Project Number	IST-2006-033789
Project Title	Planets
Title of Deliverable	Report on academic research practices
Deliverable Number	DT7-D6
Contributing Sub-project and Work-package	DT/7
Deliverable Dissemination Level	Internal PU
Deliverable Nature	Report
Contractual Delivery Date	30th January 2010
Actual Delivery Date	
Author(s)	HATII, SB, NANETH

Abstract

DT7-D6 report on a series of qualitative interviews conducted with academic/research staff at the University of Glasgow in October, 2009.

Keyword list

Interviews, communication, research, academia, preservation.

Contributors

Person	Role	Partner	Contribution
John W. Pattenden-Fail	Participant	HATII	Questionnaire design, conduction of interviews, analysis, authoring of report
Annette Balle Sørensen	WP leader	SB	Questionnaire design
Filip Kruse	WP leader	SB	Questionnaire design
Jørn Thøgersen	Participant	SB	Questionnaire design
Laura Molloy	Participant	HATII	Questionnaire design, assistance in procuring interviews, analysis
Bart Ballaux	Participant	NANETH	Questionnaire design

Document Approval

Person	Role	Partner
Annette Balle Sørensen	WP leader	SB
Filip Kruse	WP leader	SB
Jane Humphreys	SP leader	BL
Adam Farquhar	Programme Director	BL

Distribution

Person	Role	Partner

Revision History

Issue	Author	Date	Description
0.1	John W. Pattenden-Fail	11/01/2010	Initial draft version for internal review
0.2	Bart Ballaux	20/01/2010	Corrected version
0.3	Annette Balle Sørensen, Filip Kruse	21/01/2010	Commented version for further alterations
1.0.5	Jane Humphreys, John W. Pattenden -Fail	25/0108/02/2010	Commented version for further alterations Final version, submitted for review
0.9	Seamus Ross	10/02/2010	Commented version for further alterations
1.0	John W. Pattenden-Fail	11/02/2010	Final version, submitted for review

References

Ref.	Document	Date	Details and Version
DT/7-D1	Revised activity plan for DT/7	April 17, 2008	Final version
DT/7-D3	Report on User Field Studies	June 16, 2008	Final version
DT/7-D4	Report based on DT/7 questionnaire	April 22, 2009	Final version

EXECUTIVE SUMMARY

The transition from analogue to digital means of communicating and sharing information has brought with it a new set of preservation challenges. Planets has carried out a series of field studies in academic and government settings to answer the questions: What constitutes a preservable item? How easy or difficult is it to preserve digital content? Who is responsible?

As part of this, HATII has conducted a series of 18 interviews with academic staff at the University of Glasgow. The studies aimed to explore attitudes towards research communications and information-sharing, research outputs, and their preservation. The interviews ranged from approximately 30-40 minutes to one hour in length and were structured to stimulate conversation and encourage anecdotes.

The views that emerged are diverse, though some patterns emerged across disciplines. The respondents have established distinctive attitudes relating to the value of outputs and preservation, which are influenced by personal and departmental values and agendas rather than by demographic factors such as age or length of service.

Email plays a central role in the work of all researchers interviewed, though there are diverging opinions about the aspects of e-mail that are worth preserving and the blanket need to preserve all of it. Researchers in general value their time and are sceptical of too much preservation; while there may be valuable information in emails or meetings, they are generally confident that the meaningful information will emerge elsewhere. This is especially true of researchers in scientific disciplines.

Researchers are concerned with accessibility of their outputs in the long-term. Traditional results such as journals and book chapters are of particular value to researchers in arts and humanities, social and health sciences. Raw data is of particular value to researchers in science disciplines. Databases and datasets, especially when published on the internet are valued by all disciplines that produce them. Ephemeral events – conferences, seminars, meetings, or even telephone calls – are rarely placed in formal preservation schemes, and the value of doing so is questionable.

The preservation of working papers, notes, or other intermediate materials may be desirable or useful to others. Opinion is divided by discipline over whom should gain access to intermediate results and more generally about whom should be responsible for their preservation.

Few are concerned that elements of their published work may become lost, although there is less confidence about work that is born digital. In recent years, the standardisation of formats such as PDF has led to a vast decrease in problems related to accessing older files. Most access problems now are due to rights or legal reasons instead of obsolete file formats or decaying media.

Finally, scientific practitioners are generally more comfortable with digital resources, and much less governed by academic orthodoxies. Researchers in the arts and humanities are grateful for the ease of use that digital materials provide, and slowly moving into a completely digital age.

Social networking, blogs, and other features of modern web-based life are beginning to creep into the research environment. The need for at least some preservation of these materials is generally agreed upon, but there is no consensus about whose responsibility it is.

TABLE OF CONTENTS

Appendix A ; Planets Interview questionnaire

1. Introduction

Any digital preservation system must reflect the goals and values of its end-users. The use of digital media in contemporary academia for research-related communication and to share the results of academic research presents new scenarios in relation to long-term access, assessment and preservation. As academic environments become more digitally-based, even entirely so in some cases, the range of attitudes towards these scenarios becomes difficult to ignore. The inevitable differences in backgrounds result in diverse viewpoints that may often contradict and challenge each other. When designing systems and policies for digital preservation, it's important to have an awareness of this multiplicity.

The Planets project's Dissemination and Take-up subproject has sought to survey these various attitudes towards the use of digital media for communication and disseminating research and preservation of the results. This report details the qualitative study conducted at the University of Glasgow, home of Planets partner HATII. It follows a quantitative study conducted at Århus University. A parallel qualitative survey has been carried out with civil servants in the government of the Netherlands.

Spanning eighteen researchers across a variety of departments, the Glasgow interviews investigated topics directly and indirectly related to digital preservation. The interviews were conversational, with open questions designed to stimulate discourse and extract anecdotes from the vast experience of the researchers.

The opinions expressed, as chronicled in this report, provide an insight into the minds of University researchers who may not always be thinking about preservation as an end-result. These viewpoints will hopefully prove beneficial to those seeking the perspectives of end-users in developing and executing policies and managing information so it can be accessed in the future.

2. Methodology

The DT/7 User Field Studies work package began with a series of focused, qualitative interviews conducted by the three partner organisations (HATII, NANETH and the SB). This first iteration comprised of probes and revealed the following findings:

- Academics work in personal environments that encourage creativity and individuality

- Digital communications are vital to researchers' work supported by face-to-face and telephone communications
- E-mail is pivotal to researchers' communications. Professional networks are also vital to the initiation and development of ideas and substantially enabled by e-mail.
- Researchers consider it important to preserve the intermediate results and data as well as final results.

- There are few procedures to determine how work should be carried out and preserved. Where standards are defined, they are not necessarily enforced.

- Approaches to preservation are individualistic.

- Researchers have personal and differing opinions about what to preserve and who is responsible.
- Researchers are responsive to using standardised sets of tools to preserve the results of their work.

After the first iteration of the study, a quantitative follow-up questionnaire was launched to the entire research communities of Glasgow and Århus universities. The Århus study proved successful, surveying 2800 researchers and yielding over 400 responses. The results are detailed in deliverable DT7-D4. The University of Glasgow questionnaire had a low number of respondents. The follow-up qualitative interviews, detailed in this report, were designed as an alternative to the questionnaire. They aimed to take the results of the Århus study and explore them in greater depth.

Subjects were selected to span as wide a range of research roles as possible and to incorporate researchers from different stages in their careers. Respondents were split evenly between arts and humanities, health sciences, social sciences and natural sciences and were a mix of PhD students, associate lecturers, and full tenured professors.

The interviews followed a questionnaire (see Appendix A). However, the interviewer encouraged the respondents to speak freely about any related topics they felt were relevant. The questions, though always asked, were used as guidelines to provoke responses rather than as a rigid proforma. Topics of particular relevance to a researcher were investigated more deeply, while topics that did not apply to one's situation were discussed quickly. Because the interviews were designed more as conversations than interrogations, the respondents were made to feel comfortable and open with their opinions. Full anonymity was guaranteed to reassure everyone that they could speak freely about their institution and/or research environment.

Most interviews lasted for a half-hour, though some went on for as long as an hour. All interviews were recorded, and later transcribed by the HATII research team, where the most interesting statements were highlighted. This report attempts to construct a structure around the diversity of opinions presented by these 18 researchers. A quantitative weighting was attempted for some issues, but many researchers, due to the conversational nature of the discussion, did not always assert a clearly attributable opinion. In other cases, they expressed an understanding for both sides of an issue, which also made quantitative scoring difficult. Despite this, some clear trends emerged, which sometimes related to demographic background such as research discipline and age. (as described in Section 4).

3. Demographics

The original goal was to target 20 respondents with as much balance and diversity as possible. Due to the timing of the interviews (October 2009, at the beginning of the academic year), the response rate was slightly below the target of 20. Twenty were scheduled but one was cancelled last-minute and another interviewee did not turn up, leaving eighteen.

Because of the small sample size and the qualitative nature of the project, it was not the intention of the designers to attempt any statistical analysis. Quotations or comments found in this report are those of the individual, and not intended to be representative of all practitioners in their discipline or all researchers of their age range. Taken individually and as a whole, they may well be indicative of general thinking. Demographic information is presented here primarily to show the overall picture of the cohort.

3.1   Academic discipline

The Århus study, as described in 'Report based on DT7 Questionnaire'[1], was slightly biased in favour of the sciences.

Pie chart legend:
- Natural Sciences
- Health Sciences
- Arts and Humanities
- Social Sciences

---
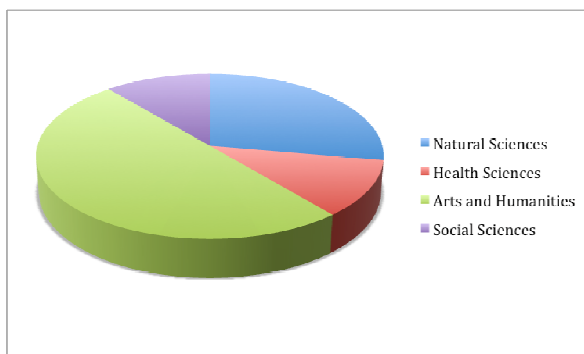
[1] Sörensen, A; Kruse, F; Thøgersen, J.: *Report based on DT7 Questionnaire* ,22 June 2009, [www.planets-project.eu]

Of the nine respondents we have categorised into arts and humanities, four come from History of Art or Music and the rest come from English literature or English language. These could likely be subdivided into separate arts and humanities categories but we have chosen to keep our classifications consistent with the Århus study detailed in the previous report.

3.2   Age of respondents
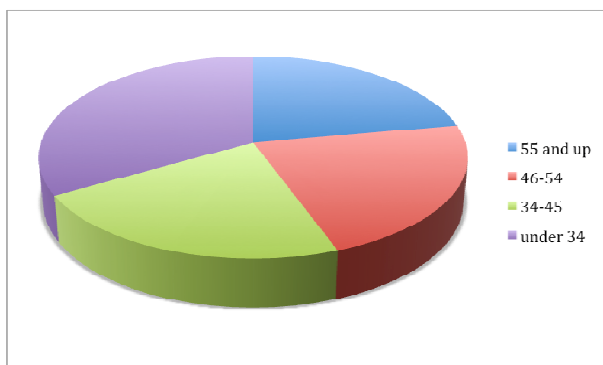
The average age of the respondents was 42.83.

3.3   Years of experience
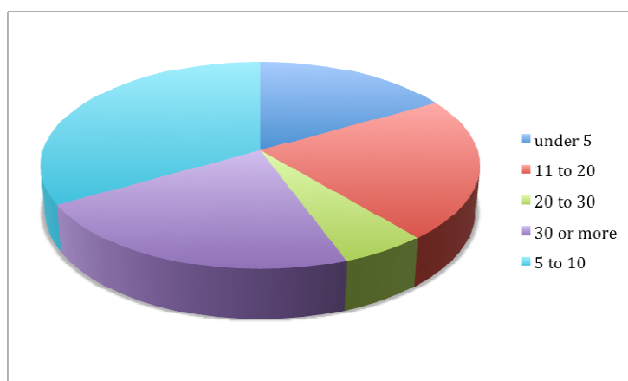
The average number of years of experience in research was 16. There was some ambiguity about when "research experience" began; most began counting after they finished their PhD, while one respondent was still completing her PhD and another had just completed it. One respondent was clearly counting undergraduate work as research experience, but the given figure was not adjusted. Additional questions were asked regarding how the researchers split their time between teaching, research and administrative work (as well as their ideal balance). This question, as well as questions about which software packages were used on a daily basis, were to familiarise the interviewer with the interviewee's work environment at the beginning of the interview As these results varied wildly based on the each respondent's role within the University, so no statistical summaries have been applied. (Appendix A).

4. Findings

The interview was broken into four sections: demographic information (covered above), research communications, research work/outputs, and a miscellaneous section about the digital/analogue divide and non-traditional information resources. The interview questions are available in their entirety as an appendix to this report.

Though the conversations roamed freely, an attempt has been made to present the findings in a cohesive structure. If there is one conclusive fact to report from these interviews, it is that the 18 respondents had distinctive points of view. The attitudes covered in this section are varied and

diverse; for any one opinion, there is almost certainly a disagreeing view present in another interview. While these respondents were treated as representatives of different scholarly subjects, even within the same department, opposing views were often expressed.

4.1   Key Findings

Detailed descriptions of these findings are contained in the three sub-sections 4.2 – 4.4 to follow.

- Email is very important to researchers in all disciplines. Many feel that preservation of some email is important. However, there is no consensus on what aspects of an e-mail should be preserved and how much preservation would be beneficial.

- Ephemeral events such as telephone calls or face-to-face meetings play distinct roles in the research flow, but there is little demand for recording or transcribing these.

- Researchers generally develop their own archiving systems for research-related communication and research outputs and rely in varying degrees on systems provided by the University.

- The research outputs that are most beneficial for others to use, re-use and adapt are databases, datasets and raw data. Traditional written/published research outputs (books, journal articles) are valuable as well, and for some disciplines the primary forms of output.

- While there is a general confidence that traditional, published research outputs are being looked after, there is no consensus on whose responsibility this is.

- Producers of datasets and databases are generally more concerned about preservation than researchers who do not create these materials.

- There is little concern that published research outputs will become lost. However, there is less confidence about the longevity of results published digitally.

- There is no consensus about the value of preserving and sharing working papers, notes, or other intermediate materials. Some researchers feel very strongly that these materials are private, while others see great benefits in exchanging them with others. There is also disagreement about who is responsible for preserving them.

- Scientific practitioners are generally more comfortable with publishing, re-using and citing digital resources, and much less governed by the academic orthodoxies. Researchers in the arts and humanities are grateful for the ease of use that digital materials provide, and slowly moving into a completely digital age. However, they are unwilling to cite online sources such as blogs, personal web-sites and wikis.

- Researchers reported few problems in accessing digital information due to age or obsolescence; old, potentially decaying digital media is generally stored by the individual, with little concern for its long-term accessibility.

- Social networking sites, blogs, and Wikipedia are beginning to impact the research flow. Researchers in scientific disciplines are more willing to refer to and cite online sources. The need for at least some preservation of these materials is generally agreed upon, but there is no clear sense of who should be responsible for this.

4.2   Attitudes towards communications preservation

4.2.1   Email

One of the few consensus opinions expressed throughout these interviews was regarding the significance of email to research. Although it sounds like an obvious statement in this modern age,

email is vital to today's research communication. Almost all of the respondents indicated that email is integral to their work, usually describing its role as "vital" or "essential". Email is used as both a tool for administrative business as well as the exchange of ideas.

Two younger respondents – a PhD student in English literature and a post-doc in Music – both expressed a strong reliance on email. The PhD student saw email as not only important to her work, but as the overall *most* important element of her communications to support all aspects of research. She used email to exchange ideas with other researchers, organise meetings and interviews, and also to discuss her own work with her supervisors. The musician used email more as a means for organisation and administration, rather than a method for exchanging research ideas, but still relied heavily upon it. This was attributed to the individual nature of composing as well as the structure of music as an academic discipline. A second composer also felt that his research process was fairly individual and personal. "Email is the first port of call. After that it's more human, face-to-face," he said.

The most notable exception to the importance of email was a 66-year old sociologist (with 42 years of experience in research), who only found email to be "a bit" significant to her work. She felt that email was helpful only to keep up-to-date with mass announcements about new periodicals and books. She was quick to attribute this attitude to her own age and background, saying "I think you are marked by the mode of communication that was prevalent when you were a student, say. Which in my case is this, - the printed non-fiction book."

Interviewees were asked whether e-mails should be preserved (question B1). While many respondents were enthusiastic about preserving their emails, most had reservations about preserving it all.

Researchers with a 'hard' science background were least likely to see value in preserving e-mails. Generally, they hold the view that information in e-mails ends up somewhere else. Respondents from the disciplines of physics, physiology, cardiovascular medicine and biodiversity informatics all expressed the same concern about email preservation – that it is largely unnecessary because any valuable information in email ends up migrating to another place. A physics researcher puts it succinctly: "A lot of the information in emails is important, but that information tends to make it to our wiki. The rest is probably irrelevant."

However, the value of archived email discussion lists is recognised by researchers in scientific disciplines. One researcher said he usually finds the answers to his technical questions via mailing lists that are archived on the web. A biodiversity informatics professor goes as far as to say that the only important emails in his work are those from a taxonomist mailing list, precisely because they are archived and indexed:

> I'm quite keen on having all of these conversations in public, because there's more value, in a sense – people discover them. People you don't expect. Some communication is by email, but if you're going to spend all that effort writing something, why not let people read it? … [Non-public] email is a pain – most of my email is either spam or administration … Email is a chore now, it's not terribly useful.

While this reservation was only made explicit by those working in the health and natural sciences, other respondents shared their own reservations. To some, preserving emails is just a means to an end, or even a system of filtering information – separating the wheat (of valuable content) from the chaff (of mundane, day-to-day trivia).

There is a grey area in relation to which aspects of an e-mail have value and whether the content of email messages carries value for future use, or whether the email is only valuable as a medium for sending attachments. An English literature lecturer thinks that "the actual emails are trivial," seeing no need to preserve anything beyond the attachments.   A cardiovascular science researcher initially stated that email is primarily a transport system for attachments, but then changed his mind, saying, "It can be both [attachments and actual email body, that are important] … the interpretation of the attachment is in the email, and both are important."

Others, mainly in the arts and humanities, were concerned that their email was a private matter. They were uncomfortable about these details being made public, and believed that a preservation system would compromise the way in which they communicate. "I would say that I'd be hesitant [to make them available] because emails are sort of off-the-cuff," says a researcher in English language. 'There really isn't anything too personal really, just maybe a bit … loose?"

By contrast, others recognised that there may be inherent value in preserving e-mails. Many were aware that the personal correspondence can shed light on the collaborations and dialogue to

support the research. An English literature professor stated that "Personal, trivial and incriminating [emails] should not [be preserved] but there's also the swapping of draft work between collaborators, changes, [and] suggestions for meetings." This use of email as a working area for ideas was shared by other respondents from the arts and humanities, and the value of preserving this was widely understood.

One viewed email as correspondence, that is always worthy of archiving and study: "Correspondence is a very valuable way to gain insight into how projects developed, in a very nuanced way instead of official documentation. I see email as an electronic form of that." An art historian agreed, but felt that the institutional attitude towards correspondence had changed: "In the old days correspondence was passed on to your successor. In most jobs now they tend to ignore the previous person's email, or it's considered personal."

Most of these people did not favour the 100 per cent preservation of everything, if not due to other reservations then because it felt like overkill. The problem in establishing long-term email preservation is that of organisation and curation. A natural sciences researcher put it succinctly: "A lot of the emails you send are disposable, but everyone has emails of archival importance as well."

A PhD student attempted to make sense of this on her own: "I started to set up subfolders on my own email account. One for supervisions, one for thesis-related stuff, one for teaching related stuff and one for personal." An art historian, on the verge of retirement, expressed concerns that her emails would need a proper archivist:

> I didn't think [email should be preserved] but I've realised it probably should be and probably isn't going to be. I've realised the records of the [artist's] correspondence are there, and I don't know what we're going to do because some of it is private. I have no idea how to properly archive it once I leave … There's stuff in there that definitely should be preserved, but the main mechanics of when we have lunch I'd get rid of … Maybe we have to send them all to the [museum] and then they have the problem …. There are two bits that matter – one are the records which are in various collections and I have to print it out, and the other is a more private thing. I get asked to authenticate works of art, and whether I say yes or no, it's a pity to let that go.

Respondents were asked what steps they take to preserve research-related communications and how easy their department makes it.

Almost all the researchers had developed some form of personal archiving, from simple USB-drive backups to more structured systems. "It's prudent to do that," stated a literature professor. Archiving is usually focused on the research outputs themselves (see below), but may involve research-related correspondence too. The personal archive is often haphazard and incomplete, a weakness that many of the researchers were aware of. "I'm never entirely sure that I properly archive my communications," said a physiology professor. The responsibility for data security is not always clearly articulated – some look to have control over their own research legacy, while others are fearful of the University's own ability to protect communications for the long term.

The interviews provoked conflicting responses about the University's email system. Some researchers were incredibly satisfied, citing the University's "generous" disk quota. Others seemed to have the opposite experience, claiming that they frequently have to delete their old emails to save space. Two researchers in the same physics project described how all staff in their group uses Google Mail for email correspondence, completely eschewing the University-provided services. One felt that the University did not make it easy to preserve email communications, while the other admitted that he never gave the University a chance. "We've sidestepped the issue – we haven't addressed it," he reported. Google's system is preferred, though they can not articulate exactly why.

In summary:

- Email is essential to research and may be used for small talk, administration, or the exchange of serious ideas

- Preservation of e-mail should be selective; not all e-mails should be preserved

- Scientific practitioners feel that any information of value in their emails ends up in another place which will be preserved.

- Many people keep some form of personal communications archive largely to preserve research outputs but which may also include research-related communication

4.2.2   Phone and voice communication

Unsurprisingly, the telephone has a diminished role in contemporary research communications among the researchers who were interviewed. It is often used specifically where there is a desire not to preserve a record of communication for example when it is confidential or security-related communication. Respondents also indicated that a phone call could be used for very quick communication where even an email would be too much trouble, such as verifying a fact or where it is easier to talk through an issue because it is complicated.

The respondents from the natural science disciplines expressed greater willingness and familiarity with digital means of voice communication such as Skype, Evo and other voice-over-IP systems. "Audio communication tends to be digital now, partly because the University is encouraging people to use voice-over-IP services because they cost less," reports a physics researcher. Research across multiple institutions, such as the physics project that two respondents worked in, has gravitated towards voice-over-IP solutions.

Interviewees were asked (question B3) about the value of recording telephone conversations. Many were against such recordings. An English literature professor saw no value in this, as his phone calls are "just about the banalities of work of sending it to a deadline". Likewise, an astronomer says that he "can't see if that would be worth the effort." A medical researcher felt that his own written notes, taken by hand during phone calls, had the same value as a full audio transcription.

However, there are instances where recording calls could be useful. An art historian referred to an ongoing "problem" in her department that could have been avoided if a phone conversation had been recorded. A composer from the department of Music also reported that he took written notes during phone conversations, but thought a full recording could be "useful". A physics researcher believed recording conversations "could be useful", though only if this would lead to a specific outcome. "Conversations can be a waste of time unless you come to an action point at the end. What do we do, what can you do, can you do this for me?"

The majority of respondents were unaware of any departmental or University mechanisms to facilitate the recording of telephone calls (question B4). One English literature professor said he was aware that such tools existed, but had never used them. Everyone else was either unsure or uninterested in the University's assistance in such behaviour. An English language researcher saw the University's role in this matter as irrelevant, and identified, as with e-mail, that recording a conversation could constrain the discussion:

> I think if I wanted to record a telephone call, I wouldn't know how to go about it. So in that sense, [the University is] not actively making it difficult. Telephone calls are sometimes chosen as a method for that very reason – because it doesn't leave a lasting record. I think people may become uncomfortable; presumably they'd have to give consent, but nevertheless, I think it might put people off.

In summary:

- Scientific practitioners are more comfortable with electronic voice communications
- The value of recording / preserving telephone conversations is questionable
- There may be some specific and limited value in recording them in particular situations

4.2.3   Face-to-face meetings

Personal meetings are still at the heart of research culture, even if these meetings are less fundamental to the development of ideas than they were in the past. Face-to-face meetings are inherently difficult to preserve beyond video or audio recordings, and this is rarely done. Like telephone calls, the question of recording or preserving meetings is something that academics have rarely thought about (question B5). Full transcriptions or audio recordings are perhaps a case

of "too much preservation" for something that is best left as an ephemeral event. And like email preservation, there is some need to filter out incidental information.

Many respondents indicated that they take notes or minutes during meetings, and most felt that these notes would be sufficient record. A PhD student admitted to taking notes of conversations she had with other academics, primarily her supervisors, and also felt that an audio or video recording would be unnecessary. One project leader commented that these notes work best when paper-based: "I tend not to take my meeting documents on a computer – I prefer a hard copy so I will print out agendas and papers, because it's quicker to refer to, use and annotate during a meeting." These written annotations do not always make it back into a preservable, digital format.

There may be specific occasions when it would be beneficial to record face-to-face meetings. One lecturer in social and economic history created a dataset with a group of others from his field. He believed that it would have been beneficial to have recorded the meetings which were about the design of the database, though he added that he was "not sure an oral record would be any better than the detailed notes [that were] made."

A music post-doc shared a specific anecdote of a face-to-face meeting that would have benefited from being recorded. "I had a conversation while walking around the middle of Trondheim, Norway that would have been useful to my research," he related. "And both of us said 'Hey, we should have recorded that, let's do that someday.'" Meetings and casual conversations often contribute to ones personal development as an academic, though impromptu interactions may be compromised if recording is premeditated.

In terms of how far the University makes it easy or difficult to preserve face-to-face communication, the interviews indicated that, as with telephone conversations, the University does nothing to make the recording of ephemeral events difficult. This does not concern most respondents, as few have ever inquired about such facilities. Those who are aware of it see the University as a neutral figure – neither facilitating recordings, nor actively preventing them. Institutional preservation does not extend this far. "Occasionally, the technology is lacking but there are no other obstructions," says the music post-doc.

One researcher reflected on the challenges of capturing face-to-face interactions:

> We all have meetings face-to-face where you can walk away from it feeling really successful, and then you think about it and you sit down, and you think 'Why?' Taking notes immediately afterwards actually helps me, on a personal level. Whether they end up being integrated into some kind of email report, if this was some kind of important meeting, one that was going to shape the future of the project… but it was just a kind of update, something exploratory with somebody else, I think it's all too easy to lose that information.

Face-to-face meetings exist not only to discuss perfunctory matters but also for the genesis of thought and ideas - there is something ungraspable, yet ineluctable about working with others; it cannot be replicated through virtual telepresence. This is especially the case in the performing arts disciplines, this is more the case, though ironically, these meetings are the ones most likely to be recorded.

In summary:

- Face-to-face meetings are rarely documented or preserved beyond the form of meeting minutes
- Many respondents are satisfied with minutes as a written record of a meeting

4.3   Attitudes towards research practices

4.3.1   Research outputs

The fruits of research are the staple diet for preservation systems. Planets has an inherent interest in the outputs of research that are valued and should be preserved for the long-term. The subjects in these interviews all produced traditional and non-traditional types of output, and discussed these products in terms of preservation. Additionally, there were many thoughts about research work that

does not directly lead to publishable output, and the value (and challenges) of preserving this material.

Interviewees (question C1) were asked: In your field, which research outputs are the most important for others to use, re-use, and adapt into their own work? All respondents indicated that traditional formats such as journal articles and book chapters were essential to their field. However, researchers who had produced datasets or online resources felt these materials were more useful in the work of others, regardless of whether the data is scientific or humanities-based.

Traditional written outputs still dominate in the unyielding academic world of arts, humanities, and social sciences. A sociologist saw value in only two types of research output in her field: books ("important") and articles ("less so"). The field of English literature is likewise focused on the monograph - books, published articles in journals or other sources, and book reviews are still the most valuable currency. The mechanisms for re-use and adaptation are rigid, following accepted standards of quotation and citation.

In the natural sciences, raw data tended to be valued as the most significant output for others. In both astronomy and physics, respondents indicated massive quantities of simulation data and observations to be the most valuable resource. Of course, this data underpins the numerous journal articles that propel academic careers, but it's these underlying data that allow the articles to be written in the first place.

In the health sciences, journal articles are also dominant and are paramount to the development of the field (as well as a means to underscore one's career). A respondent indicated that books and book chapters are significantly less visible than journals, due in part to the frequently evolving nature of the field. Additionally, the "cut-throat" competitiveness of his area of research made book chapters less necessary: "People do get asked to write book chapters, but it falls way down the priority list, because book chapters don't tend to get cited." This researcher also commented on how journal articles are a means of judging the quality of an individual researcher, based on the number and quality of published papers. Bluntly, he states "You gotta have the good papers to get the grant to keep you employed." However, he later indicated that the raw, experimental data that is produced through laboratory testing was perhaps the most important research output in his field in terms of use, re-use and adaptation by others.

Datasets and databases, particularly when web accessible, have wide and enduring value to others. Even among our small sample, dataset output was not limited to the scientific discipline. Respondents in art history, English language, and economic/social history all produced databases or sets of information that could be cited as a source by others, usually without access restrictions.

These datasets are often evolving or changing, making preservation an ongoing process. One database was created by an art history researcher as his career-spanning project. This database is the end product of the project, and can be freely accessed by anyone via the web. To him, any academic papers written as part of his project are significantly less valuable to related researchers, and certainly more difficult to find. The website, however, provides a rapidly growing resource that can be found by anyone with an Internet connection.

A researcher in English and Scottish languages who has created a web-accessible database feels similarly about his project. Any presentations, journal articles, or conferences that result through his work, while still "important", are significantly less valuable for others to use. "Researchers really do hit [our website] and I think that will have the biggest legacy, if you like. If we could only keep one thing, it would be that because it's not just raw data but something that can be interpreted and understood."

Another respondent, who maintains a different art history web database, valued her site because of the way it is freely accessible to anyone. "I do like exhibitions – this is art history," she says. "A lot of people see it, and you get a catalogue. But the last exhibition we did, we've also put it online, which has a longer lifespan." She has focused on the website in more recent times, often to the dismay of more traditionalist colleagues: "A number of people in my editorial board would like stuff in hard copy -- so I said they can print it off our website!"

However, there are instances where the value of databases. may be limited. A lecturer in Economic and Social History had worked on a large project to create a social sciences dataset that is freely available for anyone to use, but he is sceptical that it has ever been of value to anyone else. "In terms of the dataset, I think it's so complicated, with over 200 pages of explanatory notes, it would be very difficult for anyone to actually use it. So traditional methods of communication – ie: writing, are by far the most important."

Few of the respondents initially identified conferences, seminars or other events among their research outputs. (One social scientist explicitly disagreed, however, seeing these events as activities, not outputs.) When pressed however, most agreed that these events did constitute research outputs, though there was little consensus about their value. For example, one respondent who felt strongly that live events were a key mechanism for making manifest research is a composer who writes only for recordings and concerts. Never publishing scores, his concerts (and their preservation, through the form of recording) were considered to be the main output of his scholarship.

Conference proceedings, perhaps because of their citable format, were more often listed as outputs than the actual conferences. One lecturer saw conference proceedings as somewhat obscure publications, comprised of material that is usually refined into something better. However, researchers across several disciplines (astronomy, English literature, and physiology) mentioned their own conference papers as valuable, though other respondents were less concerned with them.

A music post-doc believed that some of the most valuable output he had created while at the University was written "non-academically" for art exhibition catalogues. While not strictly academic conference proceedings, the respondent believed that these writings would be far more useful to others because they "would be read more than scores, or theory of scores".

A sociologist put it succinctly:

> I do think [conferences and seminars] are important, but I have a limited amount of time, and I think you could spend a lot of time travelling to places where you didn't necessarily develop intellectually very much, as a result. I think conferences are very good for giving us the structure and feeling of a time, but for more precise, focused work I think it's better to sit and read in your room.

One notable exception is in astronomy, where conference papers are refereed and "carry the same status that journal papers do in other disciplines." The astronomer stated that after raw datasets and images, conference papers and journal articles were the next most valuable output for others.

An English literature professor sees conferences and seminars as a "pleasure", but cites that as being partially due to "place and ephemerality." He continues:

> The degree to which they are preserved is still, I think, in its infancy – certainly in the humanities. There are institutions I know who are doing a lot more, preserving things as podcasts or posting to YouTube. There is a fear in the academic community that these will be used for long-term redundancies.

In summary:

- Traditionally published research outputs remain essential to all disciplines in terms of use, re-use, and adaptation

- Researchers who have created data sets, databases or other online resources feel that these creations are more important for others than traditionally published forms. These resources are most beneficial when they can be accessed by the general public without usage restrictions.

- Events and conferences are not generally seen as research outputs, though publications created at them are, especially where these are peer reviewed.

4.3.2   Preservation of research outputs

The attitudes researchers have towards preservation of their own output develop over the course of their careers. The attitudes and requirements of any one researcher will therefore be slightly different to those of another, even if they are working in the office next door.

The preservation of research outputs is of importance to all academics, who are concerned about the development of the field as well as their own research legacy. However, a surprisingly varied set of viewpoints emerged from the interviews, indicating that the role of institutions and the level of individual involvement play a factor in preservation.

Questions were posed regarding the responsibilities for preservation, the risk of loss, and difficulties in accessing other research outputs. Unsurprisingly, different attitudes came from those

with different daily practices, as well as those with past experience in institutions and research projects.

Researchers were asked if they feared their own work was being lost (question C2) and in follow-ups, the notion of their own research legacy was brought up for discussion. Surprisingly, the idea of their research legacy being lost did not seem to be a concern for the respondents with published research outputs, as they had confidence in the longevity of material in printed form. Most respondents indicated that they often think about the long-term availability of published material, but they did not often judge loss to be a relevant risk. Researchers were aware of online publication resources, even if they were not aware to what extent their own work was being preserved.

Attitudes towards these differed. The emergence of Google Books as a freely accessible resource was met with enthusiasm by an English literature PhD student, but with extreme scepticism by a lecturer in the same department who is about ten years older. His concerns were rooted in the financial implications of such accessibility:

> I know more and more over the last year or so people have been reading my work over Google Books instead of purchasing it. I wasn't aware how it would benefit a publisher ... [My book that was published recently] costs about £90. Most of it is available freely through Google Books. How is a publisher supposed to generate sales if most of it is available? That's how most people will read my work these days, or through electronic journals like JSTOR and LION.

The same lecturer however recognised the potential benefits of open access for students:

> Some books haven't had any revenue attached to them. I don't know that that's specifically due to Google Books or not…. It's helpful for students especially, who can't afford to buy the editions – I don't have any problem with that. But it is an anxiety when a book is published in a very limited run of 100 or 200 copies. I can't quite see the publisher's logic in passing that on to Google which would further erode any publishing royalties they might expect.

A sociologist was unaware if her books are being preserved by a third party such as Google, but was also enthusiastic about the opportunities it presents for widespread dissemination, saying "I think I would be quite keen [to have work available for free via Google Books or other service]. I think the main objective is to diffuse the book – or the thoughts in the books. It wouldn't worry me very much if the publishers won't make great profits from that."

The comments relate to concern about publishing industry revenues, or the desire to make work accessible to as wide of an audience as possible; they do not relate to fears about published work disappearing. The idea of their research legacy being threatened does not seem to be a concern to these respondents, as they have confidence in the printed form.

However, the situation is viewed differently for content that is born digital. The English literature lecturer quoted above, who is relatively young (with 10 years of experience in his field), is quite reluctant to publish articles in electronic-only formats. At the time of the interview, he had never published in any such form because he was concerned about loss:

> There's a kind of unspoken prejudice about the permanence of an essay being reproduced through electronic form. Colleagues will often grumble that an essay they've had published on the net only – the website has been removed or updated, or their work has been badly damaged by a webmaster. Among English Literature lecturers, there's always a feeling that the e-journal is somehow less reliable …. I hope [that attitude] is changing. I would be happy to publish in an e-journal, but there's a lot of negative feedback surrounding it, simply because it's an idea that a website-based magazine or journal is prone to disruptions that a print copy is not.

A researcher in social and economic history echoes these sentiments:

> I think anything that isn't finished when one dies is likely to die with one – that's a fact of life. If I were to go under a bus tomorrow, there are lots of unwritten articles and book proposals… What does concern me is publishing in online-only outlets, particularly websites without stable URLs, for example. There are some things I have done that have gotten lost or moved, and are very difficult to find.

Researchers in scientific disciplines seem more ready to abandon printed formats. Though the issue of trust in digital media was asked explicitly in question D2, the theme crept into answers across many of the questions in section C. These attitudes are explored further below. Preservation of less traditional forms of research output, such as databases or websites presents their creators with more immediate challenges. The time-bound nature of project brings a particular set of constraints that were less relevant to paper-based environments (though with

many journals heading into an electronic-only format, more traditional forms of outputs are facing them too). One database maintainer understands that his efforts are only valuable as long as the databases can be accessed:

> We've got to make [our website] available and always available. It's interesting to note now that the funding body who just funded this cycle, we have to promise to keep it going three years after it finishes. The big institutional, national repositories have folded, or aren't taking on new deposits. This puts projects under a slight concern. The strategy we've taken … is to keep it going, mainly through extending it. We can justify to the department that whatever cost it takes to keep this thing running is acceptable if we keep adding little bits of content to it as we go along. I'm not sure how sustainable that is; that will take somebody's time. That's fine when nothing's going wrong. We've got to worry about technological obsolescence.

Despite the impermanence of electronic content (either real or perceived), few respondents felt that elements of their own work were being lost (though some replied that they had never considered the issue before). However, one arts and humanities researcher was extremely fearful about data loss, following an experience where he nearly lost the entire manuscript of a not-yet-published book. Others, however, reported that they willingly destroyed their own notes or elements of their work in order to simplify, save space, or assist in organisation. Some respondents seemed unconcerned about losing any unpublished material, seeing this as part of the nature of academic work.

Those who write source code (for scientific programming) use version control systems such as Subversion or CVS to preserve their work incrementally. One physics programmer did admit that he wrote some code that was not being formally preserved, and could be lost: "I have developed code that is just sitting in my home directory." But this was of minor concern to him. Another physics-based respondent even admitted to using Subversion to store written drafts of papers and journal articles, thus ensuring as a means to preserve interim drafts and version control. A biodiversity informatics professor used Google Code instead of any University-based resources partially because he was worried about his own work getting lost, but also because he wanted to easily share his work with people outside of the University system.

As discussed above in relation to preservation of research-related communications, and research, many researchers use personal backup and archiving systems. These idiosyncratic systems, in conjunction with varying degrees of trust in University server space (network drives, backups, etc.) are primarily responsible for the peace of mind.

In summary:

- Researchers are concerned about the long-term availability of research outputs, but generally do not actively worry about their unpublished work being lost. There is a high degree of confidence in published research continuing to be available in the long-term, if published in print form. Arts and humanities researchers are very hesitant to publish in electronic-only formats because of their perceived impermanence, but scientific researchers make no such distinction.

- There is less confidence in research published online, especially among arts and humanities and social science researchers.

- Attitudes differ regarding the free availability of previously published books via Google Books or similar services.

4.3.3   Preservation responsibilities

Discussion about long-term preservation of research outputs leads to the question: Who is responsible for such efforts? The participants in these interviews were asked to discuss their thoughts on this subject (question C7) and this led to a variety of viewpoints.

Surprisingly few respondents (a PhD student and professor of physiology) held the publishers of research outputs responsible. More felt that the University should be accountable for the long-term availability of its output, including the physiologist (who believed that unpublished material fell under the University's domain, and published to the publishers).

Some scientific researchers believed that the funders of their research should be at least partially accountable. One database creator believed that the preservation and backup responsibilities should be delegated to the funders (who pays for them) or the University (who maintains the

servers). "Rightly or wrongly, [preservation problems] are not something I think of as my problem because it's always another party responsible," he said.

Others believed that the individual should be responsible for ensuring the preservation of the material they create (including the PhD student, who believed that unpublished material was the responsibility of the individual).   One physics researcher had the nuanced view that the responsibility of preserving research outputs "is always going to shift".   He continued: "There are departmental and University priorities and you can't avoid the differences."

The comments point to a degree of uncertainty over roles and responsibilities which will need to be clarified through the articulation of organisational digital preservation policies.

As far as the University's efforts go, the majority of respondents believed the University was taking care of their research outputs at least somewhat. A few scientific researchers expressed doubts, and some were not looking for the University to be involved with long-term preservation of their work.

For formally published outputs, Glasgow University Library maintains an internal repository called Enlighten. The repository is available to any University staff in any department. Only some of the staff who were interviewed for this report were aware of its existence   Of those who were aware about it, some mistakenly believed that depositing in Enlighten was mandatory as a University-wide policy – though it is in fact only 'strongly encouraged'.

Some regarded Enlighten as a useful facility. One researcher in English Literature said: "I quite like Enlighten … It's good that the University is there to do that."   Another could see benefits and risks: A researcher in the social sciences feared for the future of journal publishers because of repositories, saying; "I can see why it's good – it increases our profile and citations, and people will hopefully link to the proper location. But we will get to a situation where refereed journals get marginalised and people will start using these repositories [instead]."

However, others were very critical of this system. A professor of English literature once tried to remove earlier essays of his from Enlighten.

> I didn't want certain publications placed on the library record because they were essays that might have been out-of-date, or they had arguments that I had subsequently revised quite dramatically. So I was not happy that they were being placed on the library website. The items that they listed there were not the ones that I wanted to have listed there, but I couldn't have them removed. The library refused … I just feel that it's problematic that I don't have a great deal of control over how the library markets or advertises my research repertoire … There's an element of unhappiness about how the University represents me to the public, or to the wider scholarly community … It's difficult to tell who's accessing them. They don't provide any feedback on how many hits my essays are getting.

The two oldest participants in the interviews, both with over 40 years of experience working in academic research, were also significantly less trustful of the University's ability to look after their efforts long-term. One expressed doubts based on an experience where her email archive was lost; the other feels generally satisfied about the protection of her outputs, but does not think her notes will be maintained by the University (see section 4.3.4, below). And for the English literature lecturer who was unsuccessful in removing his old publications from the Enlighten system, the University is guarding his output *too* closely.

In summary:

- Few respondents hold the publishers of research outputs responsible for their preservation

- Some scientific researchers believe funders should be responsible for preservation

- Many researchers look to the University to safeguard their work although some believe it is their own responsibility

- Researcher may desire control over their published work, even after the publication date

4.3.4   Intermediate results

While structures are often in place to preserve formal research outputs, there is still an enormous amount of research effort that does not result in an end product. This content, which we have called *intermediate research results*, may include notes, drafts, or unexplored tangents. It may also include results of laboratory experiments, source code, unpublished manuscripts, sound

recordings, or raw data. Essentially, anything at all that was not seen as an "output" was considered, for the purposes of the interview, to be an intermediate result. Interviewees were asked (question C5): "Should intermediate research results be preserved, such as drafts, notes or incomplete tangents - and why? What specifically in your own unpublished work do you think might be beneficial for future use?"

Intermediate results are a grey area in terms of preservation, and thus these interviews sought to explore exactly how their creators value intermediate results. In particular, we looked for anecdotes about how intermediate results may be used or adapted by others.

Unsurprisingly, attitudes towards the value of intermediate results and preservation of them varied. Some only saw value in preserving these materials for historical purposes, in a similar way to preserving e-mails is used to file correspondence. Others thought that seeing earlier drafts or notes could open new insight into a published work. In some ways, material excised from final publication may be very important. As one English literature professor saw it: "There are a lot of sort of amplifications that you later have to cut to adapt to a word limit that I think are important to have preserved."

Respondents were asked (question C6) whether they had ever used anyone else's intermediate results in their work, and perhaps this influences the value they place on their own results. The value placed on intermediate results varied by discipline.

For a cardiovascular researcher who works in a group, intermediate results are the essential building blocks of his scholarship. His laboratory notelab book is the centre of his work, and contains all of his intermediate results that can be represented in printed form. This information must be preserved and made available to any other scientist who wishes to validate his research. "You always live in the world where someone will see your article and then want to see the real numbers, so you print these things off in the lab book."

Everyone in the lab frequently cites his lab book. He continued:

> I do find myself referring back to lab books. What I tend to do, is when I generate files, they're all given a date, so I can cross-reference with my lab book. I can flip to my lab book and find a particular graph that looks important, and then I can do a search on the computer for that date and bring up the files I created that day.

Digital data, when it cannot be placed in the lab book, is shared among the group using email. This is a virtual extension of the lab book, though it is not always tightly bound to the physical book.

Respondents in the artistic practice of music had a different take on intermediate results. A post-doc, upon completion of his PhD, sees value in preservation of them, saying "There's a map making exercise involved in keeping [some intermediate results] … it's more difficult to see how things developed." However, he happily destroyed most of his notes for practical reasons: "I just got rid of lots of old notes because either I didn't want to read them, or I couldn't imagine a situation in which I would read it. Paper notes. I didn't think there was anywhere to put them – I am also very disorganised." He is unaware of anyone ever using his intermediate results or vice-versa, "beyond conversations and anecdotes".

A second composer reported that he had used others' intermediate results, if that was taken to include elements such as handwritten notes on original scores. The idea that marginalia could contain value that enables reinterpretation has found currency in the literature discipline, in the subfield now called 'genetic criticism'. A PhD student in the field frequently looked at handwritten annotations to original manuscripts when these materials were available in archives.

Whether this material will ever actually be useable to anyone is another story. Access and organisation again become factors. The high volume of information already available to researchers means that additional academic paraphernalia may be overwhelming (if it can even be found in the first place). Another English literature professor sees this as a problem:

> How the hell would you file the various ephemera of an academic today though? [Ideas, intermediate results, etc.] you tend to let the world know. Many of these are boxed up, others are already digitised, and the notes of many important writers and thinkers are still largely neglected.

A physiologist understood the value of his intermediate results, but is aware of the reality of storing it all. He favours the preservation of intermediate results because it would be "useful", but has been living without it for his entire 36-year career.

> I think a lot of stuff gets lost. For our area, the ethical thing is that you should retain the data in a laboratory where it's created, and you shouldn't take it away with you. But students do take it away with them. The storage thing becomes physically too difficult, and you throw away old data. I think that is a loss … Ten years later I might find more data that would enable me to interpret [the old] data but its gone … it doesn't keep me up at night, because that's life. You gotta get on with that. I've been trying to find some files for a student, a post-doc who was with me 10-12 years ago. Because I've done new experiments, I now know how to interpret that data, but I can't find it. I'm never going to find that data.

Respondents were asked who should and should not have access to others' intermediate results (question C7) Scientific intermediate results are based on empirical evidence and do not have as much of a personal stamp on them as material from the arts and humanities. Consequently, there is greater willingness among scientists to share them, although this is resisted in commercially funded research disciplines such as cardiovascular medicine. Access restrictions are more likely to be desired by writers, essayists, and those from disciplines where the success of one's career depends on the originality of their viewpoints. One English literature professor strongly believes that the control over these notes should be up to the individual, and would never use material from anyone else's working papers without permission. Of his own work, he recognises the value in returning to older intermediate notes; "Those are often things that I use privately, and I will bring them out from time to time and refurbish and finesse them… for lecture notes or seminar notes. I wouldn't see them as worthy for a public domain or public platform".

One researcher points out that it is not for him to decide whether or not his own notes have value for others; another sees the idea as one of vanity and self-importance, adding: "I think the process of writing is a very individual one. Even in a collaborative group – it's still doing something that is a unique blend of our personal inputs. I don't think there are many lessons in that that will be translated to other scholars from looking at our ways of working." He is content with the current preservation structures of his department, even stating "I lose notes all the time – that's just a part of life."

Sharing of intermediate results can be surprisingly easy. For many who work entirely in the digital domain, shared network storage is a simple method of preservation that is already in place.   For those who still work in a largely paper-based structure, the fate of intermediate results is much less certain.   A sociologist constructed large tables of data by hand for her most recent book, and while she doubted that anyone would be able to read her handwriting, she stated "I would never throw away my handwritten analysis, because someone could conceivably find use in them." However, she believed that her shorthand notes "will probably be destroyed after I die".

In source code/computer programming situations, version control systems are already, in most cases, in place. These are usually run at project (or departmental) level and centralised University computing centres may not be aware of them, though the repositories are possibly backed up through standard University data storage policies.

A professor of zoology took perhaps the most cutting-edge attitude towards sharing intermediate results, viewing social networking mechanisms as one way to disseminate these. He uses the Evernote service to organise his intermediate results, which he then sends out to associates using Twitter. All of his material exists on a server run by the Evernote company and no University storage is used whatsoever. Taking it even further, he no longer publishes traditional articles or papers, instead using Internet-based discourse to share his ideas and expertise with the community. "Boundaries are blurring in publishable work," he states, and the majority of his "publications" of late have been in the format of blog posts, Tweets, or blog comments. This view was unique among our respondents, but perhaps indicative of changes to come in some disciplines.

In summary:

• There is no clear consensus about the value of preserving intermediate research results.

• Some scientific researchers exchange intermediate results frequently

• Arts and humanities researchers have a more personal connection to their intermediate results, and are perhaps a bit more protective of their ideas.

• Researchers generally are not preserving intermediate materials now and some are at peace with the idea of them disappearing one day.

- Version control systems for computer programming are in use widely as an already-functioning method of locally preserving intermediate results, though the wider preservation situation for these system can vary.

4.4   Changing attitudes towards research and information

The final section of the interview looked at attitudes towards information use, re-use and adaptation.   Seeking to draw on the identified differences between disciplines and individual researchers' comfort with technology, questions were asked comparing digital and analogue means of communicating and retrieving information. The role of the Internet in the contemporary research environment was explored, particularly through non-traditional and emerging web-based resources.

4.4.1   Digital and analogue resources

Though we have certainly been living in a digital age for some time now, few researchers would consider themselves to be 100 per cent digital users. Paper still abounds the workplace, and the role of handwritten notes or annotations does not seem to be disappearing any time soon.

All respondents were asked for examples of when they would opt for paper rather than electronic correspondence. Generally, decisions to correspond in hard copy were motivated by some sort of policy – because a request had to be made in writing, or because it is 'the done thing' for example formal communication with a distinguished academic at another University. Face-to-face meetings are valued as a back-up to electronic communication as a means to develop and share ideas and coordinate activity. Analogue voice communication tools such as the telephone was favoured for confidential information or security-related issues.

Questions were then asked regarding ease of access to digital media and trust in it: In which ways do you find digital materials easier to access than analogue and in which ways are they more difficult? (Question D1) In which ways do you trust analogue resources more and in which ways do you trust digital ones? (Question D2)

Digital resources are generally the preferred media for searching, accessing and using research-related information. Digital is championed for adaptation and re-use of data/information, even by English literature professors: "The way that you can manipulate the data is probably the best advantage." In fields such as astronomy and particle physics, there are no analogue resources in the field – they simply do not exist. The ability to search for digital materials and specific content is beneficial to researchers.   A music post-doc singled out the ability to search through PDF documents as a valuable property of digital media, and an art historian is also thankful for this when compared to her early work with note cards.

 A social scientist does offer a complaint about using and re-using digital resources, particularly when in the form of datasets – that they "require some effort to get familiar with them." But once this familiarity has been established, they are beneficial: "They are always in text format so it's dead easy; you can import them into any package you want, as tab-delineated data files."

Criticism of digital resources tended to come in situations when the digital resource is merely an electronic version of a book, article or paper. Analogue materials were still favoured in disciplines with much reading, as several respondents indicated that any document longer than about 4 pages was unpleasant to read on a screen. An English literature professor puts it well: "Reading on the screen is onerous and it gives me a headache after awhile … I find reading anything closely off a monitor to be energy-sapping after half an hour." In scientific disciplines where digital content is often a unique format that has no alternative representation, these issues were never mentioned.

In terms of trust, one researcher was quick to say that she didn't trust any information. "Any bit of information I get, I say 'but—' and question." This rigorous approach to authentication was not made quite so explicit by anyone else, though the ultimate role of the individual in assessing information was understood. One put it nicely:

> Everyone has filters, checks and balances in what they believe. I do not swear by every word in Wikipedia; also I know about the phenomenon of vanity publishing. I can tell you that [a departmental colleague] wrote his own Wikipedia entry.

Most admitted their biases towards forms of information presentation. Many in the arts and humanities were reluctant to trust web-based resources:

> If someone has gone through the trouble of commissioning, writing, editing, binding, marketing, selling a book – then clearly somebody thinks it's a valuable thing to do. In that sense it's already been through a validation process. Stuff on the web hasn't, so obviously you have to be more critical.

An English literature professor trusted analogue materials more, though he had not realised it before: "With hard copies you have something physical, material … it's just having it in your hands." A 28-year old PhD student in his department had no concerns whatsoever about trust in digital resources, only about how they were perceived by the rest of her field: "I will always be aware, in my thesis, of when I was using digital material and how so as to look scholarly in the eyes of an earlier generation of researchers… I guess I always have one eye on how it will be perceived."

By contrast, most scientific researchers trusted both digital and analogue resources equally but admitted to using few, if any analogue resources anymore. One physics programmer was aware of the potential problems of digital resources, asking "Anyone can write whatever they want online, so how do we trust it?" Of the two respondents from the social sciences, both indicated that they trust digital and analogue materials equally (though a sociologist commented on how others in her field were not as accepting).

In summary:

- Digital is generally preferred to paper for research-related communication unless there is an organisational or cultural requirement to correspond in hard copy.

- Digital is also favoured for searching, accessing, adapting and re-using information unless this involves large amounts of reading where paper is preferred. Researchers, even in more print-dominated disciplines, are grateful for the accessibility of digital resources.

- Trust is generally higher in printed media because the publishing process validates the research.

- Scientific environments were the quickest to move into an all-digital realm and likewise there exists much less bias against digital publications.

4.4.2   Problems with digital information

Planets is focused on solutions for long-term preservation and so it was of particular interest to hear any anecdotes about problems accessing or using digital resources. Questions were asked about ease of access to older, obsolete digital information that could no longer be accessed, or difficulties in adapting content to their own work.

Surprisingly, few reported difficulties relating to preservation risks. Most respondents who indicated that they had trouble accessing digital content blamed their problems on legal issues, such as licensing constraints. The few who did have format or age-related problems were generally able to solve them.

Two database maintainers and one physiologist described issues with accessing vast amounts of old physical data being stored in their office. Old floppy disks and early CDr and DVDrs are not necessarily reliable, but no one seemed concerned. Much of that data had been transferred to network storage earlier in the life of the project, leaving the original disks as physical artefacts. "I could open that cupboard and find a lot of discs I can probably no longer read," stated the physiologist, but while shrugging.

One art historian described an incident when he had to return discs of images to a data provider because he was unable to open the files because the format was unknown. This was resolved by returning the discs and all subsequent digitisations used a standard, web-friendly format that was commonly used on the web.

A medical researcher reported a problem with accessing a digital file, but it was easy to fix, requiring just a minor software installation. Another researcher received an attachment through email in an unrecognisable format, and merely asked the colleague to send it in a different format.

The predominance of standard formats such as PDF eliminates problems for many researchers. The versatility of this format is clear to one professor, who feels confident that the future will be relatively free of problems:

> In English Literature we're so grateful for the instantaneous availability of material that would have taken weeks and months of library work before – we're so glad of that – and so much of this work is historical and work that would be 100% obsolete regarding previous technology. Being aware of the advantages of that, and maybe being among the second or third generation of academics who are computer literature – they will take steps to know that anything more recent will be preserved, in a preservable format.

A sociologist who works with newspapers from the early 1900s has been generally satisfied by digitisations, primarily because she can access them from anywhere. The problems occur when some issues are not yet digitised, because it then requires a trip to London to access them.

In summary:

- There have been few experiences, at least among these respondents, where preservation risks caused problems.

- Legal issues and rights play more of a role in preventing the use and re-use of digital materials.

- The standardisation of file formats has eliminated many problems.

4.4.3   The Internet in research

The final part of the interview asked respondents to talk about the value of internet resources to research. A question was posed to all respondents asking if they had ever used material from the personal web page or blog of another researcher, and this often led to discussions about non-traditional resources in academic work. Although there was not a formal question regarding Wikipedia on the interview protocol, conversations usually progressed toward this topic.

Attitudes to re-use and citation of online research were split by discipline. Unsurprisingly, disciplines that produced traditional monograph-based outputs were very reluctant to cite any information from a website or blog, let alone Wikipedia. Though citation rules exist for web resources (and systems such as WebCite attempt to preserve some integrity), a discipline such as English literature does not tolerate such sources in essays or journal articles. One lecturer put it this way: "I wouldn't use blogs, simply because there's a problem there about how to cite from blogs, how to get permission from the blogger if I chose to quote from it. All of those hesitations tend to work against using blogging as a viable source of academic information."

The scientific disciplines were completely unconcerned with these rules, an attitude that echoed their greater willingness to publish in electronic-only journals. One physics project (where two of our three respondents from that department were based) was happy to use any piece of information that "worked", whether it came from a refereed, peer-reviewed book or a completely anonymous blog comment. Information that can be scientifically tested has its own internal means of review, which understandably sets these disciplines apart from interpretative fields like English or history.

The capacity of the Internet to support keyword searching and enable fast and easy search for information was referred to frequently throughout the interviews. The use of Google has transformed research more than any other element in modern times. This is a mixed blessing according to one older English literature professor:

> Students now are great – they can re-pen everything that is applicable to a keyword or keyword/author – and they are experts in that. What they are not expert in is any of the historical elements of the debates in which these articles take place. What they are entirely innocent of is what I call 'collateral learning' – knowing if something else is on the next shelf of the library or on the next page of the journal.

Some also referred specifically to the potential for social networking and blogging as a means to facilitate rapid and wide dissemination of research. A professor of zoology/biodiversity informatics was by far the most enthusiastic supporter of social networking, using Twitter and blog software to constantly expand and develop his professional network. To some extent this is because his research focus has shifted into an area where there is not a journal to cover it, but he also feels that social networking offers a way to build a real community that is unconstrained by academic structures:

> I don't read scientific papers. As a student I did. I don't read journals. I read blog posts, I read tick reports, I Google …[Writing papers] seems now like an incredibly clumsy process that takes a long time; you have to deal with reviewers, most of whom don't understand what you're doing – whereas I can blog about something and get tremendous feedback, from a much broader range of people than whomever would read your paper. It's much more instantaneous – the audience is much more diverse. There's still a good idea to write up a piece of work, cause you of course polish it more than you would for a blog, but…to have something appear in a few months time, where I can write it up right now and have other people actually make use of it. As much as it freaks out classical academics in a sense, it just strikes me as insane.

Since he has obtained tenure, he has been able to eschew traditional publishing routes (to the consternation of his department head). He favours complete openness in his work – he shares source code, datasets, and forms of intermediate results, and occasionally walks the boundary of copyright law and intellectual property rights. This is not an occurrence unique to his work, as he points out that "many of these bibliographic services using Google Scholar are technically breaking Google's contract." He goes on to point out:

> I haven't been in the library to actually look for a book in …[pauses] ... probably about 5 years. I stepped inside twice in the last five years – one was to celebrate the 100th PhD thesis being digitised, and the other was to get a book from Interlibrary Loan.

Other researchers also said the internet had reduced or removed reliance on traditional published research and the libraries that house it. An English literature professor refers to a colleague in his department who also brags of never going to the library, and another respondent says: "The only time I go to the library here is not to get books, but to get some peace and quiet and get some writing done." An astronomer said that he could not remember the last time he went to the library either. These anecdotes emerged when discussing the ease of accessing digital library resources from the desktop, but also when talking about the emergence of Google and Internet searching.

Wikipaedia is selectively gaining credibility as a means to disseminate and obtain information. One professor saw the denial of Wikipedia as a fear of change, and believed that other academic researchers should support it and engage with it instead of trying to fight it. "The reality is that it's fast becoming the #1 source of information for people. So the question is, do you want to be visible? If you want your information to be seen, you make damn sure it's on Wikipedia."

However, many would not admit to using it in research. Some claimed they limited its use to checking dates or getting contextual knowledge about a topic they were unfamiliar with. One respondent said he had a "healthy distrust" of it. A social scientist said that he was "inclined to believe something is true if it's in Wikipedia" but that he would not be certain until he crosschecked it; of course, he would never cite Wikipedia as a source, and used it primarily when intellectual rigor was not necessary.

Discussion led onto the value of preserving blogs and personal web-sites. Ten respondents felt emphatically that this material should be preserved; four others felt that at least some preservation would be useful. However, as with research-related communications and information, there was no consensus about who should be responsible for preserving it? Preservation of blogs and websites also presents the added difficulty that frequently no institution is involved.

One respondent felt that if blogs were hosted by a University server, then the University should be responsible. However, if academics blogged independently of the University, there should be some way that material could be anthologised, summarised, or surveyed as a meaningful repository -- perhaps in lieu of "complete" preservation of the material.

The social scientists were split, with the social and economic history professor completely uninterested in blogs, and the sociologist (perhaps unsurprisingly) seeing blogs as social artefacts:

> I don't know whether it's too Utopian to hope that you could preserve [blogs, personal websites]. I hope you could preserve them. When you think of the value that we now derive from reading ordinary soldiers letters from the first World War – I think equal value would be derived from doing studies of blogs. Which is not to say that you would necessarily use them for their accuracy, but you would use them as interesting sources, and "structures of feeling" … Somewhere it would be nice to think they were being preserved.

In summary:
- Disciplines that produce traditional monograph-based outputs are very reluctant to cite information from a website or blog, and Wikipedia is unthinkable.
- Scientific/technical researchers are far more open to Internet-based research strategies.

- Social networking has not been embraced by many in traditional research environments.
- There is uncertainty about the value of preserving blogs and personal website, and even more uncertainty about who would be responsible for doing so.

5. Conclusions

One surprising element of these interviews was how engaged the participants were with these topics. Of course, all researchers place an importance on their own work; still, many of the questions of preservation, assessment, or intermediate results were never explicitly articulated before. These topics will always produce lively discussions and debates; consensus is only likely to be found in "common sense" issues.

Though the findings and quotations (in section 4) touch on many areas that may be relevant to Planets developers and users, they are only a partial report of these discussions. Conversation also sprawled into many areas of academia not covered in this report (such as licensing and collaborations with other institutions, the role of corporate money in research projects, and the bureaucratic rumblings of the University). Such divergent issues introduce a challenge when trying to come up with conclusions, because they really lead to more questions than answers.

Sticking to the focus of this report, it is possible to see deeper themes emerge from these interviews. One is the need for organisation and selection, the flipside of wholesale preservation. Many of these researchers are reluctant to engage in further preservation procedures – particularly if they would require more time from their busy schedules – without a perceived benefit. As few of these respondents were concerned about loss or access problems, the challenge may be less about saving everything and more about making this saved material navigable. As skills of information retrieval have developed with technological advances, preservation systems (whether of email, telephone calls, or intermediate results) need to be negotiable with the user. How can institutions that operate these preservation systems manage these negotiations?

Second, there is the matter of defining clearly roles and responsibilities. Who is to be responsible for this curation of content? In practice today, it seems to be the users themselves. Many of these respondents did not see any need to record phone calls because they already took notes on what they deemed as important. An essential element of each person's own research rhythm is the ability to conduct these decisions oneself. While preserving emails for future use may catch information that slips through the cracks, it also second-guesses the researcher's own judgements and risks making a mountain of material that will never be accessed again. What are the implications for institutions in terms of laying out roles and responsibilities in policies?

The concept at the core of this is "value". All of these questions regarding what is important and what should be preserved (be it blogs, emails, or early drafts) require some sort of value judgement. All research work is personal, and these value judgements will be different for everyone, as these interviews show. Some people think preserving intermediate notes would be incredibly useful because these notes carry value; others see it is a frivolous activity. A systematic approach to preservation is more likely to take a "save everything" mentality, but people may be reluctant to adapt such systems because their own judgements are subjective, not systematic.

These interviews demonstrate the need for preservation technology that builds plans and executes workflows to meet the goals and policies of individuals, research groups and institutions. Planets has made progress in this direction with the integration of conceptual and machine interpretable models that support the articulation of policy and preservation planning. Clearly, there is still important work to be done to make preservation activities, including selection and access, transparently aware of, and informed by, the goals that content owners and stewards have.

The structural (or as some would say, bureaucratic) boundaries of institutions have a much greater effect on academic research practices than we would like to admit. The everyday realities of funding research can place limitations on how expansive and complete any preservation system can be. Though the cost of disk storage is always plummeting, preservation is an active process requiring some level of human cost (as the researcher quoted on page 16 is aware). As researchers adapt to the new realities of the digital environment (at different rates, as we have seen), deeply-rooted ideas about accessibility, availability and re-use may change more slowly -- or not at all. Preservation systems must intersect with fiscal reality, and retroactively adapting ongoing projects to new preservation goals may cause friction and consternation.

One common trait shared by all of these interview subjects is their passion for their work. Though everyone had different levels of satisfaction with the University management, few indicated any such passion for funding applications, administrative timetables or other paperwork. Because a preservation system has to be closely bound with institutional policies, it may be difficult to appeal to the desires of the researcher, particularly if the benefits of a preservation system are not readily apparent. A challenge for both the designers of a system and for the institutional implementation of it is to navigate a path to the hearts of the researchers.

The bioinformatics researcher stated above that "boundaries are blurring in publishable work," and this may be the key to developing properly future-thinking preservation structures. Though many of the researchers in this study did not feel that their intermediate notes were a necessary element that required long-term accessibility, the world may decide differently for them. While no one else in this report was as willing to abandon traditional publications as readily as the bioinformatics researcher, the non-academic world is right beside him. Less formal forms of communication have begun to change the way society works; blogs, Twitter and social networking tools have forever changed the relationship between news media and the public, for example. Will academia be a bulwark of tradition, resisting these changes? Or will it inevitably give in to the tide? A preservation system needs to identify these blurred boundaries and attempt to ride their crest, allowing room to adapt for personal taste.

A final theme to point out is the orthodoxy of academic systems. It is not surprising that respondents from scientific disciplines seem to have dipped their toes in new technology before arts, humanities, or social science researchers do. What is surprising is how these respondents were so willing to discard long-held prejudices about content creation and veracity. As mentioned above, technical and experimental data can be tested for accuracy internally in a way that humanities research cannot. But beyond mere citation frameworks, scientific researchers from this interview cohort were the most enthusiastic about social networking and the full communications potential of the Internet, even if it is at the expense of so-called institutional authority. What impact this will have in institutionalised preservation systems remains to be seen, but it is surely an area to watch.

Appendix A   Interview questionnaire

Planets DT/7 questionnaire

"Thank you for participating in the Planets Project. This interview will look at your research within the University. Specifically, we will discuss communication, networks, and the ways in which you use, re-use and adapt information in your own work. We are specifically interested in problems you may have had when trying to re-use older, digital information.

When referring to digital preservation issues, we use the terms 'preservation' and 'long-term' informally; for the purpose of this interview, we mean 'saving indefinitely for future use' when we speak of preserving."

Phrases in italics should be asked to draw out the answer if the respondent has not already addressed these points in their answer. These italicised sections are attempts to open up the questions and encourage anecdotes.

A. Demographics

1. How old are you?

2. What proportion of your time do you spend on research work versus teaching and administrative work? What proportion of your time would you *like* to be spending on each of these?

3. How many years have you worked within research and/or teaching?

4. Could you briefly describe your research area? *Are you working in a group, and if so, are you working as a leader?*

5. Could you name some of the software tools/packages that you use regularly in your research work?

B. Research related communication

1. How important is email communication for your research?

2. Do you believe these emails are valuable enough to save for future researchers to refer to? *Why?*

3. Have there been any situations in which a telephone call or face-to-face communication would have benefited research by being recorded, transcribed or otherwise documented? *Should your telephone-based research communications be documented in some way?*

4. Does your department make it easy to save your research-related communications? What steps do you take to make sure? *How do they make it easier/more difficult?*

5. In what circumstances would you prefer to use printed or analogue materials instead of digital counterparts? *Please give an example of a time when you made this choice.*

C. Data/information to be preserved

1. In your own work (and in your field), what research outputs are the most important for future use, re-use and adaptation by others?

2. Are there elements of your own research data/information that you fear may end up lost? *Do you have a clear idea of the research data/information that are not being preserved in your own work?*

3. Let's think about other information that you use from other groups. Was that material created this year? Are there times when you found out about information (a presentation, a paper, etc.?) that you couldn't get access to? *Adaption and re-use?*

4. Do you think more or less preservation will benefit your work? *Why?*

5 Should intermediate research results be preserved, such as drafts, notes or incomplete tangents -- and why?   What specifically in your own unpublished work do you think might be beneficial for future use?

6 Do you often use the research results of others that are not being formally preserved? *Please give an example. Follow-ups: What challenges did you face in accessing them?   Were there problems with formats or software?*

7. Who should and should not get access to these results, and why?

8. Who do you think should be responsible for the preservation of research outputs? *The University, funders, publishers?*

9. Does your department (or the University) make it easy to preserve your intermediate research results? How do they make it easy or difficult? *Please give an example. Are the researchers outputs that you need being taken care of for the future?*

D. Digital or printed data/information

1. In what ways do you find digital data/information easier to access than printed data/information? *In what ways is it more difficult?*

2. In what ways do you trust printed data/information more than digital data/information?

3. Have you experienced problems accessing digital data/information because it was old? *Please describe some of these experiences.*

4. Have you ever found anything that contributes to your research on the personal websites or blog of a researcher? *Please elaborate.*

5. Do you believe that personal sites and other digital artefacts such as blogs, wikis etc. should be preserved?