Creating an Online Research Data Management Course: A Conversation with Data Librarians Robin Rice and Stuart Macdonald at the University of Edinburgh Data Library (UK)

Andrew T Creamer

Available at: https://works.bepress.com/andrew_creamer/8/
The University of Massachusetts Medical School’s Lamar Soutter Library has teamed up with libraries at the Marine Biological Laboratory and Woodshole Oceanographic Institute, Northeastern University, Tufts University, and the University of Massachusetts, Amherst to create an online case-based and open educational research data management course (http://library.umassmed.edu/escience_initiatives). In order for us to learn about the history of librarian and library involvement in the development of research data management curricula and resources, I traveled to the Edinburgh University Data Library in Edinburgh, Scotland in the United Kingdom to interview data librarians and pioneers in research data management education Robin Rice and Stuart Macdonald.

Robin and Stuart, well-known in the data library world, worked on the team that developed MANTRA (The Research Data Management Training Course), which was among the first online and open educational resources to teach students and early-career researchers the fundamentals of research data management (http://www.ed.ac.uk/schools-departments/information-services/about/organisation/edl/data-library-projects/mantra). MANTRA launched in October of 2011, and since that time other universities have taken notice and have turned to it as their model for developing their own interactive courses to provide guidelines for good practices in research data management. Visitors to the website will find learning modules on data management best practices that they can work through at their own pace and video interviews with academics about data management challenges.

The Edinburgh University Data Library is a library of firsts: Robin and Stuart were among the first data librarians to create a suite of research data support pages, and the University of Edinburgh was among the first to have a university-wide data policy. Robin was involved in setting up the UK Digital Curation Centre. Over the last decade Robin and Stuart have continuously found ways to innovate and advance the field of data librarianship. They have become involved in collaborative projects aimed at training University staff and students in data literacy, building the infrastructure for researchers to share and reuse data, expanding the library’s research data consultancy services, and creating new library services to support the University’s data policies.

I arrived in the beautiful and historic city of Edinburgh on a cool and breezy day in August. Crowds of international visitors were gathered in the streets for the world-famous Fringe festival. I made my way through the streets to meet Robin and Stuart in the modern offices of EDINA, the JISC-funded national academic data centre based at the University of Edinburgh (www.edina.ac.uk/about) and home to the data library. It was here that Robin and Stuart worked on a team to author and design MANTRA.
You can visit MANTRA’s website and peruse the course modules and watch the video interviews with researchers on relevant data management topics and also get hands-on practice by completing the software-specific data handling exercises ([http://datalib.edina.ac.uk/mantra/](http://datalib.edina.ac.uk/mantra/)).

The following is the transcript from our conversation.

Andrew: What projects are you working on currently?

Robin: At the University here we are working on a cross-divisional committee to put into place a plan that develops services to support the University of Edinburgh’s research data management policy- we were among the first universities in the UK to come up with a policy. Now that we have the policy we have to make it stick and develop more services, so the data consultancy services are one aspect of it, but there are other things, too. For us it’s a continuation of what we have been doing at the data library.

A lot of our peers that we know in the data library world, specialized academic libraries, are getting involved in data management and are reaching out to other disciplines—the sciences, for example. Traditionally the social sciences were more interested in sharing data.

Andrew: The UK currently offers many open educational resources for data management and preservation—what is the history behind these initiatives?

Robin: I was involved in the set-up of the UK Digital Curation Centre ([http://www.dcc.ac.uk/](http://www.dcc.ac.uk/)) and our director, Peter Burnhill, was also simultaneously the first director of the DCC. I would say that it was probably the eScience program that predated the DCC and out of that the data-driven science movement. JISC commissioned a report in 2003. Titled *Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision*, and authored by Philip Lord and Alison Macdonald ([http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf](http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf)), this report recommended that something like a Digital Curation Centre be established--and it was.

JISC put out a call for tenders and then this consortium that included Edinburgh, a library group based at University of Bath called UKOLN, and STFC, which had a different name at that time, and University of Glasgow came together and won the bid. But then after that, I think, the U.S. became the driver when the NIH and the NSF came out with their data management requirements. Then the research councils here started taking notice. From Europe I think the direction was Open Access first, and then sharing data, which is now a major initiative. And Australia is also very much involved.

Stuart: I think the JISC play a key role as a centralized funding body for these types of activities ([http://www.jisc.ac.uk/](http://www.jisc.ac.uk/)).

Robin: They have had a couple of initiatives. One was called the Managing Research Data Programme ([http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx](http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx)) that offers funds for institutions trying to come to grips with developing infrastructure for research data management. And in fact that is how MANTRA was funded because they had a training strand.

Stuart: And the DCC is funded by the JISC; EDINA is funded by the JISC.
We are part of IASSIST, an international organization, and we’re familiar with North American territory where data libraries are associated with large research universities. In Europe it is more of a data archives landscape.

Robin: Our director, Peter Burnhill, saw that as a distinction and built up that North American model of a data library here, which was unusual at the time. He did a tour of North American libraries in the 80’s, visited data libraries, and joined IASSIST. Edinburgh was a big computing center and had a role in inventing SPSS®. When the ’81 census data came out, it was machine readable for the first time—and as he jokes about it—the library couldn’t handle it because the tapes would roll off the shelves. These were some of the main reasons they created the data library here. He moved over from being a researcher statistician to take it on and found this tradition of data libraries in North America.

Andrew: Describe your career and journey to becoming a data librarian.

Stuart: My undergraduate degree was biochemistry and I did research for a couple of years. I was more interested in information—such as the science citation indexes than I was the actual subject matter. Subsequently I went to library school. I guess I should point out here the concept of the “accidental data librarian.” You may have heard reference to this. I worked at National Library of Scotland and I was working in a newspaper library, but fortuitously EDINA was looking for data librarians 13 years ago or so. I should point out that there are only three institutions in the UK that have data libraries: Oxford, LSE (London School of Economics), and Edinburgh. So it’s not like you’re applying for the same jobs in the paper advertising for data librarians! But I started here, shortly after Robin, as a data library assistant and over the years it developed into a full-time post.

Robin: I went to the library school at University of Wisconsin, Madison. I thought I was going to be a public reference librarian and then an academic librarian, but when I was a student I got a job at the data library there. There was an opening when I was graduating and I worked there for a long time, eight years. I found this job posted through IASSIST. Peter had been having trouble finding anyone that understood what the job was even about to fill the data librarian position when someone left—so the post stayed empty for a little while because he was busy building up EDINA and then they decided they should bring it to IASSIST where there would be data professionals. I just got lucky being able to move countries—getting the work visa—because no one else who applied was particularly qualified. I’m sure it would be different these days because now data is such a hot topic, but that was in ’98. So we were able to begin to build it up to the point where we are expanding. We have had a bit of funding to get more staff help here and there—partly through these projects—and now we hope the university will provide Information Services with more funds because of this [university data policy] implementation.

Career-wise I went to library school long enough ago to think I should work at the data library because it seemed like a good idea at the time to learn about computers and the internet. It was still a time when you were eligible for an email account if you were a graduate student but a lot of people saw no reason to get one, which is bizarre when you think about it. We were still learning searches on DIALOG through a dial-up subscription. And so at the data library I didn’t just learn about computers, I learned about mainframes and data.
So I kind of got mentored by my boss, Laura Guy, who had been a data librarian for quite a while. So, I wouldn’t necessarily consider myself an accidental data librarian but I certainly didn’t know that I was going to become one before I entered library school. In IASSIST we find people interested in data, not just librarians, who come from all sorts of backgrounds and disciplines and they have a variety of skills.

Which is good to see—on some of our earlier projects we were a little out there and feeling a little bit exposed. When we built the DataShare repository (http://datashare.is.ed.ac.uk/) it was part of a project that we worked on with other data librarians and repository managers from three other institutions: Oxford, LSE, Southampton, and us – DISC-UK DataShare (http://www.disc-uk.org/datashare.html). It was a great project because we learned a lot and we were among the first ones doing some things—talking about the barriers for sharing data and so on. It was a good opportunity to disseminate stuff that hadn’t been out there, but it was almost too early because the librarians weren’t interested yet; they were still getting to grips with publication repositories. And the people in the eScience world at that time did not think that librarians were going to be an appropriate way to go. In fact, there was a report that came out around that time that looked at us and others involved in the project when they had been commissioned to write about the careers of data scientists. There was this concern among people who focused on data for their research groups—it was seen like a real dive for your career because instead of publishing you’re doing the data stuff and you don’t get ahead (The skills, role and career structure of data scientists and curators: an assessment of current practice and future needs (2008) http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2007/dataskillscareers.aspx).

But one of the authors of the report took a strong interest in the data librarian aspect— it seemed from my point of view that after that report there started to be this new idea of data librarianship, which was what I’ll call this new breed of data librarianship because our traditional data librarianship is about finding and using data, and now it’s much more about putting the data somewhere, but encouraging its use can be overlooked. You can’t forget about the reuse.

Andrew: what skills does a data librarian need today?

Stuart: A colleague and I were discussing three discrete skill sets: you have your traditional library skill set, your statistical literacy, and a technological literacy. So it is a marriage of these, sort of like a Venn diagram where all three overlap and the bit in the middle is the skill set.

Robin: Communicator, too, because people are in their own silos and disciplinary worlds and you have to get people to talk to each other across these worlds, which takes confidence. So people skills have to come into it: being a good communicator, listener, and translator.

There was a poster that was done at the last DCC conference by two people at Sheffield, the head of the library school and the head of the library system (http://www.dcc.ac.uk/webfm_send/664), about what skills were needed-- competencies. You can get people coming out of different backgrounds that bring particular strengths to the team. And they might be missing in other strengths.

We’ve just gone through a hiring process here; we had a wealth of candidates and we made what seemed like a radical decision to not ask for someone with library skills. We were looking for a Ph.D.
student-type of person and we got not only that but more-- a Ph.D. in molecular biology who had research experience. And now he’s seeing all the possibilities for becoming a data scientist in the future.

And we did that because we already had librarians on the team and we wanted to balance it out. And now we think we’re going to get someone with skills from the art college who is pursuing a Ph.D. and she has skills with digital media and performance art, and so this will allow us to go out and do close work with more schools.

I love this idea of the embedded librarian. We are training liaison librarians in data literacy and having them apply it to their work. We are using MANTRA as a base and then bringing other aspects into it: group discussion, reflective writing, and applying these to their experiences working with other schools. And hopefully this builds up these librarians’ confidence to go out and work closer with researchers on future projects.

Andrew: How do you reach out to university stakeholders (faculty, student researchers) and make the Data Library accessible?

Stuart: We have affiliates and a well-established website, I think. Academics come and go and you have to build relationships and you get some kind of transference. For example, we get involved in some teaching activities at the School of Geography and the Graduate School of Social and Political Science. If staff turns over they will pass on the fact that we have been involved with them and then they will get back in contact. There is an emphasis on getting involved in building and maintaining relationships although over the last years we have been involved in so many other activities. This is a key aspect we have to develop over the next years and cement and build relationships with the research community.

And also with things like MANTRA, which can be seen as a kind of engagement tool as well as a resource for Ph.D. students. It was embedded within three disciplines so there are contacts there to be utilized. And also things like the repository. So these tools can be used as leverage to get a foothold into our researchers’ community whereas before it was more of a longhand way—sending out emails or leaflets or hand-outs to secretaries or heads of schools, heads of departments and relying on goodwill for them to be distributed. So now we have MANTRA, the DataShare repository, and our suite of data management webpages, which if you’ve seen is quite comprehensive. The data library is prominent on the IS website under the research support banner.

I have always found the term research support quite an amorphous term because almost every aspect of a university is in essence research support, but it is a conduit that can draw people to our services.

Robin: And under research support you can drill down to data support where our stuff lives. I agree about building relationships; it’s a time of change and maybe it feels like it has been for a while. A part of the reason why we got into the data management stuff, early, was in many ways in response to the same changes affecting specialized reference librarians that have had to recreate what they do. As data gets easier and easier to find and use people won’t need to call on us so much for expert help. We had to find what ARE people struggling with— information overload and data management and build on what we learned through the DCC experience and the preservation side. So we get fewer Ph.D. students
coming to us with the typical data library reference question: I want to do research on this, so where can I find the data? So we’re creating new services and now we have to create the demand for these services. It’s challenging at times but it’s exciting!

Andrew: How do students normally learn research data management?

Robin: I would say from relationships with other researchers, their supervisors. By trial and error even. For a while we were collecting horror stories from people who lost data.

Stuart: The RIN, Research Information Network (http://rinarchive.jisc-collections.ac.uk/), funded an exercise to look at information patterns, usage, access, etc. within the life sciences (http://rinarchive.jisc-collections.ac.uk/our-work/using-and-accessing-information-resources/patterns-information-use-and-exchange-case-studies). There were seven sub-disciplinary groups varying from systems biology to epidemiology and genomics, botany, etc. And what came out of that in terms of managing digital assets rather than research data was that some of these groups had a well-established system. I was part of a team observing how these disciplines were managing their data and some had protocols that were very rigid and some of the groups were even working with commercial partners. The point is there was no one size fits all—it depended upon the environment that the researchers found themselves in, and the culture, legacy and also it depended upon the content in terms of considerations regarding confidentiality or commercially sensitive content.

Robin: The people involved in Ph.D. training programs seem to agree that it is an important skill for them to learn, but I’m less convinced that all the lecturers would agree that it is an important skill for them to learn. There is no requirement for them to share their data; it’s similar to information literacy skills or IT skills that you may learn formally or you may not. Researchers want it to connect to data analysis so this influenced what we emphasized and is why we included the data handling exercises. These were very much grounded in the data analysis environment. We are still making the case—still persuading.

Stuart: It is also worth mentioning the Data Asset Framework (DAF) implementation project (http://www.dcc.ac.uk/resources/repository-audit-and-assessment/data-asset-framework), which in my mind was integral in developing services, policies, and research data management web materials. The DAF project developed an online tool that allows you to gauge the research data holdings at the department level, school level, institutional level, etc. Edinburgh was one of the pilot sites for the implementation project and there were four recommendations that came out as a result of this project: training, which MANTRA can be regarded as contributing to; a research data management policy, which we have at the university and which was passed by the senate in May of the last year; online guidance for staff, which is the suite of research data management webpages we have created.

Robin: We are streamlining these to describe more what PIs want to know instead of background information or instruction that is better placed in MANTRA.
Stuart: The fourth is the creation of cross-divisional services such as research data management and consultancy services and a data asset registry, to support the research data policy. Those were the four main recommendations.

Robin: It is a type of gap-analysis.

Andrew: How did you decide to use modules?

Robin: The recommendations for the chunking and remix of modules actually came out of a project we worked on way back in 2000 called DataTeach, which was before digital curation or data management was on the scene. It was a project with the social sciences to figure out how to get more data used in classroom teaching and what the barriers to that were. I remember Jackie Carter from Mimas, which was a partner on the project, writing about this “pick and mix” model being what researchers wanted. We also had an e-learning specialist working with us and she found the software that we used for MANTRA, called Xerte (http://www.nottingham.ac.uk/xerte/). That was an authoring tool that created these modules so that we didn’t have to design a whole website really. We just used the template. And it has the things in it that make it interactive.

And we could have used the University’s VLE (Virtual Learning Environment), which was WebCT as the time, but we wanted it open to the world and didn’t want it locked up in a VLE.

Before there was anything to look at, just some guidance pages, we white-boarded what we thought the main topics of data management would be and we tested those topics in our needs assessment interviews with the stakeholder academics.

Stuart: You can also see that they roughly mirror or map the research data lifecycle (http://www.dcc.ac.uk/resources/curation-lifecycle-model). Not directly, but it was used to inform the construct.

Andrew: Should a data management curriculum be customized to a discipline or should it be more general?

Robin: That’s a difficult question and one we have had many discussions about. On the one hand, I like the idea of the overall generic. However, the funding call was for specific Ph.D. programs, which is how we came up with the three stakeholders. And we have had discussions regarding this question and what we have learned in a pilot workshop with Ph.D. students-- they want it to be very discipline specific or else they’re not interested. And in other projects we have discovered a language issue, where you need to be speaking those researchers’ disciplinary language or else you’re going to lose them, so it’s tricky.

I think there are a lot of data management practices that are common across disciplines and things like case studies can help give you examples in their language with examples that they understand and that prove it’s relevant to them.

Some disciplines are very different. Those where researchers are doing a very particular data-intensive kind of science and using servers and command line programs and writing their own code to analyze the
data. These researchers would probably not find MANTRA relevant but hopefully it is useful to all the rest, the lone researcher with data on his or her desktop kind of model.

Andrew: What is your opinion on using case studies in data management curricula?

Robin: I think we’re going to use case studies to pilot our new services. I think they’re useful but they’re also kind of a drag to read. So you have to make sure you know what you want to get out of it and what information you want to distill from it. I feel that we should use case studies that lead into use cases based on real examples. I know it sounds like the same thing but a use case describes a situation when a user wants to do something specific. Case studies do help the person studying the case learn a lot because they have to take in a lot of detail, but for many others they do not want to take in that level of detail.

Andrew: What kind of resources did you look at before authoring MANTRA?

Robin: I think that at the time when we started we had been finding a few sets of guidance pages on the web for staff. One of the first we had found was MIT’s guidance on data management. So that was a good example. We wrote our guidance pages first, so there was our own suite of guidance pages. There were other projects in the training strand.

Stuart: Yes, we were aware of a number in the States like University of Minnesota. We were also under the impression that Melbourne was developing a course but they were going in a different direction.

Robin: Melbourne had created PG Essentials (http://www.eshowcase.unimelb.edu.au/packages/postgraduate-essentials-strategies-a-successful-start-your-phd), which in a way was a model of an online course for us. It is an online course for postgraduate students but with different content—it has information on a range of skills. We knew what they were doing in many places in Australia and I had done a study tour in Australia like you’re doing here, and I had found the libraries there were more keen in getting involved before they were here.

Robin: Something that would be good for library students that did not exist when we started is the Dutch Data Intelligence online training (http://dataintelligence.3tu.nl/en/about-the-course/).

Andrew: How did you manage the project and organize the curriculum and modules?

Robin: Xerte had the templates so we did not need to do a lot with the look and feel until the end. We knew the videos were going to be a lot of work—partly because we did not get a lot of time from the video producer to do a lot of iterative editing.

Stuart: There was also the process of identifying key researchers— we finally came up with a list of researchers that we thought had a story to tell.

Robin: At first we were ambitious that we would go out of Edinburgh but in the end decided to stick with researchers closer to home. The videos definitely had the most unknowns.
We authored the modules within our own team and each picked units where he or she was comfortable with the topics. We learned Xerte and how to create the interactive parts of the exercise.

Stuart: We also had to source images and multi-media materials: investigate licensing, find creative commons type material.

Robin: We wanted ours to be open so everything we used had to be open. We commissioned out the data handling exercises but it was a lot of work getting an agreement that what was in them was useful and coming up with an outline that we could bounce off the stakeholders to make sure what they wanted was in there. We also had to spend time editing these because these were large files. But we did find four experts who delivered the goods, so I wouldn’t say it was nightmarish just a big editing job.

Our partner, the Institute for Academic Development, which has relationships with Ph.D. training programs at schools all over the university, helped us come up with the three stakeholder groups—they advised us not to depend on the academics for very much in terms of time and input, so we didn’t. We interviewed them and tried to be true to the needs assessment we had conducted in what we created. It took a bit of chasing to make sure they were engaged with it; we launched in October so we missed the beginning of the academic year. So this year is our chance to see if it gets used and becomes more embedded in the stakeholder communities. We’ve been invited to come and give a talk in the School of GeoSciences. So that’s all--going by the plan, divvying up the work and regular project management.

Stuart: All of these discrete components were managed within a wiki environment.

Robin: And quality assurance-- we were always reading each other’s stuff and commenting and giving suggestions to make it better.

Stuart: So the wiki was a very helpful tool because you could gauge at certain times during the process and sort of see where you were in terms of the content and you could see where others were in terms of the development of the materials.

Robin: And then the user testing was another one--watching people engage with it and asking them questions about their experiences with it.

Andrew: What were the challenges you encountered creating MANTRA?

Stuart: Time - we had planned for ten hours of work for the production of a one hour module. Although it was a great deal more time in terms of polishing and preparing the final product.

Robin: [Laughing] Greater time and greater anguish--as long as we are happy with the results.

So with the work we did within our own team we could allot more time where it was needed but for the outside people you have to agree on what and when they are going to deliver, and what I learned from previous projects is that you get a contract in place that says what and when they will deliver it so they cannot wriggle out of it when things gets tough.

Andrew: What has been the response to MANTRA?
Robin: We got a lot of Twitter feedback, which was great. We have now added a testimonial page and have some quotes from people who liked it. We had someone from the DCC do an official evaluation (http://www.ed.ac.uk/schools-departments/information-services/about/organisation/edl/data-library-projects/mantra/deliverables).

Stuart: We have heard back from other librarians at other institutions that they plan to use it as a means to train and expose staff and students to data management.

Andrew: What’s next for research data management education? What’s missing or needs to be improved? What are the challenges?

Stuart: In the UK there really is no presence [in relation to research data management] in the library schools. The U.S. seems to be ahead of the game in that regard.

Robin: Our units don’t have data citation, for example, and if we were doing it now, knowing what we know now about the landscape, that would probably be a unit.

Robin: On the researcher side I would like to see it get more embedded, if that’s the right word, or ingrained in the discourse regarding research ethics because we still have a disconnect between the data sharing advocates and research ethics committees telling researchers to destroy their data if it’s not needed. And data sharing advocates’ attitude is take care of it so you can share it. So I think we still need to keep working on data sharing and it is the whole reason we did MANTRA. Our previous work has shown just how hard it is to create change within the university. We came up with different strategies—we needed a top down data policy but we also need a bottom up approach to reach researchers when they’re young enough to start changing the culture.

I think one of the reasons you can sell data management is because sharing is an inherent component in that and if you do good data management, I think, it’s really good evidence that you’re more willing to share it because you’re not afraid of your mistakes; you’ve been organized and done a good job documenting it all along. But I think we need to do more emphasizing the value of data sharing, which is tough if the older generation isn’t doing it. That’s how people learn in academe—a kind of mentor relationship, a type of pass down relationship. Although young people are sharing on Facebook, I have read evidence that shows that they do segment that part of their life and not apply it to their professional if they’re not taught not to.

Stuart: Another strategy is using high profile researchers as spokespersons to participate and champion the values of data sharing and good data management practices. For example, Professor Peter Clarke, who is a well-established researcher working at CERN, chairs the RDM Policy Implementation Steering Committee here at the University. So, getting these kinds of people involved, from an institutional perspective, must be seen as a plus for getting other researchers to take note of this as an area or activity that they should also be getting involved with.

Stuart: When we spoke to PIs as part of the RIN project mentioned above we asked them their thoughts on paying for an information specialist out of their funding to be part of their project team, and almost
every PI responded he or she would most probably put the money towards research than have a bespoke professional within their team. So this will present a challenge to selling the concept of embedded data scientists and librarians.

Robin: This might have been the first time that these researchers had encountered the idea and their responses might be different if you were to ask them again today. That’s where the funders come in with their evolving requirements for data management plans. If you don’t have a plan then you don’t get funded, and if you have a plan, then you will have to indicate if you are going to share your data or explain why you can’t. So this represents one brick in the wall coming down.