

## National Assessment Survey for the Digital Preservation Coalition

*An assessment of national need for digital preservation, carried out from August 2003 to March 2004, with a DPC workshop in November 2003*

### ***REPORT***

March 2004

Report by Duncan Simpson, consultant

#### **Acknowledgements and thanks**

*I am grateful to all members of the Digital Preservation Coalition who gave their time to helping carry out this survey project, and in particular to the DPC Secretary, Maggie Jones, for all the help and guidance she has given*

## **EXECUTIVE SUMMARY**

### **NATIONAL ASSESSMENT OF NEED – DPC SURVEY**

#### **Introduction**

1 This report has been compiled from a survey circulated in August 2003 to 27 DPC members (including the BBC, which joined in September). Forms were also sent out to Oxford and Cambridge, who are legal deposit libraries and members through CURL, a total of 29. The responses were collated in late September and October 2003. There was an excellent response from members. 26 completed survey forms were received, which is almost 90%. A workshop for DPC members was held in November 2003 which resulted in some follow-up actions, which continued into the early part of 2004.

2 The survey is part of a national assessment of need being carried out by the DPC in accordance with the objectives of its current business plan. The target has been set for the year 2003-04 to 'develop in association with members a project to carry out a UK needs survey on digital preservation'.

#### **Analysis of survey returns**

3 The survey returns, along with agreed follow-up work, offer a new and excellent source of collated information. They showed a high degree of consistency in issues identified, suggestions for action etc. All DPC members were asked to complete at least the preliminary information, so that this could be added to the DPC website. Seven members submitted partially completed returns, in line with the suggestion that the full survey was unlikely to be applicable to them.

4 Almost all respondents said (q 1.1 – 17:3, q 1.2 – 16:4) that their strategic planning documents either did, or soon would, contain reference to digital preservation. They also gave strong reasons (q 1.4 and 1.5) for their interest in digital preservation, in most cases either a statutory duty or a core business need. Against this, however, inspection of members' public websites gave (with some notable exception including, but not limited to, the British Library and the National Archives) a rather less positive picture. This suggests that though respondents organisations see themselves as giving a high profile to digital preservation, there is, as a minimum, much more to do on embedding it within the organisations and on the presentational aspects. Formal statements, through business plans and the like, should ideally align more closely with other, more public, expressions.

5 The responses to Question 2 gave excellent details of projects completed, in progress, or planned, and only need some further work to ensure that levels of detail are consistent and that non-respondents can be brought in to the picture.

6 Question 3 concerned volumes of digital material. Between half and two-thirds have made some assessment of the volumes of material they will need to preserve. Estimates are not highly specific, and some state that they have doubts as to their accuracy. External sources can be used to help give a clear sense, but

for future work it is important to be as clear as possible about volumes so some further work on this is indicated. It would also be helpful to have clearer statements of selection policy, the reasons why material is being kept for preservation.

7 Replies to question 4 give a very clear and useful perspective on the range of material respondents need to preserve.

8 Resources, both financial and human, were covered in question 5. As with question 3 on volumes, the overall picture is that respondents have started to look at likely resource need, over a five year period. None has looked further ahead. Most see problems in getting the resource they need, though none quantify it precisely. There is a strong view that digital preservation is a new activity (that is, it will not in any major sense replace current activity), and that it is not likely to generate substantial new revenue streams.

9 These two points mean that it requires new money to fund it. Also that funding will initially largely be project funding, though as things move on the need will be more for core funding to support the new, ongoing activities associated with digital preservation. As financial resource is such a pivotal issue, more work is indicated to try to clarify the size and timing of the need. There will be significant need for new staff and a pool of appropriate expertise does not exist already, so a strong training and development need is indicated.

10 Question 6 is about pressures on timing. Imminent loss of important digital material does not seem to be a major concern for members. None cite crisis-level urgency to find solutions. There may also be sectoral differences in the degree of time pressure that exists. It is very important to convey a sense, to government and funding bodies, that there is real urgency in finding solutions to digital preservation. Further work on this is indicated, to clarify the position. This should look at refining the knowledge about timescales, and logging examples of losses of digital material.

11 Question 7 looked forward to identify possible gaps in coverage and solutions to them. With a good level of consistency, respondents identified needs for:

- more 'joined up' strategic activity and partnership;
- more emphasis on the problems facing small, local and regional bodies and how to help them;
- more support activity such as a curation centre,
- infrastructure programme,
- practical projects,
- standards for content and metadata;
- involvement of commercial sectors.

12 The benefits which will help argue the case are seen to lie chiefly in cultural benefit (the existence and availability of a national record), public benefits such as better access and educational and learning advantage, business continuity

and support for important government policy such as life-long learning, inclusivity and furtherance of health policy. This moves towards a useful definition of priorities so that gaps in the planned work on, and provision for, digital preservation can be identified and action taken where necessary.

13 Thirteen recommendations for action were produced. The next process was to validate the responses and analysis of them, which was done at a DPC workshop in November 2003. The recommendations were discussed and various revisions and actions were agreed. This included proposals for some follow-up work to build on the survey returns, as part of a plan for future action, defining further objectives, leading to finding support for further work (for example through applications for grant funding support) and to completing the national assessment of need project to an agreed timetable.

14 These outcomes of the workshop were circulated by the DPC Secretary and work has since been going forward on them.

## **NATIONAL ASSESSMENT OF NEED – DPC SURVEY**

1.7 This report is in four parts:

1. Introduction: analysis of survey returns
2. General comments on the survey responses
3. Comment and analysis on responses to each question in the survey.
4. Analysis of responses to question 7, on gaps and further needs.
5. Conclusions on the survey
6. Subsequent actions

### **INTRODUCTION: ANALYSIS OF SURVEY RETURNS**

1.1 This report has been compiled from DPC Members' responses to a survey, on their digital preservation activities, on behalf of the DPC Board. The survey was circulated in August 2003 and the responses to it collated in late September and October 2003.

1.2 The survey is part of a national assessment of need being carried out by the DPC in accordance with its current business plan. One of the three strategic goals of the DPC is 'Promoting digital preservation by concerted effort to get and keep it on the agenda of key stakeholders'. To help achieve this, it is a programme objective in the DPC Business Plan to 'expand awareness of digital preservation amongst key stakeholders'. The target has therefore been set for the year 2003-04 to 'develop in association with members a project to carry out a UK needs survey on digital preservation'.

1.3 There was an excellent response from members. There are 27 members and 29 forms were sent out (which included the Oxford and Cambridge libraries, as CURL members and legal deposit libraries, as well as all members). 26 completed survey forms were received, which is a response rate of almost 90%.

1.4 The survey returns, along with agreed follow-up work, offer a new and excellent source of collated information on members' activities in digital preservation. They showed a high degree of consistency in issues identified, suggestions for action etc.

1.5 Details of the responses were prepared and submitted, and are available as numbered annexes, as follows:

[Annex 1:](#) Details of organisations responding, contact names/addresses and statement of reasons for interest in digital preservation.

[Annex 2:](#) Overall collation of responses.

[Annex 3:](#) Some selected comments on particular questions or issues, verbatim.

[Annex 4:](#) Questions 2.1, 2.2 and 2.3: members' notes of digital preservation projects completed, in progress, or planned.

[Annex 5:](#) Question 4.1: types of material being preserved.

[Annex 6:](#) Questions 7.1, 7.2, 7.3 and 7.4: suggestions for other needs and actions.

*1.6 Subsequent actions and additional work:* In addition to this report and annexes, further work arising from the survey is being, or has been, done. This includes an extended table of digital preservation projects involving members; and further work on the volumes and scope of digital preservation activity among members. An application is also in progress with the New Opportunities Fund, in conjunction with MLA (the Museums, Libraries and Archives Council, known until February 2004 as Resource) to secure funds for a sample survey of digital preservation needs at regional and local levels and among smaller organisations.

## **GENERAL COMMENTS ON THE SURVEY RESPONSES**

2.1 Because of their nature, which made certain questions irrelevant to them, we suggested that some member organisations only submit partial returns. These were the Association of Learned and Professional Society Publishers (ALPSP), the Consortium of University Research Libraries (CURL), the e-Science Core Programme (eSCP), the Publishers Association (PA), Resource and the Research Libraries Group (RLG). For the most part these consisted solely of contact details and the nature of the respondent's interest in digital preservation. The eSCP, however, also replied to questions 1 and 2, and Resource replied to parts of questions 1 and 7. The Oxford University Library Services submitted a full return as a CURL member. The Online Computer Library Center (OCLC) submitted a full and useful return, while excluding some of the questions (eg on resources) which are less relevant in view of its status as an organisation based in the USA.

2.2 Some members have a distinct position in that they have been doing digital preservation for many years. This includes the two university data archives, the United Kingdom Data Archive at the University of Essex (UKDA) and the University of London Computer Centre (ULCC). Their core business has for some years included the active practice of digital preservation. Both also offer their services to clients, on a contractual and revenue earning basis (an example is the ULCC contract with the National Archives to store and give public access to structured databases which are public records). The Arts and Humanities Data Service (AHDS) came into being more recently than the two data archives, in 1996, but its core business also centres on the storage of digital material.

2.3 The BBC, a new DPC member, has only recently begun to include digital material in its preservation programme; the stated BBC policy is not to differentiate digital from non-digital, simply to regard them as different formats. The BBC has a ten year 'digitisation for preservation' project which may strictly fall outside the DPC definition of digital preservation, for this purpose, because it does not involve 'born digital' material. However the BBC response, completed on this basis, has been included because it faces the issues of long-term preservation of the digital material thus created, and also because it will concern itself with 'born digital' preservation as this becomes an issue. The BBC has also been leading a European Union 5<sup>th</sup> Framework project to provide value-for-

money digital preservation for large amounts of at-risk audio, video and film materials. This is in conjunction with French and Italian principal partners.

2.4 Most present members of the DPC are UK organisations, though membership is intended to go beyond UK boundaries. Two current members from outside the UK are RLG and the Library of Trinity College, Dublin (TCD). RLG's response gave contact information and useful answers to questions 1 and 7, but the other questions were not relevant for its particular situation in relation to a UK based assessment of need. TCD submitted a full response and it has been included with the others. TCD's position is unusual, since it is clearly outside the UK and subject to its own national administration, but at the same time it is one of the six legal deposit libraries. If material from the survey is used for strictly UK reasons, for example, for lobbying purposes with the UK government, it would be sensible to consider whether specific information from TCD is relevant, and should be included, for the particular purpose, or whether it should not.

2.5 Most of the UK DPC membership would be able to claim some statutory duty to provide access to material under the Freedom of Information Act 2000, when it comes into force in 2005. This includes Scottish organisations, which have an FOI Act similar to the UK one, coming into force at the same time. This is not part of the intended purpose of question 1.4 b), and the FOIA stops short of creating a general duty for authorities to provide public access to material. So FOI has not been accepted as a duty when cited here. For this purpose also a similar point can be made about the Data Protection Act 1998. However, where members have cited FOI or data protection under question 1.5 (which asks about any statutory or legal issue or problem that might complicate or impede your organisation's digital preservation activities) it seems legitimate to include that reference, because compliance with the FOIA requires that organisations maintain good record keeping over time.

## **COMMENT AND ANALYSIS ON RESPONSES TO EACH QUESTION IN THE SURVEY.**

### **Question 1**

3.1.1 The purpose of questions 1.1-1.5 was to verify the extent to which digital preservation is established as a corporate issue for DPC member organisations. If it is a major concern, meriting significant attention and action including the engagement of government, and the attention of main funding bodies and others, then one would expect to see it formally flagged up in such documents as corporate and business plans and major strategy documents. Otherwise – to put it bluntly – government officials, or HE funding bodies, or whoever needs to be persuaded, could simply say, 'Your agendas are set, and your funding resolved, through your major planning documents. We see little sign there that digital preservation is a matter of high priority or great concern for you, so we will not address it, or advise Ministers to do so'.

3.1.2 On the face of it, it is indeed a recognised issue of concern. Of the 20 answers to question 1.1 (whether digital preservation figures in the corporate

plan) 17 answered 'yes' against 3 'no'. On question 1.2 (whether there is a digital preservation strategy or similar in place) 16 also answered either 'yes', or that work to create a strategy was in progress. However, the responses also indicate that many of the corporate or strategic documents cited make only partial, or tangential, references to digital preservation, and some are in subsidiary or departmental documents, rather than the main corporate ones. And almost half of the strategies are in preparation, or planning, rather than complete and available. So the coverage may not yet be quite as effective as it seems at first (see also below).

3.1.3 Digital preservation also seemed to figure commonly in corporate risk assessment and risk management procedures, question 1.3. But here, even more markedly, the references were often tangential, or not clearly defined. Risk analysis is currently a popular management tool, and a government preoccupation. Any organisation subject to direct central government control would normally be expected, either by Treasury or by its auditors, to have formal risk management process in place. So it is legitimate to ask whether digital preservation figures as a component of that process, and to attach some weight to the answer.

3.1.4 It is very clear, from questions 1.4 and 1.5 (which are about what drives the interest in and importance of digital preservation for members) that all can claim it as an issue of high importance. For many it involves a statutory duty and most see it, as well or instead, as a core business need. This is as would be expected, and gives sound basis for the need to pursue and promote the issue of digital preservation.

3.1.5 However there is a broader aspect to all of these replies, which members need to consider. I have said that replies to questions were often tangential rather than direct. As a separate exercise, to give another perspective on how highly digital preservation figured on corporate agendas, I looked at members' web pages. From these, a rather different picture emerged. It was rare, from these public sites, to get any sense that digital preservation was an issue at all, let alone a major one. Simple searches on the expression 'digital preservation' commonly got no return, and where planning documents such as corporate plans or strategies were on the websites, it was difficult to find reference to digital preservation in them. Where it could be found, the references were more often in passing than direct or emphatic.

3.1.6 There were a few notable exceptions [I should say that I did the bulk of this survey in late July/early August, in one or two instances more recently; things may have changed since]. Some examples of a relatively high profile presence: the BL website has a separate section on digital preservation, with references to relevant activities, reference to a policy document which will, when complete, be put up on the site, and a named digital preservation contact. The National Archives site has a link to a digital preservation section, with plenty of information on what is being done. The National Library of Wales has its *Digital Preservation Policy and Strategy* readily accessible on its website.



3.1.7 However, these are exceptions and as a general rule, from this exercise alone, I would have concluded that digital preservation was not a high priority issue, or perceived as a problem, at corporate level among DPC member organisations.

3.1.8 This is not to cast doubt on the accuracy of answers given to question 1. But it is to make the point that, if government, funding bodies and other major players are to be engaged with the subject of digital preservation, then it needs to be much more prominent and clearly defined on the public agendas of member organisations. And it needs to be apparent as an issue for users and the public, for example, on websites.

3.1.9 Perhaps member organisations should be asking themselves: 'If the Government Minister [or whoever] ultimately responsible for us asked, tomorrow, to see our corporate or business plans, and our main policy document(s), would he/she understand that digital preservation is a significant and high profile issue for the organisation to face, and resolve, now rather than in the future, and why, and what the scale of the issue is?'

3.1.10 At the same time, I appreciate that it would probably be counter-productive to promote alarmism through extreme references to digital preservation. But I suggest there needs to be a greater consistency about references to it, and it needs to figure with the right degree of prominence in all the formal means which exist for organisations to signal their objectives, their priorities, an appropriate level of concern, and a quantification of need. At the moment, that does not seem to be the case, in most instances.

3.1.11 I suggest also that it would be a good idea for all members to note their membership of the DPC on their websites.

#### **Recommendation 1:**

That DPC members should ensure that their corporate, planning and strategic documents include clear reference to digital preservation and its importance, and that their public statements, on websites and elsewhere, are consistent with this.

#### **Recommendation 2:**

That DPC members should note their membership of the DPC on their websites, displaying the DPC logo if possible.

#### **Question 2:**

3.2.1 Questions 2.1-2.3 were intended to produce an overall picture of current and planned practical activity in tackling the issue of digital preservation. The result is a clear list from respondents of completed projects, projects in progress and those planned. Coverage in the responses was excellent. Levels of detail, though, differed, and it might be sensible for the DPC to initiate some follow-up activity to clarify the scale and extent of activity to the same level of detail for all respondents, and perhaps also to complete the picture by approaching DPC members who were unable to respond to the survey.

3.2.2 Details of the information given in reply to these three questions is set out, verbatim as supplied (except where minor editing was necessary for clarity, away from the context of the overall responses) in [Annex 4](#).

### **Recommendation 3:**

That the DPC carry out follow-up work with members on the projects list, to ensure consistency (eg level of detail) and to fill gaps from members who did not complete the survey. This to be done by phone, e-mail, or meeting as appropriate.

### **Question 3:**

3.3.1 Questions 3.1-3.4 asked about volumes of digital material which now, or in the future, would require preservation activity. The aims were to find out what stage members were at, generally, in assessing the size of the task facing them, and to find out whether, and how, they can control (by selection) the flow of digital material for preservation.

3.3.2 The general picture is that half of those responding have carried out some form of survey of digital material already held, while more like two-thirds have assessed what volumes they are likely to need to preserve in the future and what the rate of growth in intake will be. It may seem inconsistent to have, apparently, assessed present volumes less systematically than expected future ones. More likely, though, it reflects the situation that only a minority of respondents at present hold a really significant amount of 'born-digital' material for preservation, though all are conscious that they will soon do so ('soon' meaning over a matter of some years). There are exceptions – the two long-established university based data archives, and the Council for the Central Laboratory of the Research Councils (CCLRC) – but these are organisations for which holding digital material has been core business for some years.

3.3.3 Where volumes have been assessed there seems (question 3.3) to be a general sense that the estimates are not very accurate. Only a minority of responses actually quantified the volumes. Again, there are exceptions – CCLRC, for example, holds a large volume of material, predicts rapid growth and gives precise figures. But the conclusion seems to be that, by and large, we do not have precise details of volumes of digital material either for the near or medium term future. Whether this is because it is not possible to calculate, or because the calculation has not been done, is an open question.

3.3.4 This is an important point. In order to convince government, or funding bodies, that there is an issue to be dealt with, it is essential to be able to indicate the scale of the problem. There are dangers here. First, in inaccuracy, in wrongly assessing the volumes of digital material which need to be preserved, whether through under- or over-estimate. Second, in appearing to be trying to create alarm by citing very large volumes. But both seem to me lesser dangers than of being unable to indicate – given that digital preservation is a serious issue which needs policy and strategic attention at very high levels – what the extent of the need is likely to be. Further work on this is indicated.

3.3.5 Some of this may be gathered from other, existing sources, though it is often tangential rather than direct. For example, *The Data Deluge: An e-Science Perspective*, by Tony Hey and Anne Trefethen, notes that that a large novel requires 1 Mbyte of storage, the Bible 5 Mbytes. These are relatively modest amounts. Moving up, a Mozart Symphony (compressed) requires 10 Mbytes and a digital movie (compressed) 10 Gbytes. Even higher (and more directly relevant), the annual production of refereed journal literature currently amounts to around 20,000 journals, containing 2 million articles and needing 1 Tbyte of storage. Against this, the total digital storage size of the Library of Congress seems surprisingly modest at 20 Tbytes. The estimate from the same source of the current total annual production of information (print, film, optical & magnetic media), though, is a massive 1500 Pbytes. The 2002 EPS Ltd report *The impact of the extension of legal deposit to non-print publications* (Electronic Publishing Services Ltd., October 2002) contains a good assessment of likely library volumes. For example, it forecasts a 'massive' increase in online publishing, including a near quadrupling (from 52,000 to 193,000) in the number of electronic journal issues in the UK between 2002 and 2005.

3.3.6 Figures such as these give some indication of potential scale, though still not very precise. They can be frightening in terms of the potential costs they might generate for long-term digital preservation of the material concerned. They need to be as accurate as possible so further work is indicated. This should initially be desk research using existing sources, but may also need to include further contact with certain members to get additional information from them.

3.3.7 These revised figures should also include information on likely costs, where this can be found (see also **question 5** below).

3.3.8 The estimates need to be comprehensible to senior civil servants and government ministers. This might be best achieved by relating estimates to understood concepts, such as the total holdings of the BL or of national archives.

3.3.9 There may be a further risk here. Many organisations can be selective about what they take for permanent preservation. Copyright libraries, except for the BL, are not, under present law, obliged to take every publication. Archives, as a matter of course, select down to a very small percentage of the total record created (typically, under the various Public Record Acts, fewer than 5% ever make it to the archive). The percentage they actually keep is not governed by statute. While it may be desirable for research institutions to keep everything, there is normally no statutory obligation on them to do so. Faced with a difficult problem – how to deal with large volumes of digital material needing costly long-term preservation – a typical government first reaction would include the questions: 'Why are we keeping all this stuff? Do we have to? Do we really need to?'. To counteract these questions – and also to make sure that organisations are not spending scarce resources on keeping things they do not really need – it is important to have a clear, defensible selection policy.

3.3.10 A majority of respondents say they have looked at the issue of selection policy (question 3.4). But my conclusion overall is that it is not yet as clear as it

might be, why and on what basis organisations are selecting material to keep. It is perhaps an area – like the matter of corporate policy and strategy examined in question 1 – where more overt statements of policy might help.

**Recommendation 4:**

That the DPC carry out follow-up work to define more clearly what volumes of digital material for long-term preservation are likely to accrue over the next 5-10 years. This to be done initially by use of external existing sources, looking more widely than DPC members. Then as necessary by working with members by phone, e-mail, or meeting as appropriate.

**Recommendation 5:**

Members should give more consideration to their selection policies for the preservation of digital materials, and to ensuring that lifecycle management is made explicit in them =, reinforcing the fact that long-term maintenance is most cost-effectively handled at ingest.

*[NB this recommendation is given in the amended form that was agreed at the November 2003 workshop]*

**Question 4:**

3.4.1 Question 4.1 asks respondents to define what types of digital materials they will be preserving. Answers were full and are set out verbatim in [annex 5](#).

3.4.2 The main purpose of this question is to be able to say, with authority, what the range of material is which will in future require preservation. This is all part of the process of definition and scoping also covered in question 3.

**Question 5:**

3.5.1 Questions 5.1 – 5.9 are about the crucial issue of resources, both financial and human. 5.1 – 5.4 deal with getting financial resources for digital preservation. They seem to show that, generally, the search for new resource to cope with digital preservation is still in its infancy. The general perception (and it might be worth exploring a bit further the reasons why a few respondents do not share this view) is that finding funding is going to be a big issue over the next five years or so. The majority of respondents have at least begun to assess what their needs are going to be over this period. Noone has as yet tried to look further ahead than that. This is of a piece with the general uncertainty about what volumes of material are likely to be accruing in the longer term. It also ties in with the normal timescales for bidding for funding in the public sector (but see also below on timescales).

3.5.2 The straightforward interpretation of this information is that organisations are starting, sensibly, to see what tasks and what needs face them over the next five years – a perfectly normal scope for a planning period in government, educational, research or cultural organisations – and are clearly nervous of what they are starting to find. Some are further down the road than others, some have barely started the task, but most have started to move.

3.5.3 There may be a risk here for DPC and its efforts to raise the profile of digital preservation. Another possible interpretation of this information is that member organisations cannot say for certain that there is, or is going to be, a resource problem here, or what the size of it is. Governments, in particular, tend to think in the short term and to decline to deal with problems which appear neither urgent nor quantified. It is important therefore to be sure that there is no such pitfall here.

3.5.4 Question 5.5 is interesting. The stock response from government officials to an issue such as digital preservation is likely to be that, however new and radical it may be, it is only a technically different way of doing what most DPC members already do, and have been doing for years. If, for example, publishing moves from being chiefly a paper based operation to chiefly a digital one, then you can cover at least a large part of your new, digital costs from the funds you no longer need for dealing with paper publications. Question 5.5 presents a strong and generally consistent view that this is a false argument, and that for most practical purposes, and for the foreseeable future, digital preservation is not going to be 'instead of' traditional preservation, but 'as well as'. This is going to be a crucial and difficult area of debate when tackling funding sources for new funds, so it is vital that the arguments for it are sound and acceptable. Some of the comments from respondents on this issue appear in [annex 3](#), and it is important to ensure that they will stand up under examination.

3.5.5 Question 5.6 contains no surprises; most responses see the need for both project and core funding, with an emphasis initially on the former, moving over time (as digital preservation activities and programmes become more embedded in the organisations' operations and functions) to the latter. This is as one might have expected, but it is useful to have confirmation of it. However, we need to manage the risk that it will be difficult to make the move, when it is needed, from procuring one-off 'project' expenses, to requiring continuing, year-on-year core funding.

3.5.6 The answers to 5.7, also, are unsurprising. Most respondents see the likely source for funds as being their usual primary funders, whether that is ultimately government (as it usually is for DPC members) or other. While it may seem disappointing that there are few proposals for more original or different funding sources, it would be unrealistic to think that there might be some magic solution waiting in the wings.

3.5.7 It is disappointing, if not surprising, that there seem to be few possibilities for developing significant new revenue streams from the digital material taken in for preservation, as emerges from replies to question 5.8. The numbers break down fairly evenly yes/no, but even the 'yes' replies do not see substantial revenues being developed, on a scale to offset the development costs of digital preservation processes and systems. A question such as this is always asked by government when it is approached for new money. So, as with question 5.5, we need to be sure that the arguments people have put forward as to why this is so (and again, some of their comments are set out in [annex 3](#)) will stand up under scrutiny.

3.5.8 Finally, question 5.9 asks about the staff resource for dealing with digital preservation. There are clearly issues here which need further work, and the most obvious (apart from funding) is training and development.

3.5.9 Even if you have the money, you can buy in trained staff only if they exist. There may be two risks. The first is simply not being able to find the right skills when they are needed. The second is that a 'bidding war' develops, even (or especially) between allied organisations (as happened, for example, with IT staff when every organisation was trying to deal with the expected 'millennium bug' problem in 1998/1999). This, if it happens, can soak up large proportions of what little cash resource is available to pay specialist staff, without really making things better. Some respondents talk about the need to develop their own expertise in house, which is a positive response to the problem. But it will not avoid the 'bidding war' problem unless it is done by everyone.

3.5.10 In order to alleviate the expressed anxieties about shortages of trained staff, and to help in developing staff in post, it would help if some centralized resource were available, to identify and give training in the right skills. Clearly also, at some stage, the relevant teaching organisations (for example, professional preservation, library and archives qualifications) will need to encompass new digital preservation knowledge, skills and techniques at the right level for those training to enter the professions. The issue is where to find the impetus to get things moving, and whether this can be a role for the DPC or whether someone else might take it on.

*[Nb recommendation 6 also includes the question of urgency of timing, which is discussed under **question 6**]*

#### **Recommendation 6:**

That the DPC carry out follow-up work with members to clarify the scale, likely cost and timing pressures of their digital preservation activity, to see if these can be more clearly defined, and if not, why not. This to be done by phone, e-mail, or meeting as appropriate.

#### **Recommendation 7:**

The DPC should develop an ITT for a consultancy to develop a training package and workshop, possibly using the Cornell NEH funded online tutorial and workshop as a model.

#### **Question 6:**

3.6.1 Question 6 asks whether there are any specific timing pressures facing members. The point is to consider whether there are timing priorities, or pressing reasons why some sectors, or types of material, need faster solutions than others. Several organisations refer to pressing business plan targets or similar. These are valid reasons for action but, because they are internally set, fall short of establishing priorities that government would recognise. The most cited external factor seems to be the impact of the legal deposit legislation, which bears on a number of members. Archives also refer to possible difficulties over legislation, and a public consultation paper on this has been issued by the

National Archives. Nobody cites imminent problems with the safety of material, which is on the face of it comforting. The BBC does refer to possible degradation as a driver for action, but no one actually refers to loss of material as an immediate issue.

3.6.2 Sectoral differences might also be significant here. Archives point out that their accessioning is done on a fairly long timescale (the 'thirty year rule'). So they will, for some years yet, be dealing largely with the paper records of the last three decades. They will, though, also need to deal with the issue of newer, digital material, before the thirty year watershed, as much of it may have degraded or become inaccessible before that point is reached. However, the point remains that there may be more immediate urgency for libraries, universities and research institutions to deal with current material, than for archives.

3.6.3 DPC members need to acknowledge and take steps to avoid the risk of a perception that action of the right quality and scale can be deferred for any significant period of time. This is especially true, given the timescales involved. Government funding works in three year cycles mostly, so it can take at least that long - often longer - for a bid for funding to work its way through, even if successful. Add to that the time then required to get a substantial project under way, and it is clear that, even if starting now (or already started) it could easily be five or six years before the necessary systems and processes are in place – and that assumes that all goes reasonably well.

3.6.4 What does this all mean? DPC members share the view that there is real urgency in finding solutions for digital preservation. But the evidence seems to show that they are not at present actually losing important digital material, or even close to doing so. Coming back to the theme of how government and funders may react in an unhelpful way to such information, there may be a risk of someone using this to conclude, however mistakenly, that there is, for the time being, no real pressing urgency for action.

3.6.5 Another possibility is that there may be an unwillingness, among organisations holding digital material, to acknowledge that they have actually already lost important digital material, whether through neglect or lack of skills and resource.

3.6.6 I suggest that, to counteract any impression that there is less urgency than DPC members know to be the case, it might help to build a dossier of information on losses. Quotable instances, even anecdotal, of such loss would be useful, provided they are reasonably reliable. The best solution, in view of the sensitivity, might be to look for overseas examples.

3.6.7 Several members cite the 2005 full implementation of the Freedom of Information Act 2000 as a driver. That, taken in conjunction with the UK government's 2005 electronic service delivery targets (which I believe are mirrored in the Scottish Executive) is certainly an extra argument that government will recognise, as to why speed is important in the short term.

*[NB the important issue of timing and urgency has already been included as part of recommendation 5]*

### **Recommendation 8:**

That the DPC considers establishing a log of examples of loss of digital material in order to build up a clearer picture of vulnerability and to assist in strengthening the case for action.

### **Question 7:**

3.7.1 Questions 7.1 – 7.4 ask respondents to flag up any missing issues which need attention, and any broader benefits that digital preservation programmes may bring in their wake. Responses were good, with several consistent themes appearing. They have been noted for each organisation in [annex 6](#), and are discussed fully in the third section of this report, immediately below.

### **ANALYSIS OF RESPONSES TO QUESTION 7, ON GAPS AND FURTHER NEEDS.**

4.1 Question 7 was designed to identify members' perceptions of gaps in present activity of provision for digital preservation. This included a question on whether members had any suggestions concerning the needs of non-DPC members.

4.2 All replies to the four questions are recorded, verbatim, in [annex 6](#).

4.3 There was a reasonably high degree of consistency in the main topics which were raised. This section of the report attempts to brigade the suggestions, and summarise them as a series of recommendations, for further discussion by DPC members. The main opportunity for this will be the seminar on 7<sup>th</sup> November.

4.4 Under each question, I have listed what seem to me to be the issues, divided into 'key' and 'subsidiary'. This is a subjective judgement at this stage. I have then produced a list of draft recommendations for members to discuss and ratify, or amend, at the seminar.

### **4.5 Question 7.1: Are there any national priorities or actions which you think are needed, which are not being actioned?**

*Key issues:*

'Joined up' national strategic planning, for example, working towards a distributed digital archive with good access.

Continuing/developing a high level advocacy campaign, to influence thinking on the importance of digital preservation and the need to implement solutions.

A digital curation centre [as first part of a UK national preservation infrastructure programme - cf NDIIPP]

A national register of digital materials, to aid access and avoid duplication

[NB as things are already being done to achieve the two issues above, I have not suggested any additional recommendations]



*Subsidiary issues:*

Continuing emphasis on practical implementation of digital preservation, in addition to awareness raising and research.

**Recommendation 9:**

That the national assessment of need project include a strategic overview, a review of existing and emerging infrastructure developments (e.g. the Digital Curation Centre) and also an evaluation of gaps needing to be filled.

**Recommendation 10:**

That a tightly focused and targeted advocacy campaign, accompanied by appropriate publicity material, be undertaken to draw the attention of key stakeholders to the emerging results of the assessment exercise.

**4.6 Question 7.2: Do you have any suggestions, or views on priorities, for digital preservation needs affecting non-DPC members?***Key issue:*

Support/enabling for digital preservation at regional/local/small organisation level eg guidance, advice, software tools, services, long-term care, skills etc. (NB this might be a role for the digital curation centre).

*Subsidiary issue:*

Emphasising benefits of digital preservation in terms, eg, of potential re-use of material so that it offers cost and value benefits and appeals to small organisations.

**Recommendation 11:**

That further work, including a sample survey of small, regional and local organisations be undertaken to assist in identifying what digital material they currently hold and what their needs are.

**4.7 Question 7.3: Is there anything, other than increased financial or staff resource, which would help you to meet your digital preservation requirements?***Key issues:*

Strategic partnerships at national level, including sharing of information and solutions.

Support/interest/involvement of the relevant commercial sectors (IT, publisher etc) for the objectives of digital preservation

*Subsidiary issues:*

Information on efficacy and costs of mass storage systems.

Definition of minimum metadata standards for the creators of electronic materials

A shared tools/file format repository

Standards for content and metadata.

[It is already one of the strategic goals of the DPC to foster strategic alliances. It includes players from the commercial sector in its membership and has arranged events designed to bring together all sectors with an interest in digital preservation. It is also part of its operational goals to promote and develop standards, and to produce and disseminate information on all aspects of digital preservation. Other activities are helping achieve these, so the issue is how they can be further helped by the assessment of need project]

#### **Recommendation 12:**

That the advocacy campaign (recommendation 10) includes targeting of relevant parts of the commercial sector, with a view to engaging them more closely in further work.

#### **4.8 Question 7.4: Can you specify any broader benefits arising from your digital preservation programme?**

*Key issues:*

Facilitation of general on-line access [but, must be sold as a general good, not an elitist one].

Continuing (ie permanent) existence of, and easy/wide access to, the national published and non-published (records, images etc) archive.

Facilitating government policy eg e-government, (lifelong) education, health, social inclusivity, better/ more democratic debate eg on health and other issues.

Removing barriers/improving access to information for all.

Public benefits arising from research, learning and teaching.

Assurance of long-term business continuity.

*Subsidiary issue:*

More effective implementation of new legislation eg FOI, Data Protection, Environmental Impact Regulations.

#### **Recommendation 13:**

That the national assessment of need project develops and includes in its outcomes a clear statement of the benefits digital preservation can bring.

### **CONCLUSIONS ON THE SURVEY**

5.1 DPC members' responses to the survey were excellent in terms of number and quality and a great deal of useful information has been assembled and collated. It will bring benefits, not just in the national assessment project, but in other work.

5.2 Overall, we can take comfort from the evidence of really substantial work and projects which are dealing with the huge challenge that digital preservation represents, both for members and for others. Many positive ideas have been brought forward for discussion, initially at the November seminar, for future action

and improvements. One of the real additional benefits of the survey results is that it begins also to broaden out thinking, and open up issues which are of concern to non-members, and to suggest some ways in which these might also be tackled.

5.3 The main problem emerging is that there is still a lack of quantification, over likely volumes of digital material requiring long-term preservation, the resource need, and the likely timing of that need. It is very desirable to do some more work to try to pin these issues down as accurately as we can.

5.4 There are therefore two main strands of future work emerging. The first is to refine the survey results, chiefly by going to members to clarify and expand some of the emerging results. This need not be too onerous, as it will be selective and done largely by face-to-face meeting or e-mail or phone conversations. The second is to begin to draw up a programme of future work and actions to progress the national assessment of need project.

## **SUBSEQUENT ACTIONS**

6.1 The recommendations in this report on the survey of members were incorporated into presentations at a DPC members workshop on 4th November 2003. Members were offered twelve recommendations based on the findings in this report, and divided into three categories:

- Recommendations requiring action by DPC members
- Recommendations relating to enriching the data gathered from the survey
- Recommendations suggesting new initiatives mostly requiring additional funding

6.2 The recommendations were agreed, some in slightly amended form. The key recommendations for future work were those in the third category. From those a plan of future work was evolved. The workshop also endorsed an outline implementation plan and timetable for completing the national assessment project by the end of 2004. A summary of the outcomes of the workshop was circulated by the secretary of DPC, and work has continued since on pursuing the agreed outcomes. In terms of the substance of this part of the overall project and this report, this has mostly been to do with refining the material gathered from the survey and devising ways to make best use of it both for DPC members and others, and in pursuit of the overall objectives of the project.

6.3 Additional documents have been produced as a result of further work, and submitted to the DPC for use on its website and for other purposes related to the national assessment project.

Duncan Simpson

March 2004