarking on research data management training. Along with highlighting issues about the value of the RDF and Seven Pillars models themselves, the combined mapping of the projects' course outputs to the models suggested that there was consistency in the data management skills required across the disciplines. The mapping identified core generic data management skills central to each project's modules, 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 subsequent synthesis work explored these findings further, by analysing the final course materials, feedback from students and the projects' own conclusions and recommendations to see if there was agreement with what the mapping had suggested. The following findings were drawn out of the process with regards to generic versus discipline-specific training. The number at the end of each point refers to the subsequent recommendation made in section 4 of the report:

- The generic principles apply across all disciplines, but some disciplines are more ready to recognise this than others (2);
- All projects agree on the need for discipline-specific definitions, examples, exercises, case studies etc. to illustrate the relevance of the principles to the audience (1,3);
- Participants respond well to discipline-specific examples and the opportunity to discuss issues with tutors and others in similar disciplines (3,7);
- A discipline-specific approach is more likely to engage students – in many cases principles are the same across disciplines but are more interesting to students if these principles can be seen in the students' own context (3);
- A data management plan (DMP) needs some discipline-specific interpretation or customisation to be understandable by students (6);
- Some disciplines may benefit from greater focus on specific areas of data management; e.g. DataTrain's archaeology material gave more detail that other RDMTrain courses on different types of data used in the field, and DataTrain's social anthropology module gave more detail on ethical issues and consent (4);
- Echoing the findings of the earlier JISC MRD Incremental project[13], many researchers don't understand much of the specialist language from the information or preservation worlds (5);
- Care is required to balance the need for discipline-specific detail in training with keeping training relatively brief and concise, if it is to be incorporated into existing research methods courses. The necessary brevity suggests there is only time to train the generic basics. Courses which found this balance showed better delegate retention (2).

In addition to the generic and discipline-specific issues, a number of points were also identified in relation to the course delivery approach and skills development beyond the training sessions themselves:

- *Timing and length* - Training should be at an early stage in the PhD process but key messages and pointers to local support should be reiterated as often as necessary throughout. Training should be concise and ideally integrated into existing institutional research methods courses. This is less of an issue with online/self-directed learning (8,9).

- *Delivery methods* - Course attendees liked the face-to-face element as they were able to ask questions and share experiences with others; feedback to MANTRA online training, for example, was that whilst well-received, the face-to-face element was something it lacked (10).

- *Take-up* - Researchers are still reluctant to put the effort into DMPs or keep plans up to date unless required to, despite understanding their usefulness. The attendance at some courses was also low, suggesting a continued lack of understanding of the significance of good data management practice (11,12).

- *Skills development* - Students would like to be able to customise the DMP, or perhaps as part of their skills development process; projects which placed a focus on the development of individual plans received good feedback. DataTrain archaeology's approach, where each exercise built towards the completion of a personalised DMP for each individual PhD student, was useful (6).

[13] See http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmi/incremental.aspx and http://www.lib.cam.ac.uk/preservation/incremental/.

RDF and Seven Pillars models

After completing the combined mapping exercise and soliciting feedback from the projects on the individual mappings, we were able to identify a number of ways in which the two models appeared useful for describing and supporting data management training, as well as several drawbacks.

Potential value and usefulness of the two models
From the DaMSSI perspective, mapping the projects' course details to the two models was relatively successful. The courses all mapped to the same or similar areas, which demonstrated the models' usefulness in indicating consistency and generic elements in describing training across the projects and disciplines. The wide scope of the RDF is such that most aspects of the courses fit somewhere within the framework, and personal skills and attributes which the courses developed were also adequately covered. The Seven Pillars process is immediately recognisable within a research environment, and so offered an understandable cycle to which the data management process could be applied.

From the projects' points of view, feedback indicated that both models are *potentially* useful for describing and planning data management skills and training, although projects showed limited appetite for deep engagement with the models during the life of the project, largely down to time constraints. General consensus was that the models provided a stable overall structure for a set of training modules, which could then be developed further for discipline-specific requirements. The descriptions of skills and abilities within the models were potentially helpful for describing and comparing the content of courses, and for providing a list of consistent training outcomes that were relevant to students, supervisors and course providers.

The projects felt that there was potentially greater value in making use of the models to help embed data management courses into an institution's research skills or continuing professional development curriculum. All projects could see the value of using a model such as the RDF (whose summarised version, the Researcher Development Statement, has been endorsed by a wide range of organisations[14], including the UK Research Councils and the HE funding bodies, and is being taken up by many HEIs) to describe and map their training outputs, as an institution would be more likely to include a course within its permanent curriculum if it had immediate relevance to the model. The RDF provides a common framework for supporting providers such as university staff training services and developers such as the RDMTrain research projects, helping them to talk to each other more effectively. The project teams understood that having a common frame of reference can help immensely with assessing institutional training provision and offerings, and that the models help to identify pathways through existing courses, enable new courses to be fitted into the framework, and can identify areas where new courses are required.

Whilst the RDF has obvious use at an institutional level, one project also saw value in the Seven Pillars as a way of engaging with institutional libraries about data management, and even soliciting their involvement in the development and delivery of data management courses. As libraries offer much support to researchers in terms of information-handling it is important that they are engaged in this work and are able to effectively assist their users. Through the mapping to the Seven Pillars, DaMSSI was able to define some initial recommendations for LIS course providers on what skills may need to be included in courses to ensure information professionals have training in the skills required to support researchers in managing their data. Both models do also have the potential to be successfully mapped to other models such as the DCC curation lifecycle model which would provide even clearer identification of any gaps in the support provision offered by services such as HEI libraries.

The results of the mappings supported claims by Vitae about how the RDF is valuable for students, supervisors and course providers. A student can easily see the knowledge and skills that are required to be an effective researcher and how to progress towards attaining further competency in each area, and if data management training courses map to this, students can understand how each course is relevant within the overall structure. The DaMSSI career profiles help to provide evidence to students

[14] http://www.vitae.ac.uk/researchers/278641/RDS-endorsements.html

that the skills developed through participation in these courses are indeed essential for undertaking day-to-day activities within a number of professions.

A model such as the RDF also provides a means of showing the importance of data management skills within the researcher's overall career development for the supervisor or mentor, who may not themselves be familiar with some of the key principles. In addition, a course provider can understand the importance of data management training if it is placed in the context of a researcher's overall career development, offering further opportunity for roll-out and embedding of courses. Again, we must stress here that there is a need to engage with professional bodies to secure their endorsement for data management skills and to provide access to courses – their own or those of others – to support graduates as they enter their professions and seek ongoing development opportunities.

In its project plan, DaMSSI identified that there are a variety of information-handling and data management courses being delivered within HEIs, but there is still no consensus on what constitutes a basic set of postgraduate data management skills. Having models such as the RDF and Seven Pillars offers the possibility of identifying the generic skills being taught across disciplines; this is the first step in linking disparate courses from various institutions or disciplines into one more contextualised pathway. The models can be used to define introductory level courses, with students then moving on to more advanced or specific topics identified as being relevant to their particular discipline. The DaMSSI team has also reviewed the Digital Preservation Outreach and Education (DPOE) pyramid[15]. There may be potential to make use of the three audience levels described there as an additional means of classifying course offerings more granularly.

Challenges and limitations affecting the potential value and usefulness of the two models
The mappings equally highlighted a number of limitations to the models in describing data management skills and training. Whilst many of the course details could be related to some aspects of the models, there were elements which did not fit easily into either model. The 'creation' phase of data production and management is one example. As the Seven Pillars looks at information literacy, the focus is on gathering and analysing published information and the skills required to do this. The phase of creating new data and the necessary data management skills which should accompany this do not have a clear place within the model, despite there being some overlap in the skills required to manage this information. The RDF does offer more scope for describing these skills, but even then there were some abilities that were difficult to map to this model. For example, the 'preservation' aspect of data management was not obviously reflected in the RDF, and difficult to specify through the descriptors available. The mappings therefore demonstrated the need for data management processes to be more explicit within the models if training is to be mapped to them successfully, which will then allow students and supervisors to realise how aspects of a data management course relate to the various skill-sets defined by the models.

Following on from this there were some issues with the language used in both models, on which a number of the RDMTrain projects commented. In the first instance, many project staff members were confused by the acronym 'RDF' for the Researcher Development Framework, this acronym already being widely understood in this community to denote a completely different concept[16]. In addition, each of the Seven Pillars has a name that has immediate relevance to data management, but the definition of these terms is at times different for different audiences. For example, the 'Plan' pillar in the Seven Pillars model focuses specifically on search strategies for locating information, whilst 'plan' within a data management lifecycle has a broader and earlier definition of planning how data will be managed at the same time as a research project is outlined. That process, however, would currently be more aligned with the Seven Pillars' 'Scope' pillar.

Course details using the same language do not always clearly map to the part of the model with the same name. This adds to the obfuscation created by terminology in data management in general; different disciplines frequently have different understandings of terms used. A few of the disciplines covered felt that the models were too restrictive in their descriptions and view of research, focusing more on either text-based or scientific materials. In particular, disciplines which relied on practice-

[15] http://www.digitalpreservation.gov/education/
[16] Namely the 'Resource Description Framework', a widely-used set of World Wide Web Consortium (W3C) specifications.

based research felt the model did not really encompass this type of research, and as a result mapping data management skills specific to these disciplines was not fully satisfactory.

Whilst most of the projects could draw parallels between their modules and aspects of the models, and found the initial mappings interesting, there was little real enthusiasm for using them to any further extent. Projects recognised the potential usefulness of the models for mapping training in order to recognise gaps and commonalities and for providing a common language to use with training providers, but appetite for use was limited when working in the context of a short research project timescale with deliverables to develop. There was recognition from members of the project teams that institutional training services were possibly more familiar with and more likely to implement the models, but there was limited evidence in some cases of detailed awareness on the part of the research project team of whether the institutional training provider already used or would be likely to use such reference models.

A number of projects did comment that timing was a major reason they would be unlikely to engage deeply with the models, as many module outlines had already been drafted by the time the projects learnt about the models or viewed the initial mappings. When consulted at the close of the projects about using the models for further embedding, only two projects showed any interest in using a model, specifically the RDF, for future activities.

Feedback to Vitae and SCONUL about the potential value and usefulness of the two models
Our initial findings about the usefulness of the models were conveyed to SCONUL and Vitae shortly after the site visits were completed. Both Vitae and SCONUL were sent a short report describing the work of the Initiative, the mapping process and feedback from the projects about the models. It was important to provide immediate feedback to SCONUL and Vitae rather than to do so at a later stage in the project as both SCONUL and Vitae were in the process of revising the models and had limited timeframes for making any substantial changes based on DaMSSI's findings.

The findings described below are those that were provided to SCONUL and Vitae as part of our feedback and are distinct from DaMSSI's overall conclusions and recommendations about effective training for research data management, which are reported in section 4. It is important to restate here that these findings were communicated to SCONUL and Vitae as they emerged during the project.

- *Useful for consistently describing learning outcomes of courses* – the combined models have proved potentially useful for giving structure to data management courses and for describing learning outcomes in a consistent and structured way. This is helpful for both adhering courses to a wider continuing professional development curriculum, but also in consistently describing courses for deposit in course and teaching repositories such as Jorum[17].

- *Useful for embedding courses within institution's continual professional development (CPD) curriculum* – according to our testers, one of the biggest advantages of using the models, most obviously the RDF, is their potential to enable data management courses to become embedded within broader continuing professional development curriculums within institutions. The endorsement of the RDF by RCUK and subsequent uptake of the RDF amongst HEIs is increasing significantly and courses which can demonstrate their relevance to the model may be more easily integrated into other training activities. In a similar vein, demonstrating links with the Seven Pillars gives the opportunity for data management courses to find a place within information literacy courses delivered by university libraries, which gives further potential for their uptake amongst researchers.

- *Need for an information-handling/data management lens for the RDF* – the original RDF is very broad in scope and this was seen as a limitation by our testers when we attempted to collectively map the training course details fully to the model. Testers agreed that an information-handling or data management 'lens' would give a much clearer focus on the relevant aspects of the model, and where individual training courses would fit. In the time since the mapping was undertaken with testers and feedback was provided, Vitae has begun development of their RDF information handling lens. It is important for the lens, as it is

[17] http://www.jorum.ac.uk/

currently being developed, to reflect as precisely as possible the wider data management lifecycle and how this fits with the RDF. The Seven Pillars model would have been one way to start the development of such a lens, but it would need further revision to reflect the wider data management lifecycle and how this fits with the RDF.

- *Need for improved language/terminology for models* – the terminology used in both models requires further refinement to make sense to research audiences and to more accurately reflect data management skills. The RIN Information-Handling Working Group has produced a taxonomy for information literacy for the RDF, which provides substantial progress towards answering this requirement.

- *Need to reflect all stages of the data curation lifecycle* – As the Seven Pillars looks at information literacy, the focus is on gathering and analysing published information and the skills required to do this. The phase of creating new data and the necessary data management skills which should accompany this do not have a clear place within the model, despite there being some overlap in the skills required to manage this information.

Response from SCONUL and Vitae
SCONUL reported that our feedback was very useful, and included a number of changes in their revised model as a result. One of the most notable changes was a bigger emphasis on data collection as well as information gathering; descriptions in the model now include able to 'identify different *data collection methods*', 'select most appropriate…*data collection techniques*', and '*assess credibility of the data gathered*'. This followed on from the feedback that the model was focused on published material and lacked acknowledgement of the creation element of data management.

Overall, Vitae welcomed the findings we provided to them in our feedback document. Vitae is confident that a new information literacy lens on the RDF will address many of the limitations of the framework identified by the projects and reported by us; Vitae is currently working to develop this lens by the end of 2011, and has invited the Initiative to provide feedback on it when the lens is available for review. In addition, Vitae has accepted our observation that the acronym for the Researcher Development Framework, 'RDF', caused some confusion among research and IT staff as this is already a well-established acronym for a completely unrelated concept, and that for this reason they are now avoiding using the acronym whenever possible.

Wider recognition and endorsement of data management skills development

The cycle of knowledge represented by the Seven Pillars model is widely taught within LIS courses. As the RDMTrain data management training was mapped to the Seven Pillars, a number of skills were identified which may be valuable to information professionals with regards to data management, particularly those involved in academic institutions and supporting students and researchers. The areas listed are not fully addressed by the model, but require a level of knowledge and understanding by the information professional if they are to offer information-handling support for digital data.

- *Knowledge of what constitutes digital research data* – a general awareness of the types of data which a researcher might be handling across a variety of disciplines;

- *Understanding of data created by the researcher* – an awareness of how this will be different to published data in terms of gathering and evaluating, and knowledge of appropriate file formats for different types of data;

- *Understanding of how to organise digital data* – an ability to structure file names, organise folders, and add metadata;

- *Knowledge of data sharing options* – an understanding of the various avenues available for sharing of data and legal restrictions;

- *Awareness of data preservation and curation options* - an understanding of alternatives including storage, repositories and ability to advise on preparing and depositing materials.

If information professionals are effectively to support researchers in the future, providers of information science courses must understand the changing needs and practices of researchers and update their courses appropriately.

Engagement with professional bodies is also a key part of skills development throughout a successful career. Communication with various professions as part of the career profile work has taught us that major professional bodies usually offer accreditation and continuing professional development courses for their members, often as a condition of membership. Encouraging them to acknowledge and endorse data management abilities is an important step towards ensuring these skills are developed beyond postgraduate level and maintained throughout working life. Work with LIS professional bodies such as CILIP is similarly the best way to engage with the LIS sector in including data management skills within their qualifications and training, as such bodies oversee individual LIS courses and members look to them for continuing professional development activities. Initial contact was made with CILIP during the project but the ongoing assessment of their accreditation criteria and cancellation of some of their training activities made this a difficult time to progress this work with them.

It has been challenging to engage with large professional bodies and to find relevant contacts who might be able to help take the work of the project forward. Smaller interest and regional groups seemed more receptive to our work and run more specialist courses in which the results of DaMSSI could be disseminated at a later date. Feedback from interviewees as part of our career profile work revealed that professionals are happy to attend for training courses run outwith their professional associations if the topic is relevant and the course is deemed credible. This highlights the need for agreed benchmarks for training course content and the skills taught within courses. The RIN WG is, at the time of writing, developing a draft set of criteria for self-assessing training courses. This may be a good starting point for benchmarking course offerings.

Practical limitations
A few issues with regards to running a support project such as this have been noted:

- *Working with a number of projects across institutions* – feedback and suggestions from the projects were essential for moving some of the DaMSSI work forward (e.g. the career profiles and synthesis work) and ensuring we offered the best support possible. This was at times difficult to solicit owing to the projects' own time constraints or in individual cases lack of interest in or engagement with DaMSSI and the models. Closer work with Vitae was also limited, as our key contact unexpectedly left at an early stage of the project.

- *Timing of the project* – As DaMSSI was scheduled to finish at the same time as the RDMTrain projects, our ability to optimise the support we could offer the projects was somewhat limited. A number of the projects commented that if DaMSSI had continued past their completion they could have provided more timely and detailed feedback on the models, and made more use of the support we might offer such as taking forward their work of embedding courses into institutional programmes and Jorum. Much of the mapping work and guidance such as the originally proposed case studies were also dependent on the projects completing their own work first. At the time of the site visits, some of the projects' modules were still in the design stage and so early mappings were somewhat preliminary. Equally, the ability to draft case studies of how each project had used the models was severely limited by the fact that this would only be clear at the end of the projects, by which time DaMSSI would also be completed.

2.4 Immediate Impact

The work of this nine-month project has led to a number of immediate changes:

- **RDMTrain projects** can show the relevance of their training in relation to the RDF and the Seven Pillars and use this to embed their courses in larger institutional CPD courses, and when engaging with institutional library services;
- DaMSSI has begun the work of mapping data management skills development throughout a researcher's career, and across a number of disciplines, which can be taken forward by the **DCC** and other projects;
- Data management skills are now more widely recognised as an integral part of a researcher's career development through our collaborative work with the **RIN information-handling working group** in developing the information literacy taxonomy and the 'Informed Researcher' guidance booklet;
- The value of data management skills has been recognised within the wider information literacy framework through our contributions to revisions of the **Seven Pillars** research lens;
- There is wider recognition amongst our stakeholders of the importance of good data management skills in a variety of professions, through the production of our **career profiles**;
- Both **course providers** and the **wider data management research community** have a clear set of recommendations to take forward when designing and delivering future data management training.

2.5 Future Impact

DaMSSI has merely begun some of this work in the data management skills career development area. Indeed it is anticipated that much of the long-term impact of DaMSSI's work is how it will be developed and taken forward through subsequent projects and collaborations. We anticipate the following future impact:

- Through the mappings, RDF guidance documents and career profiles, students and mentors can be shown the relevance of data management training in relation to career development both within academia and a wider range of professions;
- Institutional course providers can demonstrate the relevance of data management training courses in relation to the RDF;
- Through the synthesis work, the recommendations on generic and discipline-specific training can be taken forward by course providers and other projects to develop a clear set of data management training materials best suited to postgraduates within a given discipline;
- Professionals will be able more effectively to plan skills development over the course of their careers – both while they are studying within an HEI setting and more importantly after they graduate into their chosen profession;
- Professional bodies will be provided with documentation that demonstrates the significance of good data management, and will begin to include this within their CPD training;
- LIS course providers will be made aware of the potential gaps that may be emerging between information specialists' skills and those of the researchers they may be supporting;
- DaMSSI's work may lead to greater co-operation on training profiling work through DPOE, DigCurV and other similar projects. The DaMSSI team attended the International Curation Education (ICE) Forum and a pre-forum meeting with the Closing the Digital Curation Gap project and partners in June 2011 to discuss taking our findings forward with similar projects.

3 Conclusions

The overall conclusions to be drawn from the project's work are:

1. Postgraduates require training in the basic skills of data management which can be delivered as a set of core generic principles. Discipline-specific examples and references should be included alongside the generic to engage the audience and illustrate relevance and context;
2. The RDF and Seven Pillars models have potential value in mapping data management skills and courses;

3. Each model has potential to assist in the embedding of data management skills training into institutions, whether this be in CPD training supported by the RDF or in information literacy training by HE libraries supported by the Seven Pillars;
4. Each model is only useable in this context with refinement and guidance that allows the relevance of data management to be understood;
5. Postgraduates find a data management plan particularly useful and the more customisable it is to their individual circumstances, the more they are able to make full use of it;
6. Data management skills can be shown to be relevant to a range of professions outside academic research, as demonstrated by the career profiles;
7. There still needs to be more collaborative work done to help professionals find and assess a disparate range of course offerings to find the right course for them at the right time in their careers.

4 Recommendations

Recommendations for institutions and projects embarking on future data management training:

General:

1. **Review existing training initiatives and skills models before creating new resources** – this minimises instances of recreating existing work, allows projects to learn lessons from previous work and can contribute to greater cohesion of terminology; this clarity will increase usability and uptake of the resulting resources.

Generic vs discipline-specific training approaches

2. **Work closely with disciplinary experts to ensure that terminology used within courses is accurate and clear** – this includes agreeing a basic definition of core concepts such as what 'data' can be within the discipline. This is particularly helpful for non-science disciplines;
3. **Keep overviews and central descriptions of topic areas basic and generic** – this introduces the topic at a level that is interesting but digestible for PhD students. It also allows modules to be more easily integrated into existing larger research methods courses;
4. **Interlace generic with discipline-specific examples, references and case studies** wherever possible – this highlights relevance to the audience, engages them, puts basic points into context and makes them understandable;
5. **Acknowledge accepted research practices within the discipline and work to develop training materials that reflect these practices –** e.g. kinds of data handling, research funder expectations, popular archives and repositories, etc;
6. **Translate jargon for the audience** –training providers should avoid using acronyms and jargon and explain principles and issues in a language researchers/students can understand;[18]
7. **Make use of recent funding body mandates for data management planning to embed data management planning as part of research practice** –students wish to produce a plan specifically relevant to them, often as a learning outcome of the course or as part of their wider skills development. Self-directed learning with access to customised guidance for the discipline and moderated exercises around the development of a data management plan works well;
8. **Have extensive knowledge of the discipline** – trainers who know the discipline well can provide the context and interlaced examples that engage students and make the topic seem relevant to them.

[18] For more on the value of understandable language to the success of research data management training, see the *JISC Incremental Scoping Study and Implementation Plan*, p18, point 2. Available online at http://www.lib.cam.ac.uk/preservation/incremental/documents/Incremental_Scoping_Report_170910.pdf.

Timing and delivery

9. **Offer training in the basic principles of data management at an early stage in postgraduate studies –** this allows students to begin their project using best practice and the production of a meaningful data management plan;
10. **Be concise** – basic modules should be short enough to maintain interest and be integrated into larger research skills courses;
11. **Students prefer face-to-face training** – 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 (e.g., distance learners).

Take-up

12. **Stress the potential benefits associated with good data management practice** – in addition to being part of good research practice, evidence of good data management can also help researchers to secure funding, respond credibly to questions regarding their research findings, and ensure that legal requirements associated with their research activity are met;
13. **Work with professional bodies and funders to endorse and promote good data management practice** – students and researchers may lose incentive to undertake data management activities without endorsement and potential reward for their efforts from leaders and funders within their discipline;
14. **Classify course offerings–** make sure that the anticipated outcomes of attending training courses are clearly defined to allow participants to select the training that best meets their learning objectives. The DPOE pyramid[19] may be a useful model for identifying courses aimed at executive-level strategic planners, operational managers and practitioners;
15. **Agree benchmarks –** work to secure agreement on benchmarking learning outcomes and means of assessment so that courses from a range of training providers can be effectively compared.

It should be noted here that our recommendations relate directly to our findings from working with the five RDMTrain projects and the disciplines and student levels they were targeting. The balance between generic and discipline-specific training needs may be different in other disciplines and indeed for researchers who are further along in their career paths. DaMSSI recommends that these aspects be investigated further as funding allows.

Wider community recommendations:

1. **Reuse JISC RDMTrain outputs to help embed data management training within other HEIs** – the outputs of the RDMTrain projects and DaMSSI should be promoted for reuse within other institutions; in particular the newly funded JISC 07/11 projects may be well placed to tailor and reuse the JISC RDMTrain outputs within their own institutions;
2. **Engage with professional bodies** - engage with professional bodies to highlight the importance of data management skills throughout professional careers and how information professionals can offer support through development of their own skills.

Recommendations for JISC:

- **Allow support projects to run for a minimum of two months beyond the life of the programme projects** – DaMSSI was originally scheduled to run for one month beyond the end of the RDMTrain projects to allow time to reflect on final findings and outputs from each of the projects. We agreed a time extension with the programme manager to allow for more effective synthesis of the RDMTrain projects' results. This extra month allowed us to more effectively assess and disseminate the projects' outputs into the support project findings. We recommend that JISC considers allowing future support projects to run for two months beyond the life of the projects they are supporting as standard practice. This helps to ensure that all the relevant project information is publicly available and can be fully assessed.

[19] See http://www.digitalpreservation.gov/education/

Recommendations for Vitae/SCONUL (beyond those reported in section 2.3 above)

- The RDMTrain projects recognised the value of the models but perceived implementation challenges, largely due to their project timescales and production deadlines. DaMSSI appreciates that familiarisation with and adoption of a new model or framework requires a certain input of time and effort. Some level of sustained support from those bodies which develop the models will help to facilitate implementing the models within institutions.

5 Implications for the future

As stated in the 'Future Impact' section, there are several strands of the work started by DaMSSI that can be built upon by future projects. Discussion of the development of some areas will take place in the coming months, but the following are what we see as useful future work:

- 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 RDF and Seven Pillars to draw out longer term data management skills development for specific disciplines;
- The EU-funded DigCurV project may incorporate some findings of DaMSSI into their design and development of a digital curation training curriculum;
- 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 to provide further recommendations on convergence of these with graduate LIS courses;
- Further collaborative work is planned on training profiling with projects and initiatives including DCC, DigCurV, CDCG and IDEA;
- There is potential to extend the suite of career profiles, possibly in consultation with professional bodies. Expressions of interest in further work have been made variously by the DCC, members of the RIN IH WG, the Society of Biology and some international partners such as University of Wisconsin-Madison;
- Investigate the potential use of the JISC funded eXchanging Course Related information (XCRI) specification as a means of aggregating and making more accessible the JISC 04/10 projects' course descriptions.

6 References

Digital Curation Centre (DCC)
http://www.dcc.ac.uk

Humanities Advanced Technology and Information Institute (HATII)
http://www.glasgow.ac.uk/hatii/

JISC Managing Research Data training projects
http://www.jisc.ac.uk/whatwedo/programmes/mrd/RDMTrain.aspx

Research Information Network (RIN)
http://www.rin.ac.uk

RIN Information Handling Working Group
http://www.rin.ac.uk/info-handling-group

RIN. *Biocurator career profile.*
http://www.rin.ac.uk/resources/factsheets/managing-research-data-guide-biocuration

SCONUL. *The Seven Pillars of information literacy: a research lens for higher education.*
http://www.sconul.ac.uk/groups/information_literacy/publications/researchlens.pdf

Vitae. *The Researcher Development Framework (RDF)*

- Homepage http://www.vitae.ac.uk/policy-practice/234301/Researcher-Development-Framework.html
- Full model http://www.vitae.ac.uk/CMS/files/upload/Vitae-Researcher-Development-Framework.pdf

Swan, A., Brown, S. (2008) The Skills, Role And Career Structure Of Data Scientists And Curators: An Assessment Of Current Practice And Future Needs - Report To The JISC. Available at http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dataskillscareersfinalreport.pdf