Digital Curation and Doctoral Research: Current Practice

Daisy Abbott
Digital Design Studio
Glasgow School of Art

Abstract
This article considers digital curation in doctoral study and the role of the doctoral supervisor and institution in facilitating students’ acquisition of digital curation skills, including some of the potentially problematic expectations of the supervisory relationship with regards to digital curation. Research took the form of an analysis of the current digital curation training landscape, focussing on doctoral study and supervision. This was followed by a survey (n=116) investigating attitudes towards importance, expertise, and responsibilities regarding digital curation. This research confirms that digital curation is considered to be very important within doctoral study but that doctoral supervisors and particularly students consider themselves to be largely unskilled at curation tasks. It provides a detailed picture of curation activity within doctoral study and identifies the areas of most concern. A detailed analysis demonstrates that most of the responsibility for curation is thought to lie with students and that institutions are perceived to have very low responsibility, whilst individuals tend to over-assign responsibility to themselves. Finally, the research identifies the most common types of support system for curation and suggests ways in which students, supervisors, institutions, and others can effectively and efficiently address problematic areas and improve digital curation within doctoral study.

Received 2 October 2014 | Revised 6 January 2015 | Accepted 10 February 2015

Correspondence should be addressed to Daisy Abbott, Digital Design Studio, The Hub, Pacific Quay, Glasgow, G51 1EA. Email: d.abbott@gsa.ac.uk

An earlier version of this paper was presented at the 10th 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/

International Journal of Digital Curation
2015, Vol. 10, Iss. 1, 1–17

DOI: 10.2218/ijdc.v10i1.328
http://dx.doi.org/10.2218/ijdc.v10i1.328
Introduction

More than ever before, doctoral students are engaging in research data creation, processing, use, management, and preservation activities (hereafter referred to as digital curation). For most research students digital curation is an intrinsic part of their study and a skill that they are expected to acquire to an appropriate level by completion.

‘I look upon the role of data management for a new researcher as being one of those essential skills that you really ought to get at the same time as you learn how to handle your references, as you understand methodology, as you get to grips with the theory that is going to set the frame by which you do your research. And it sits alongside those and it’s equal to them. […] I think it’s an essential skill, it’s not one which I think you can neglect.’ (Haywood, n.d.).

The Joint Statement of the Research Councils’ Skills Training Requirements for Research Students emphasised “that training in research skills and techniques is the key element in the development of a research student” (RCUK, 2001). Two particular expectations in terms of the research training provided by institutions are that students should develop the skills to “design and execute systems for the acquisition and collation of information through the effective use of appropriate resources and equipment” and “use information technology appropriately for database management, recording and presenting information” (ibid). Despite the now somewhat dated focus on data acquisition (as opposed to the whole data curation lifecycle) and databases (as just one form of digital asset), these 2001 guidelines (now incorporated into the Vitae Researcher Developer Statement as “information literacy and management” (2011)) demonstrate the firm expectation from funders that researchers engage in sound digital curation practices.

In short, doctoral students are expected to have all the necessary skills that are required by a professional researcher by the time they graduate and that, if they do not have these skills on commencement, the process of doctoral study will include the acquisition of this expertise.

Research Context

It is clear that, as digital data forms an increasingly large part of our lives in general and our research in particular, the integration of digital curation into expected research skills has become explicit. This is unsurprising: in recent years the research environment has seen a move towards an acknowledgement of the increasing prominence of large and/or complex data (just one example is the reworking of the AHRC Technical Appendix into a Technical Plan in 2013), robust data management plans as funding requirements (Burgess, 2013), and an increasing emphasis on Open Access (RCUK, 2014). Furthermore, there has been significant investment in the production of training resources to increase digital curation knowledge and expertise in the UK research base, in addition to the ongoing commitment to “promote and support good research data management and sharing for the benefit of UK Higher Education and Research” (Jisc, n.d.).
Extracts from the AHRC Technical Plan guidelines demonstrate the digital curation requirements of funded research, both before and after a project takes place. A full Technical Plan is required:

‘for all applications where digital outputs or digital technologies are an essential part to the planned research outcomes. A digital output or digital technology is defined as an activity which involves the creation, gathering, collecting and/or processing of digital information. […] The AHRC requires a minimum of three years after the end of project funding for both preservation and sustainability, but in many, if not most, cases a longer period will be appropriate. […] The AHRC normally expects digital outputs that are preserved and/or sustained to be freely available to the research community’ (AHRC, 2013).

Following quickly behind these expectations at funder level, institutions demonstrate similar expectations of their researchers. The need for the people who will be actually managing data on a day today basis to develop good practice is well established and, although previous evaluations have found that “training for researchers on information seeking and management is uncoordinated and generally not based on any systematic assessment of needs” (Research Information Network, 2009), in general the digital curation expertise of research staff is slowly improving. However, despite an identified need for training to be provided at an early stage of a researcher’s career (Jisc, 2011; Molloy, 2012; Ward et al., 2011), it is untypical for PhD students to be treated the same way as research staff when it comes to digital curation. Even taking into account recent advances in digital curation training and resources aimed specifically at doctoral students (e.g. MANTRA) and short courses run by the Digital Curation Centre (DCC), support is generally rare, patchy, ad hoc, or simply not promoted directly to students or supervisors. Doctoral-level digital curation education has been specifically identified as an area requiring further development and it is clear that supervisors are expected to include support for digital curation alongside all other elements of their supervision role (Poole et al., 2013). This can be seen in line with the general expectations placed on PhD supervisors by educational policies – “The candidate’s relationship with his/her supervisor is key to a successful research degree programme” (QAA, 2011) – and is sometimes stated explicitly by institutions. An early draft of the Open Access and Research Data Management Policy for PGR Students at the University of Exeter (an institution which is more engaged with this subject than most) stated that “the responsibility for research data management lies jointly with the Main Doctoral Supervisor and the PhD student” (University of Exeter, 2012). In later drafts and the final version of this document, the responsibilities of the supervisor have been softened, with a clearer delineation of the expectations from student, supervisor and institution:

‘4. The lead PGR Supervisor and the PGR student should discuss and review research data management annually, addressing issues of the capture, management, integrity, confidentiality, security, selection, preservation and disposal, commercialisation, costs, sharing and publication of research data and the production of descriptive metadata to aid discovery and re-use when relevant.
5. The University is responsible for the provision of training, support and advice on Open Access and research data management as well as the provision of a backed-up storage service for completed digital research data and for Open Access research publication, including papers and doctoral theses.

6. Responsibility for research data management lies with the PGR student, and if relevant, jointly with the Principal Investigator (PI) of the research project. The lead PGR Supervisor is responsible for advising the PGR student on good practice in research data management.’ (University of Exeter, 2013).

This example demonstrates that institutions are still in the process of negotiating exactly how to support doctoral students in digital curation, but it is clear that a burden of responsibility to mentor students in best practice remains with doctoral supervisors, who may not themselves have the necessary expertise. It is not my intention to suggest that every doctoral student should become an expert in data curation, nevertheless, it is a rare piece of research which does not create, manipulate, or record digital data in some form. Whilst some digital curation tasks are provided by others (for example, regular backup of data held on university servers or long-term migration of formats once deposited in a specialist repository), it is still unlikely that a student or supervisor would complete their research without any engagement with digital curation activities.

Literature Review: Data Curation and Doctoral Supervision

There is a great deal of literature aimed at both PhD students and their supervisors that aims to improve the quality and efficiency of supervision. Across much of this literature it is common to find practical advice: tips, specific procedures, and examples of techniques to implement that benefit both supervisor and student. However, despite the growing importance and prominence of digital data and outputs within doctoral study, digital curation advice remains firmly in the realms of specialist resources. Whilst there are a small number of published sources of information on the subject of data curation during doctoral study, they tend to be focussed on a top-down model of training provision, aimed at librarians and other information scientists or institutional management. For example, Piorun et al. (2012) offers a curriculum framework for the delivery of data curation teaching (in a librarianship journal) and Alexogiannopoulos et al. (2010) report on the implementation of a Data Asset Framework assessment aimed at University of Northampton research managers and Information Services. The value of a bottom-up approach is acknowledged by Ward et al. as they argue that “top-down, policy-driven, or centralised solutions are unlikely to prove as effective as clear, appropriate and practical support delivered to researchers in a timely manner” (2011). Whilst this publication is not specifically aimed at PhD students or supervisors, the authors highlight the importance of targeting doctoral and early career researchers in order to embed sound data curation practice within research and, pertinent to the supervisor-student relationship, describe the importance of one-to-one advice and support from a data curation mentor.

In terms of published works dealing primarily with the supervision process, various texts imply or refer specifically to the wide variety of data that can be produced as part of research (e.g. Barrett and Bolt, 2007; Cryer, 2006) and data generation is widely accepted as being at the core of much research. However, there is a significant gap in
the literature about why and how to manage, curate, and preserve digital data as part of a PhD. Of course, both doctoral students and supervisors can benefit from resources aimed at researchers in general, but the form that this information takes is not typically as published books or articles. Instead, the majority of guidance on digital curation takes the form of online resources and training programmes, many of which will be familiar to readers already.

- Digital Curation Centre (DCC)
- Research Data MANTRA
- Visual Arts Data Skills for Researchers (VADS4R)
- Incremental
- DataTrain
- Open Exeter
- Curating Artistic Research Output (CAiRO)
- Research360 and institutional data management plans

This list is not intended to be a comprehensive list of all organisations and resources that provide data curation services, simply those resources which focus on training – particularly skills development for post-graduate research students in the Arts and Humanities.

**Methodology**

This research developed a short questionnaire with the main aim of capturing a range of information from a statistically credible group of respondents. The survey was aimed at doctoral students (past and present), doctorate holders, and supervisors. It consisted of ten questions, along with a description of the research. Questions 1 and 2 were about the individual respondent in order to enable the comparison of responses across different phases of a research career and different disciplines. Question 3 asked respondents to estimate their level of digital curation expertise. Question 4 was aimed at students only and gathered information on the type and frequency of digital curation activities. These categories of activity were based on those described in the DCC Curation Lifecycle Model. Question 5 gathered data on the perceived importance of long-term preservation of doctoral research data. Questions 6-9 were about the perceived

---

1 Digital Curation Centre: http://www.dcc.ac.uk/
2 MANTRA: http://datalib.edina.ac.uk/mantra/
3 VADS4R: http://www.vads4r.vads.ac.uk/
4 Incremental: http://www.lib.cam.ac.uk/preservation/incremental/
5 DataTrain: http://www.lib.cam.ac.uk/preservation/datatrain/index.html
6 Open Exeter: http://as.exeter.ac.uk/library/resources/openaccess/openexeter/
7 CAiRO: http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmtrain/cairo.aspx
9 DCC Curation Lifecycle Model: http://www.dcc.ac.uk/resources/curation-lifecycle-model
responsibility of different roles (student, supervisor, institution, and other) in curating data both during and after the term of the PhD study. Finally, Question 10 assessed awareness and use of different resources and services for digital curation.

Results and Analysis

Overall survey results show 116 respondents with a good rate of questionnaire completion (abandon rate was < 25%). The range of research stages/roles represented was roughly equal between early doctoral students (comprising imminent and first-year students) (21%), mid or late students (28%), doctoral graduates who do not supervise students (23%), and doctoral supervisors (28%), as shown in Figure 1.

![Respondents by research stage](image)

**Figure 1.** Breakdown of respondents by stage of research or research role.

Perceived Importance of Digital Curation

As Figure 2 shows, very few of the respondents (around 10%) consider long-term preservation of their data to be of little or no importance. In fact, for doctoral supervisors and non-supervising PhD holders, over 50% consider long-term preservation to be extremely important. For students, taken as a single group, this figure was under 40%. However, under 5% of students felt that long-term preservation of data was of little or no importance, compared to just under 10% of PhD holders and supervisors. The students on average tended towards the middle of the scale when reporting on perceived importance, with a much greater number considering long-term preservation to be moderately important.

---

10 NB: Few questions had a 100% completion rate, therefore when data is shown as percentages these are of the total number of valid responses, ignoring any null values.
Perceptions of Digital Curation Expertise

Overall, the perception of respondents of their own digital curation expertise was very low. 74% of respondents stated that they had limited or no skills in digital curation and only 10% stated that they were “fairly skilled” or “expert”.

Figure 2. Perceptions of the importance of long-term preservation of data, comparison by role.

Figure 3. Perception of skills and expertise for digital curation.
It is useful to assess whether there are differences in the self-reporting of expertise between the supervisory role and the supervisees. Alexogiannopoulos et al. state that PhD students “were found to be generally less experienced in managing data than more senior researchers. For many of them it was their first time conducting research on such a large scale.” (2010) and this is borne out in the detailed results shown in Figure 4.

**Figure 4.** Comparison of digital curation expertise across different categories of researcher.

Aggregated into two broad categories (students and doctoral graduates), around 20% of PhD holders and supervisors report “no skills or expertise” compared to over 40% of students, and just over 60% report that they have limited or no expertise, compared to an average of over 80% of students. There was more variance in reporting a moderate or large amount of digital curation expertise, again with the first group reporting a greater expertise average than the students (around 40% compared to under 20%). Only one respondent reported that s/he was an expert in digital curation (a non-supervising graduate). Therefore, graduates report more confidence in digital curation than students, even though overall expertise levels remains low. Nevertheless, some have knowledge that could be fruitfully introduced to their students. Whether or not this knowledge is being effectively used remains a question. Alexogiannopoulos et al. note that in their study “some PhD students did say that they had received advice on storing data from their supervisors, however most interviewees seem to ‘go with what feels right’” (2010). This suggests that there seems to be a role for supervisors who do have expertise to advocate the student’s development of critical faculties in this area, or at a minimum to simply encourage the student to develop their digital curation expertise.

---

11 An interesting side note is that supervisors, on average, report slightly lower levels of digital curation expertise than PhD holders who do not supervise students, particularly in the “fairly skilled” and “expert” categories. This may indicate that PhD holders working in higher education have slightly lower expertise levels than PhD holders working in non-academic positions.
Breakdown of Digital Curation Activity in Doctoral Study

The questionnaire gleaned information about the type and frequency of different types of digital curation activity undertaken by students. An analysis of Figure 5 and Figure 6 leads to detailed information about the data lifecycle in doctoral research. Here, I focus on the most relevant issues. Overall, the responses show an active picture of data creation, use, and management taking place within student research. Clearly, despite the very low levels of reported expertise discussed above, students are, in general still undertaking digital curation tasks. These results demonstrate two main issues. Firstly, a relatively large number of students are aware that some digital curation tasks are relevant to their research but are not doing them (as shown by Figure 6, the main concerns are the omissions in deposit of data into managed repositories, technical preservation, and creation of descriptive metadata). Secondly, that some of the digital curation activity undertaken frequently by students themselves may in fact be being performed without the necessary skill levels to minimise both risk to data appropriateness, accuracy, and preservation, and effort on the part of the student. This has significant consequences for the quality of the research itself and the ability of the student to complete his/her PhD. For example, creating representations of artworks for inclusion in a written thesis is a highly skilled task with a direct causal link to the assessment of the overall research.

![Figure 5. Types of digital curation activity by frequency.](image)

12 This question was targeted only at students and therefore based on a much smaller sample than the majority of the survey (the question was answered by 57 out of the 116 respondents).
Figure 6. Frequency of digital curation activity, expressed as percentages of total relevant answers.

Perceptions of Responsibility for Digital Curation

The survey gathered information about the perceived responsibility for digital curation in order to identify potential problematic gaps in the roles undertaking particular activities. Respondents were offered four choices to assign responsibility: student, supervisor, institution, and other. As “Other” was very much an atypical choice, the bulk of the analysis focuses on the first three roles.

On average, respondents assigned 20% of the responsibility for digital curation to the supervisor both during and after the period of study. Unsurprisingly, the majority of responsibility was assigned to the student: 75% during study, dropping to just over 50% after the PhD ends. The responsibility of the institution is perceived to grow after the period of study but remains relatively low at just 33%.

These data demand a more thorough analysis to reveal the differing perceptions of each role. Figure 7 shows the range of responses for digital curation responsibility during doctoral study, broken down by the role of the respondent. Two main clusters can be observed, a tight cluster (A) showing that a large number of respondents (particularly students) believe curation responsibility to lie almost entirely with the student. The second looser cluster (B) shows responsibility shared largely between the student and supervisor with a low institutional responsibility. A formal cluster analysis corroborates these categories and shows that roughly 75% of respondents fall within cluster A.
Figure 7. Perceptions of responsibility for digital curation by research phase of respondent (during period of study).

Figure 8, which plots perceived responsibility after the period of study, shows a much wider spread of responses with two clear clusters, one of which (C, with approximately 30% of the respondents, mostly PhD holders) shows responsibility lying largely with the institution and the other (D, with approximately 60% of the respondents, mostly students and supervisors) showing responsibility remaining with the student. The fact that there are two clear schools of thought, and that the majority of respondents believe responsibility to remain with the student, is notable.

Figure 8. Perceptions of responsibility for digital curation by research phase of respondent (after period of study).
The differences in perception revealed by these two figures (and shown clearly by the boxplot quartiles and medians in Figure 9 and Figure 10) are particularly interesting in that they are somewhat unexpected. It was expected that respondents would assign greater responsibility for digital curation to each other, when in fact the converse is true. The data show a significant shift towards respondents assigning themselves more of the responsibility, not less. Over three quarters of both student groups believe that over 60% of curation responsibility during doctoral study lies with themselves, whereas PhD holders and supervisors show a significant shift towards responsibility lying with the supervisor. Based on the median response, students do not believe that institutions have any role at all to play in supporting digital curation during study. Three quarters of PhD holders and supervisors believe that under 20% of responsibility lies with the institution (see Figure 9).

![Figure 9. Responsibility data showing range, 1st and 3rd quartiles, and median (during).](image)

The picture is somewhat different for perceived responsibility after the period of study, as shown by the wider ranges and multiple outliers in Figure 10. The only clear conclusions from this data are that once again, students and supervisors each take on more of the responsibility themselves and that those people who hold doctorates but do not supervise (a group which is likely to largely work outside academia) are the only group who perceive a large role for the institution to play in long-term preservation of their data, with a correspondingly low personal responsibility. This wide range of responses may indicate a lack of clarity in understanding of who is, in fact, responsible for digital curation after the PhD ends, and contributes to the impression that data are neglected once a project is complete (Alexogiannopoulos et al., 2010).
Other entities with curation responsibilities.

In addition to the three roles above, respondents were asked to nominate others with a duty towards digital curation. On the few occasions that respondents noted other entities with responsibility they tended to be assigned under 50% of the responsibility during the period of study, but over 50% once the PhD had been completed. These included external research organisations with a direct connection to the subject of study and funding bodies (both during and after the period of study). One respondent stated that the departmental technician and the student shared equal responsibility for digital curation, both during and after the PhD. Other entities named as having responsibilities only after the period of study were library/repository staff, journals, and, interestingly, future students (although whether in the same discipline and/or the same department was not clear).

Awareness and Use of Digital Curation Resources

Several examples of existing literature acknowledge a general lack of researchers’ awareness of the digital curation support services that are on offer (e.g. Alexogiannopoulos et al., 2010) and the difference between mere awareness and embedding of services is discussed in Molloy (2012).

The approach taken in this research was to ask doctoral students and supervisors simply if they had heard of the resource and if they had used it. Figure 11 demonstrates a very low usage of all external resources across all roles. Under 10% of respondents had used dedicated resources such as the DCC, MANTRA, and VADS4R and all resources of this type had only 10-20% awareness. Surprisingly, both awareness (nearly 50%) and use (over 20%) of the policies of other institutions is considerably higher than any of the dedicated digital curation services. It should be noted that subject specific resources were excluded from the survey unless they were directly related the arts, however there was little variance between the arts-specific resources mentioned and broad resources such as the DCC.
The picture is somewhat different with sources of support closer to home. The most used resource is the informal advice of friends and colleagues (60%). This is followed by IT services at the researcher’s institution (52%) and then doctoral supervisors and advisors (44%). Whilst three quarters of the respondents were aware of institutional policies, they have been used by under 40%. This relatively low awareness and use of institutional-level support may be partly the cause of the low perceived responsibility of institutions in digital curation discussed above.

The large difference between external sources of advice compared to those provided by the home institution and informal advice from friends and colleagues very clearly demonstrates the value of direct contact when providing digital curation support to doctoral researchers. With the exception of institutional IT services, the awareness and use of sources of support correlates directly with their nearness to the researcher. These findings strongly support the assertion by Ward et al. (2011) that top-down solutions for training provision may not be as effective as those that embed themselves in the day to day activities of researchers.

Conclusions and Recommendations

Working with digital data is an intrinsic part of a great deal of doctoral research. Developing the appropriate skills and knowledge to create, access, use, manage, store and preserve data should therefore be considered an important part of any researcher’s development as they move through doctoral study developing the necessary skills for employment. However, the relatively recent emphasis on digital curation in research combined with its highly-specialised and fast-moving nature present problems for supervisors, who are generally expected to mentor their students in this topic alongside other research skills.
Ensuring Practical Digital Curation is Understood

90% of doctoral students and supervisors alike consider digital curation to be moderately to extremely important. Digital curation activities are regularly undertaken as part of doctoral research with most students creating and using data on a regular (at least monthly) basis, and performing technical and administrative actions slightly less often. However, the analysis of student digital curation activities identified several areas for concern where researchers are aware of the relevance of an activity but rarely or never do it. These concerns could be fruitfully addressed by increasing the understanding of doctoral researchers and supervisors of data curation as an iterative lifecycle (via DCC information and training for example) or, more specifically, communicating clearly to students the risks and pitfalls of omitting curation activities which are particular to their own research and the potential consequences on completion of the PhD.

Expertise and Responsibilities

All groups of respondents considered themselves largely unskilled at digital curation, yet assigned themselves some responsibility for doing it. Students were assigned the largest responsibility (particularly by themselves) and have the lowest levels of expertise. This is likely to result in a great deal of pressure on individual students and the supervisors supporting them. Supervisors could encourage students to, where appropriate, relinquish responsibility for digital curation activities to appropriate supervisory or institutional support structures. This model already works well for automatic backups run by institutions and could potentially be widened to cover other appropriate curation tasks.

Making the Most of Sources of Support

Use of digital curation support information and services within the institution ranges between 40% and 60%. Use of specific external resources is low at under 10% and awareness for all specialised external resources was under 20%. This represents a missed opportunity in terms of outsourcing as much training as possible to dedicated experts. Supervisors should aim to familiarise themselves with both local and external support in order to minimise the burden of digital curation expertise within the supervision relationship and support the student in sharing curation responsibility as suggested above. Both supervisors and students should aim to communicate their support needs to institutions in order to demonstrate demand for information and training appropriate to their needs, and institutions should commit to meeting these needs as early as possible in the doctoral study, where possible including library or repository staff in the process. For their part, external sources of support should aim to improve their discoverability within this user group but also specifically tap into the sources of support who have a more direct link with the doctoral student. In general, more emphasis should be placed on bottom-up models of sharing curation expertise.
Acknowledgements

I would like to acknowledge the advice provided by Dr Robin Burgess (Glasgow School of Art) and Sarah Jones and Laura Molloy (Digital Curation Centre) in identifying data management resources and commenting on a draft of the questionnaire. I would also like to thank Mark Chevallier for advice on some of the more complicated statistical analysis.

References


University of Exeter. (2013). *Open access and research data management policy for PGR students (Final version).* Retrieved from http://hdl.handle.net/10036/4279
