

Delivering Innovative RDM Training: The immersiveInformatics Pilot Programme

Anna Shadbolt
University of Melbourne

Leo Konstantelos
University of Melbourne

Liz Lyon
University of Pittsburgh

Marieke Guy
Open Knowledge Foundation

Abstract

This paper presents the findings, lessons learned and next steps associated with the implementation of the immersiveInformatics pilot: a distinctive research data management (RDM) training programme designed in collaboration between UKOLN Informatics and the Library at the University of Melbourne, Australia. The pilot aimed to equip a broad range of academic and professional staff roles with RDM skills as a key element of capacity and capability building within a single institution.

Received 13 January 2014 | *Accepted* 26 February 2014

Correspondence should be addressed to Anna Shadbolt, Library Research, University of Melbourne, Victoria 3010, Australia. Email: annams@unimelb.edu.au

An earlier version of this paper was presented at the 9th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution (UK) Licence, version 2.0. For details please see <http://creativecommons.org/licenses/by/2.0/uk/>



Introduction and Objectives

The advent of data-intensive research in the *Fourth Paradigm* (Hey, Tansley, & Tolle, 2009) has led to a requirement for enhanced support to promote good research data management (RDM) practice within universities and research institutes. In some higher education institutions, professional support services, such as academic libraries, have begun to consider the range of issues, tasks and services that are required. Many stakeholders are involved in RDM and the range of roles and responsibilities have been documented (Lyon, 2012).

New roles have been proposed, such as data librarian and data scientist, whilst in other cases existing roles have been extended to embrace the growing data challenge, such as subject/liaison librarians. A recent Research Libraries UK (RLUK) report (Auckland, 2012) highlighted the RDM skills gaps of subject librarians. An older report by the Australian Partnership for Sustainable Repositories (APSR) (Henty, Weaver, Bradbury, & Porter, 2008) identified a recurring requirement from Australian Universities for data management training for researchers. The 2011 Strategic Roadmap for Australian Research Infrastructure states:

‘...the most pressing issue that impacts on research infrastructure is the training and development of skilled staff to provide technical support to the research community.’ (Commonwealth of Australia, 2011).

In each case, a requirement for dedicated, cost-effective training programmes has emerged to build RDM capacity and capability within the community. This paper presents the findings, lessons learned and next steps associated with *immersiveInformatics*: an innovative RDM training programme designed in collaboration between UKOLN Informatics and the Library at the University of Melbourne, Australia.

The programme objectives were to:

- Develop an innovative, multimodal approach to teach the foundations of good research data management practice;
- Demonstrate practical research workflows within the research lifecycle across different disciplines;
- Apply basic data management tools to curate datasets;
- Facilitate an interactive experience with opportunities for shared learning, networking and reflection;
- Plan for a cost-effective course, which is also of high quality and fully evaluated to ensure continued programme refinement and development.

Programme Structure

*immersive*Informatics¹ was conceived as an innovative training methodology for enabling researchers and professional support staff in academic libraries, IT services, and research support offices to become familiar with current data curation issues and good research data management practice. The training programme structure aimed to embed the principles of data management both within the context of the research lifecycle and in varying international RDM policy frameworks.

Following agreement of a collaborative project plan, the partners focussed on selecting appropriate learning tools and platforms, as well as generating the module content. Learning tools included online forums, online data diaries and online delivery of module resources via the University's eLearning Blackboard platform. The project teams at both sites shared the content development tasks and reviewed all the modules as part of a quality assurance (QA) process, with revisions made where appropriate.

The programme was structured to run for ten weeks with one module completed each week over two mornings. Eight modules focussed on policy, context, RDM tools, services and best practice. Two modules focussed on the practical *immersion* of the themes covered in the other eight modules in a 'real world' research data context within the University.

The immersive component aimed to enrich the training in three novel ways:

1. The placement of participants in a research context to work alongside researchers or data managers. This provided participants with first hand observations of research processes throughout the data lifecycle, ideally in a domain relevant to their work.
2. Participants engaged in the research lifecycle by "co-curating" datasets in partnership with the researcher. Participants were expected to apply their RDM training, record their findings and provide feedback on how to best describe, manage, access, preserve and re-use the research dataset to the host researcher/data custodian.
3. Participants were invited to maintain a personal "data diary" as a reflective learning tool, to:
 - i. Describe experiences,
 - ii. Critically analyse situations,
 - iii. Develop new perspectives, and
 - iv. Evaluate the learning process (Fitzgerald & Chapman, 2000) associated with the *immersive*Informatics pilot.

The ten modules sequentially covered the following themes:

1. The RDM Landscape (local, national and international);
2. Understanding the Research Process (Immersive);
3. Requirements Capture;
4. Funders, Roadmaps and Data Management Planning;
5. Informatics Advocacy;

¹ *immersive*Informatics is based on an un-funded 2012 Jisc project proposal led by UKOLN Informatics.

6. Storage Infrastructure;
7. Data Publication, Sharing and Re-Use;
8. Preservation;
9. Legal and Ethical Issues; and
10. Revisiting and Reporting on RDM issues in the Research Context (Immersive).

Module Delivery

Modules included presentations, optional (advance) readings, group work and data diary reflection. Each module was anchored by an expert speaker and amplified by practical exercises, reading materials and other resources. Presentation slides and audio/visual recordings were available as online resources on the eLearning platform at the conclusion of each module.

A typical Module delivery structure divided the content across two days per week as shown in Table 1.

Table 1. A typical structure for the delivery of *immersiveInformatics* modules.

Component	Activities	Mode
Day 1: Three hours		
Introductions, objectives, goals	Introduce topic, exposition of specific objectives of the module, unpacking of specific participant goals, brief discussion, agree common goals, air concerns and issues.	Facilitation
Topic content	Expert/stakeholder representative speakers show good practice exemplars, case studies, demonstrating tools.	Presentations
Practical exercise	Complete a range of learning activities including quizzes, applying RDM tools to the working disciplinary datasets, critical reviewing peer materials, and discussion round-tables.	Group work
Day 2: Two hours		
Topic content	Focus on some of the data challenges associated with the module domain. Expert/stakeholder representative speakers show good practice exemplars, case studies, demonstrating tools.	Presentations
Practical Exercise	As above	Group work
Data Diary	Links to resources, participant lessons learnt, discuss issues, self-assessment exercise, feedback channel.	Moderation

Due to their immersive nature, Modules 2 and 10 followed a different format that included interaction with the host researcher and progress reporting to programme organisers (Table 2).

Table 2. Structure for the delivery of the immersive modules.

Component	Activities	Mode
Module 2: Understanding the Research Process		
Introductions, objectives, goals	Introductory meeting with host researcher, identification of a suitable dataset to curate, discussion of RDM issues and requirements within the context of the immersive research environment.	Immersion Facilitation
Module 10: Revisiting and Reporting on RDM Issues in Research Context		
RDM documentation	Finalise RDM documentation for immersive placement, including: Data Management Plan provided by the University of Melbourne ² ; Data Asset Framework ³ audit profile; other documentation relevant to the dataset. Consult with programme organisers.	Individual work; Facilitation
Presentation to host researcher	Report on completed documentation in a 30-minute presentation to the immersive placement researchers and course leaders. Receive feedback and discuss issues.	Presentation; Moderation
Final session	Prepare and deliver a joint 10-minute participant/host researcher report back on key outcomes and learning from the placements to all participants and host researchers. Open discussion: Research Data Management training within current role.	Presentation; Facilitation; Moderation

Modules 1-9 ran over a nine week period from August to early October 2013. Following feedback from the formative evaluation stage, Module 10 was pushed back by six weeks to provide participants with sufficient time to complete data diary entries, revise the topic content and reading materials, and finalise RDM documentation for their immersive placement.

Programme Evaluation

Participation

Fourteen participants registered for the pilot programme, with twelve of these completing all ten course modules. Two withdrew due to time clashes with scheduled training days or workload clashes with other duties. The remaining participants

² Participants used the University of Melbourne DMP template: <http://researchdata.unimelb.edu.au/how>

³ Data Asset Framework: <http://www.data-audit.eu/>

identified their role(s) as: Librarian (2); Research Assistant/Librarian (1); Research Archivist (2); Research Training Librarian (1); Archivist (2); Research Database Manager (1); and Research Fellow (3).

Evaluation Methodology

The *immersive*Informatics pilot was evaluated both internally and externally, based on a full expert evaluation in collaboration with the Australian National Data Service (ANDS)⁴. The programme assessment was undertaken through a mixed strategy of formative and summative evaluation techniques. Specifically, the evaluation methodology consisted of two stages and a total of five data collection instruments:

- At the formative stage, an online questionnaire was administered to trainees that aimed to collect feedback on five core areas of the training pilot – content, reading materials, session presentations, session activities and the data diary. Taking the form of a mid-term review, the questionnaire was available for completion during Week 5 of the training. The feedback was complemented by a focus group session chaired and facilitated by an external expert evaluator from ANDS.
- At the summative stage, a comprehensive online questionnaire was sent to trainees in order to collect feedback on the perceived impact of the course; the marketable skills attained; the suitability and relevance of the pilot content, structure and multimodal tools for the participants' needs; as well as information on highlights, potential omissions and future plans. Following on from the formative stage, a focus group session was organised between participants and the external evaluator to provide more insight into the data collected via the questionnaire. In parallel, a short survey was administered to participating researchers/data custodians as a means of evaluating the pilot experience from the perspective of the immersive hosts.

Besides the instruments in the evaluation stages, the data diaries were an additional source of – direct and indirect – feedback. In reflecting upon the weekly modules, some participants also tended to express their viewpoint on improving content delivery or emphasised on areas of the course where support was lacking. Although some of these musings may have eventually been reported in formal evaluation, the data diary entries offered a means of assessment that was rooted in the learning process and therefore incorporated a degree of introspection in the feedback that could be attained by neither the questionnaires nor the focus groups.

Evaluation Results

On the whole, the response rate was consistently high, with nine out of twelve (75%) trainees and four out of ten immersive hosts (40%) having completed the final online surveys. All participants attended at least one of the two focus groups with the external evaluator.

⁴ The Australian National Data Service is an Australian Government funded agency supporting best practice in research data management across Australian research institutions. More information: <http://ands.org.au/index.html>

Rather than disaggregating the evaluation feedback by evaluation stage, this section collates and reports on the evaluation results in four thematic groups; namely: course and module structure; multimodal tools; immersion; and impact and future plans.

Course and module structure

A fundamental aim of the *immersive* Informatics pilot programme was to evaluate the suitability, appropriateness and effectiveness of the proposed modular curriculum, as well as to gauge the relevance of an immersive, work-based training programme for an array of research data management needs.

Overall course relevance

Commencing with the latter, the pilot showed that more than half of the participants found the course at least 75% relevant to their needs – a pattern recurring in both formative and summative evaluation stages (Figure 1).

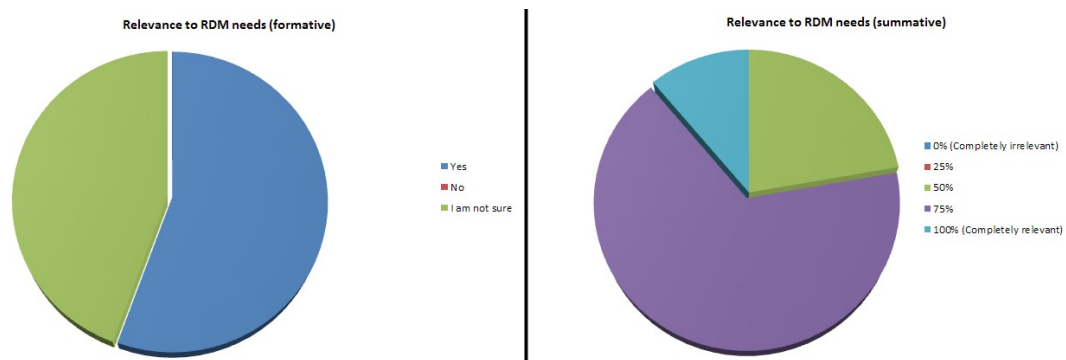


Figure 1. Perceived relevance to RDM needs during the mid-term review (left) and the summative evaluation stage (right).

The components pertaining to the local (i.e. Melbourne University or Australian) context and initiatives in research data management were deemed most relevant, while the components relating to the larger scale data management considerations – including the development of institutional strategies, policy and roadmaps – were reported to be less relevant.

Module components

The participants evaluated the four key components of each module (excluding the immersive modules, which were evaluated separately): content, reading material, session presentations and session activities. In particular, the assessment centred on the amount of content and reading material, and the role of session presentations and activities in facilitating the participants' learning.

Across the eight taught modules, there was general agreement that the course content adequately covered the topics necessary to build up a holistic view of RDM. However, concerns were raised during the focus groups that the amount and complexity of the content required some form of “study time” to be allocated by supervisors and managers during the weekly working schedule.

Similarly, the amount of reading material that was suggested in each module was found to be overwhelming. However, almost 50% of the participants identified that the reading material provided a “useful resource and the broader range of readings is more likely to apply to the broad range of course participants” and that it comprises “a great

resource [...] to track down the articles at a later date if and when needed to refresh the acquired knowledge.”

Face-to-face presentations were assessed as being central to the learning experience of the *immersive* Informatics pilot. Participants commented positively on the variety of invited speakers, the quality of delivery and the coverage of the presented topics. Session activities were less favourably received, with the majority of responses tending to disagree with their role in supporting the learning process.

The analysis on module components clearly pointed towards a case of balance between theory and practice. Many participants felt that the course was weighted in theory and they would have benefitted more from a course weighted in practical components. The theoretical constructs were seen as “too abstract” and thus made the activities more difficult for the group to complete. Several commented on the time constraints and felt that the emphasis on theory detracted from the (more valuable) practical examples and case studies. The lack of time to absorb/translate between theory and their applying that theory to their project was seen as a barrier to their ability to successfully transfer the learning into practice within the training period.

The value of a broad overview of content was acknowledged but it was at the institution level of example and practice where most participants gained value from the course. To some participants, it felt as being more like a component of a scholarly “Masters-like” course and thus was not the training experience they expected. If it were recreated as a “for credit” course participants felt that this would have significant benefits including:

- They would have spent more effort on the readings,
- Supervisors would have treated it more seriously and possibly made more concessions to the time required to complete the course,
- They would have gained a recognised and transferrable qualification which they could use in job applications,
- It would provide a signal of the high value data management has to the wider university and university executive.

Course structure

The participants were asked to evaluate five key areas of the pilot programme structure, including delivery format, types of schedule, starting point of the immersive component, and modules to remove or extend.

There was agreement that the preferred delivery format would be one day per week – either a full day (67%) or less than a full day (33%). The two half-day format adopted in the pilot was deprecated, mainly due to the time quotient. In terms of schedule, the delivery of modules on a weekly basis was preferred (67%) as opposed to a Bootcamp style/crash course/intensive training alternative (22%) or a combination of both (11%). Delaying the start of the immersive component to half-way through the course was suggested as more suitable (78%) than starting in week two of the course (22%).

In terms of modules that the respondents suggested should be removed in future iterations of the training programme, Informatics Advocacy was the most selected, although a number of participants felt no module should be removed (Table 3). The module relating to legal and ethical issues was most selected as a module to extend, closely followed by Module 5 (Storage Infrastructure) and none (Table 3).

Table 3. Number of respondents selecting various modules to be removed or extended in future iterations of the training (n=9, N=12)

Module	Which modules would you remove? (Select all relevant)	Which modules would you extend? (Select all relevant)
M1: RDM Landscape	1	0
M3: Requirements Capture	1	1
M4: Data Management Planning	0	1
M5: Informatics Advocacy	4	1
M6: Storage Infrastructure	1	3
M7: Data Publication, Sharing and Reuse	0	2
M8: Digital Preservation	0	2
M9: Legal and Ethical Issues	1	4
The immersive modules (M2 and M10)	1	1
None	3	3

Multimodal tools

Two online forums (Immersive Informatics and Social) and the data diary were set up in the eLearning environment to support participant learning. Six participants (67% of respondents) reported that they used the forums, with all reporting that they did not gain any real benefit from using them. The three respondents who did not access the forums attributed lack of time to their decision not to engage with the forums.

Five respondents reported using the data diary, with three of these stating that completing diary entries was helpful to their learning. Again, lack of time was identified as a key factor that limited engagement with these online tools.

Immersion

Participants were asked how the immersive part of the course matched their interests and needs. Participants reported varying experiences, with most finding that immersion supported their learning. Some excerpts from participant responses include:

“...[It] gave me time to think about our RDM needs...very useful.”

“... I was much more aware of what the difficult issues were and was able to focus on how to make it work.”

“[It] matched my needs, not sure if it matched my interests.”

“...I can see that there is a natural progression from being more involved with the planning at the initial stage and also at the end of a research project.”

The ‘success’ of the participant immersive experience varied across the group and that it was dependent on the right fit of project and participant expertise, experience and

enthusiasm. Despite the apparent difficulty of finding project partners, many in the group would have preferred to attempt to find their own.

Participants were asked to indicate their perception of the impact of their training on the immersive host environment. Eight participants responded to this question, with five being unable to determine their impact despite positive interactions. Four respondents believed that their participation made a noticeable impact. Some comments include:

“[I have] greater certainty that best practice data management practices will be implemented.”

“The greatest impact was establishing a long-term preservation plan.”

Immersive hosts also assessed the impact of the placement on their practices. Four responded (40%) with all stating that:

- Some of their data management needs were met by having a trainee;
- They contributed to the learning of the training participant;
- They had the opportunity to learn new things from hosting a training participant, including:
 - “An expanded framework for considering data management.”
 - “More systematic management of data.”
 - “Assessing the characteristics of datasets; pros and cons of sharing; benefits of data citation, and recommendations for preservation.”
 - “The issues facing others and how they relate to our work.”

Three of the four responding hosts stated that the immersive placement helped to instigate workplace behaviour change, with two adding that the placement helped to instigate a culture change in their team.

Impact and future plans

Participants indicated a desire to continue to learn more and be an advocate for good practice in research data management. All considered options of how this could be integrated into their current roles. Some examples include:

- “Assisting PhD students with DMPs.”
- “Improving data management practice for my projects and with the centre as a whole.”
- “I would like to be able to include RDM as part of my responsibilities with the Faculty...I have included it in my performance plan for 2014.”
- “I would be interested in helping researchers with RDM processes if there were positions available.”

All participants reported a desire to maintain contact with the group and stay informed about new courses, refreshers and new resources as they become available.

Future Developments

Refinements to the module content will be made in 2014 to incorporate detailed feedback and recommendations. A review of the programme structure will build in more flexibility across content selection and delivery modes to better align with varying needs of stakeholders. Ongoing engagement with pilot participants will be used as a springboard for an RDM Community of Practice at the University of Melbourne.

An “immersive” RDM Doctoral Seminar Programme and two “immersive” Masters courses are also being designed and delivered in iSchool at the University of Pittsburgh.

The 2013 pilot course materials will be released during 2014 under a Creative Commons licence.

Acknowledgements

Special thanks to our expert external evaluator, Ms Karen Visser, Program Leader, Resources, Policy and Skills, at the Australian National Data Service, Canberra, ACT.

References

- Auckland, M. (2012). *Re-skilling for research: An investigation into the role and skills of subject and liaison librarians required to effectively support the evolving information needs of researchers*. Retrieved from Research Libraries UK website: <http://www.rluk.ac.uk/wp-content/uploads/2014/02/RLUK-Re-skilling.pdf>
- Commonwealth of Australia, Department of Innovation, Industry, Science and Research. (2011). *2011 strategic roadmap for Australian research infrastructure*. Canberra: Author. Retrieved from <https://www.education.gov.au/strategic-roadmap-australian-research-infrastructure>
- Fitzgerald, M., & Chapman, Y. (2000). Theories of reflection for learning. In S. Burns & B. Chris (Eds.), *Reflective practice in nursing: The growth of the professional practitioner* (2nd ed., pp. 1–22). Oxford: Blackwell Science.
- Henty, M., Weaver, B., Bradbury, S., & Porter, S. (2008). *Investigating data management practices in Australian Universities*. Canberra: Australian Partnership for Sustainable Repositories.
- Hey, A., Tansley, S., & Tolle, K.M. (2009). *The fourth paradigm: Data-intensive scientific discovery*. Redmond, WA: Microsoft Research.
- Lyon, L. (2012). The informatics transform: Re-engineering libraries for the data decade. *International Journal of Digital Curation*, 7(1), 126-138.
[doi:10.2218/ijdc.v7i1.220](https://doi.org/10.2218/ijdc.v7i1.220)