

A global approach to a global resource:
An investigation into the archiving of Web 2.0
records around the world

Ruth Kusionowicz

A dissertation submitted to the University of Wales in partial fulfilment of the requirements for the degree of Magister in Scientia Economica (MSc Econ) under Alternative Regulations

Department of Information Studies

Aberystwyth University

Submitted 2013

Abstract

The purpose of this dissertation was to explore how different countries are approaching archiving Web 2.0 records and to ascertain if a global approach would be appropriate or practical. Archives that capture and preserve web-based material from three different countries were approached. Data was collected by analysing their policy documents and conducting semi-structured interviews by email. The results showed that there were no significant differences in the approaches and attitudes of different countries when archiving Web 2.0 records. Participants agreed on the lack of direct influence of society on appraisal decisions, the potential of Web 2.0 records for new types of researchers, and the need to engage with creators of Web 2.0 software and records to enable these resources to be archived successfully. All participants expressed concerns with various technical and practical issues, but didn't see these as complete barriers to archiving web-based material. Legislation and legal issues were regarded as less of an active concern by respondents than was expected. Collaboration and cooperation was a recurrent theme and played a role in various stages and areas of archiving Web 2.0 records. It was concluded that the different archival traditions of various countries would not cause a problem in establishing collaborative practices, and that a global approach to archiving Web 2.0 records would not only be possible, but would actively be encouraged and embraced by institutions.

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed (candidate)

Date

STATEMENT 1

This work is the result of my own investigations, except where otherwise stated. Where ***correction services** have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged (e.g. by footnotes giving explicit references).

A bibliography is appended.

Signed (candidate)

Date

[*this refers to the extent to which the text has been corrected by others]

STATEMENT 2

I hereby give consent for my work, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed (candidate)

Date

Acknowledgements

Firstly I would like to express my deep gratitude to Kirsten Ferguson-Boucher, my dissertation supervisor, for all the help, advice, inspiration and encouragement she provided throughout the research and write-up process.

A huge amount of thanks must go to all the participants who agreed to take part in my study and made the effort to get their answers and policies to me in a timely fashion. Without their input this dissertation wouldn't have been possible.

I would also like to thank Dr Julie Mathias, my personal tutor, for not only developing my skills as an archivist, but also for her constant support and helpful advice throughout the duration of my Masters. And to Aberystwyth University, for giving me the opportunity to study on such a fantastic course.

I can't forget my course mates and fellow baby archivists, who's collective support, mutual moans and bizarre sense of humour have kept me sane. I couldn't have survived the course without you!

And finally, a big thank you to family and friends who have helped, encouraged or supported me while on my course and especially while writing this dissertation.

Table of contents

	Page
Abstract	i
Declaration	ii
Acknowledgements	iii
Table of contents	iv
<u>1. Introduction</u>	1
1.1 Background of the topic	1
1.2 Definitions	2
1.3 Aims and objectives	2
1.4 Scope and limitations	3
1.5 Structure	3
<u>2. Literature review</u>	5
2.1 The concept of 'value'	5
2.2 The influence of society on appraisal	8
2.3 The influence of technology on appraisal	8
2.4 The use of policies in appraisal	10
2.5 Future researchers and their desires	11
2.6 Creators and users of Web 2.0 records	11
2.7 Practical issues with digital and web-based records	13
2.8 Legal issues with web archiving	15
2.9 Appraisal in different countries	16
2.10 Collaboration	19

<u>3. Methodology</u>	22
3.1 Methodological approach	22
3.2 Methods	22
3.2.1 Literature review	22
3.2.2 Selection of participants	23
3.2.3 Interviews	24
3.2.4 Data analysis	25
3.3 Limitations and lessons learned	25
3.3.1 Literature review	25
3.3.2 Selection of participants	26
3.3.3 Interviews	26
3.3.4 Data analysis	27
3.4 Methods summary	27
<u>4. Results</u>	28
4.1 Collection policies	28
4.1.1 UK policies	28
4.1.2 USA policies	29
4.1.3 Australia policies	29
4.1.4 General observations	30
4.2 Interviews	30
4.2.1 Types of Web 2.0 records	30
4.2.2 How Web 2.0 records are selected	31
4.2.3 The use of policies	32

4.2.4	Value	32
4.2.5	The influence of society	32
4.2.6	Technological issues	33
4.2.7	Access	33
4.2.8	Amount of material	33
4.2.9	Legal issues	34
4.2.10	Future researchers	34
4.2.11	Collaboration	34
4.2.12	Issues with creators	35
4.3	Results summary	35
<u>5. Discussion</u>		37
5.1	Discussion of the results	37
5.2	Discussion summary	42
<u>6. Conclusion</u>		43
6.1	Review of this study	43
6.2	Lessons learned	44
6.3	Future research	45
<u>Bibliography</u>		46
<u>Appendices</u>		49
A	Participation request	49
B	Interview schedule	50
C	Example of theme framework	52

Chapter 1: Introduction

1.1 Background of the topic

The Internet was in use from the late 1970s and in 1993 a new service called the World Wide Web (www) was launched (Vossen & Hagemann, 2007, p.2). According to Vossen and Hagemann (pp.40-41) there were four stages in the development of the Internet, from a research and academic focus (1980-1991), to early public uses (1990-1997), then achieving domestic and international critical mass (1998-2005), before becoming a mature, universal and worldwide accessible network that continues to grow and advance technologically rapidly (2006-present).

The term Web 2.0 is credited to Tim O'Reilly, who coined it in 2003 and later defined it in his blog in 2005 (Black, 2007, pp.1-2). It referred to a set of technologies and principles for utilising the Web, giving the ability to integrate information in new ways, the desire to harness distributed knowledge, and the need to engage users as co-developers (Yakel, 2006, p.160). The Web 2.0 term was used to signify the Web had begun fundamental changes in the way people were able to use it, moving from a medium for passively sharing published content to a vehicle for people to interact with and re-use that content (Theimer, 2010, pp.9-10). Although the concept and the term 'Web 2.0' has elicited much debate (Black, 2007, p.3), the prevailing idea is that previously passive recipients of content can now engage with, combine and share information in new ways, and contribute, not just view, content (Samouelian, 2009, p.43).

Web archiving is still in its infancy, in comparison to archiving physical records. There are few institutions who attempt to capture web-based material, and its rapid changing nature and frequent technological advances all play a part in causing new challenges and points for discussion in the archival world. The development of Web 2.0 resources have enhanced these issues and created yet more problems for trying to record society's online presence and the new way the world works and connects. As Web 2.0 records are produced on a global scale and interactions are no longer constricted by location, perhaps records professionals from different countries would benefit from approaching the issue of archiving these records together. However, their archival traditions have developed separately and are influenced by various factors due to their differing cultures, beliefs and governing structure. This could therefore be a major obstacle to finding a global approach to archiving Web 2.0

records, despite their universal nature. This study has been carried out to discover how different countries have approached the problem of Web 2.0 and whether a collaborative attitude would be realistic, practical and effective.

1.2 Definitions

There is still debate about what certain terms in relation to this topic mean. Therefore, for the purposes of this study, the following terms have been defined and these definitions should be referred to when reading this research.

Web 2.0:

a perceived second generation of web-based services that emphasise online collaboration and sharing among users; is the network as a platform; and is interactive and encourages participation, where end users upload as well as download.

Digital curation:

maintaining and adding value to a trusted body of digital research data for current and future use, encompassing the active management throughout the research lifecycle

1.3 Aims and objectives

The main aim of this study is to explore how different countries are approaching archiving Web 2.0 records and ascertain if there is a global approach which can be identified. The nature of the problem to be investigated is how archives choose, collect and appraise Web 2.0 records.

The objectives of this study are:

- To discover which kinds of Web 2.0 records are being preserved by archives.
- To establish if there are similarities or differences between countries' approaches and why.
- To observe how cultural, regulatory and technical frameworks impact the curation of Web 2.0 records.

- To ascertain if a global approach to preserving Web 2.0 records would be appropriate or practical.
- To make recommendations for appropriate approaches for archives.
- To identify areas for future research.

1.4 Scope and limitations

The scope of the research is to examine the current archiving practices of three countries that have a strong archival tradition and where English is the primary language, focusing on Web 2.0 records. The methodology employed was a combination of semi-structured email interviews and an examination of policy documentation.

The limitations of this study are mainly concerned with the location and number of archives analysed. Although the aim of this research involves looking at a global approach to archiving Web 2.0 records, only three countries have had data collected from them. They are all Western countries, whose main language is English, and who generally have a high rate of regular internet usage among their citizens, businesses and government. This could suggest that the findings therefore cannot be applied to other countries whose societies don't fit these criteria. The limitations of this study are further discussed in Chapter 3.

1.5 Structure

To address the aims and objectives outlined above, a qualitative methodological approach was taken. Three countries were focused on, namely the UK, USA and Australia, and data was collected from two archives in each country that have or accept Web 2.0 records into their collections. For each of these archives, an analysis of their policy documents was undertaken, and a suitable representative was interviewed on their institution's approach to archiving Web 2.0 records and their general thoughts on Web 2.0 resources.

A thorough literature review, presented in Chapter 2, reveals the main challenges in archiving and appraising web-based content, the differences in archival approaches of different countries, and the use of collaboration within the archival community. Chapter 3 introduces in detail the methodologies used for the literature review, collection of data and

analysis. The collected qualitative data was analysed and categorised into different themes, which are presented in Chapter 4. The outcomes of the research are discussed in Chapter 5, with respect to the different hypotheses discovered in the literature. In Chapter 6, final conclusions are drawn and recommendations for future research are made.

The planning and implementation of this research follows the Aberystwyth University policies for Ethical Practice in Research and DIS Ethics Policy, which are based on the code of professional practice set out by the British Sociological Association.

Chapter 2: Literature Review

The purpose of the literature review was to discover previous research and studies that have been carried out in relation to archival appraisal, the different archive approaches of different countries, and the issues surrounding the curation of Web 2.0 records. The information gathered was then used to determine where knowledge was lacking and where further research was necessary. There were limitations in the literature that were quickly identified, particularly regarding material about Web 2.0, which tended to focus on this as a form of access to archives or as outreach, rather than as a type of record worthy of preservation. There are more details about the limitations in Chapter 3, as well as an explanation about how suitable literature was found.

Various themes were identified in the reading, which have been separated into sections in this chapter. One theme that is not discussed in detail is archive appraisal, which is a concept that has been debated since the beginnings of the profession, and sparked copious amounts of writing on the subject. Unfortunately, there is not the space here to go into all the arguments, discussions, practices and theories that have been presented over the years. However, a few of the principles and issues that are particularly relevant to the topic of this study have been touched upon.

2.1 The concept of ‘value’

When talking about appraisal, archivists almost always use the word ‘value’ (Boles, 1987, p.367). The idea of value being integral to appraisal began with Schellenberg (1956, p.28), who’s approach was based on the concepts of primary and secondary values, and evidential and informational values of records (p.139). He realised that appraisal was a subjective process, with informational values being relative to time and place, and that “what is valuable to one archival institution may be valueless to another”. He believed complete consistency in judging informational value was undesirable as it was impossible to achieve, but thought that “diverse judgements...may well assure a more adequate social documentation” (p.149).

Schellenberg’s ideas have since been hugely debated among archivists. Lockwood (1990, p.395) believes his approach needs to be adapted, as when he was writing the methodology and subject content of the field of history had remained relatively consistent for

nearly 50 years. She says that evidential value is relatively easy to access, but informational values can cause widely differing judgements (p.397). When assessing informational value, the training, areas of interest and philosophies of the individual archivist have the most potential influence (p.398). Newton (2001, p.89) states that there is a lack of definition and a vague interpretation of the 'value' of a record in relation to secondary informational and research uses. Maclean (1995, p.75) believes that the valueless should be eliminated when deciding what should be preserved, rather than attempt to select the valuable. Bailey (2008, p.92) indicates that archives are filled with items considered largely trivial at the time of creation, but represent a goldmine today for historian or researcher. However, he reiterates that not all information is of equal value (p.106). He recognises that "while all records are information, not all information sources are records" (p.59), but believes that this mantra should be replaced with "all information sources are potentially records" (p.65). According to Eastwood (2002, p.65), "we build our decisions on the knowledge of the value of past preservation of archives to project what will be valuable in the future". However, post-modernists say archivists should acknowledge that records only have the value attributed to them by the people who act on them, including the archivists who appraise them (p.62). Reed (2003, p.74) believes that inherent in the notion of the evaluation of records is the notion of framing or shaping the form the archival record will take for future generations. Neumayer and Rauber (2007, p. 1) warn that "in trying to keep the most important or most valuable content, appraisal actively favours mainstream values, whilst sub cultural influences are effectively eradicated". McLoughlin sums up the problem with using the concept of value to appraise, stating that "even with guidelines and criteria, the attribution of value to records is subjective, entirely dependent on the framework, ideology, context, training and background of the person(s) making that determination" (Newton, 2001, p.92).

Determining the 'value' of Web 2.0 records can be very difficult. A piece of 'useless' information may prove useful in a yet unknown future combination of data (Bailey, 2008, p.100). "We can never know which of our children will go on to become world leaders and global icons", so how can we predict which blogs or social network profiles are best to keep (p.93)? As Pymm (2010, p.22) points out, "the price for getting unfiltered public input is accepting that it will be wide ranging and no doubt contradictory and arbitrary". Spam emails may not be worth preserving, but the fact they existed should be kept for future generations (Neumayer & Rawber, 2007, p.2). A future wide-scale attitudinal research project could use a

few randomly selected samples of blogs rather than preserve every blog ever written (Bailey, 2008, p.93).

Despite these difficult and complex issues, the potential value of Web 2.0 records cannot be ignored. For example, the early threads of Usenet (archived by Google) show the evolution of online dialogue (Cahill, 2009, p.13). Organisations are increasingly recognising the value of keeping records of websites and social media sites (Dionne, 2012). Dionne gives three main reasons to save websites as records – they're an organisation's public face; they provide context and relationships of other records; and they capture a picture of an organisation's presence. Eastwood (2002, p.68) believes that records of private provenance often complement the public record. Private archives reveal a good deal about how democratic freedoms are enjoyed and show the kinds of human action and creation that only individuals can make and leave residue of. Even Jenkinson (1980, p.243) agreed with this in 1948, recognising the need to study the faring of ordinary people for the enlightenment of our own conduct today. As Bailey (2008, p.32) says, "the ease of content creation and dissemination offered by Web 2.0 technology empowers the gifted amateur, the iconoclast and the maverick, all of whose voices can offer so much, but which often go largely unheard". Cahill (2009, p.4) agrees, characterising blogging as "a direct report from the front lines, an often deeply personal interpretation of events with a significant historic or political impact". Vossen and Hagemann (2007, p.63) see the value in analysing people's homepages, which can help discover commonalities, common interests, relationships, and entire communities. Richardson and Hessey (2009, p.29) believe Web 2.0 allows individuals to become authors of their own narratives, and that social network sites (SNSs) alter the day-to-day organisation of individual and collective life (p.27). They see SNSs as providing a means of collecting a trail of associations, and that "the virtual world is becoming the way we learn about other life-worlds" (p.34). They state that Facebook "acts as an archive of social relationships and provides a means of recording ongoing interactions" and that it is "a way of archiving the self, storing biography and enhancing social memory" (p.36). Possible projects using an individual's Twitter archive include the creation of a 'virtual diary', as fodder for a Twitter based memoir, or as a digital scrapbook (Greenfield, 2012). However, having an individual's tweets in isolation doesn't recreate a picture of the conversation. Matt Buchanan sees tweets as "nuggets of information and wit" that should be kept somewhere. He champions "reclaiming our tweets", however Greenfield doesn't believe there is really any utility for this beyond fairness.

2.2 The influence of society on appraisal

Libraries and archives have had a long acknowledged role in preserving social memory (Harvey, 2010, p.5). Eastwood (2002, p.66) asserts that “we preserve what works for us in our society”. Boles and Greene (1996, p.310) agree, stating that “it is society that defines what is expected to be found in an archives”. According to Pymm (2010, p.23), “collections are not ‘neutral’ but reflect the society which created them”. Newton (2001, p.95) believes that people relate most to records that remind them of events of their everyday lives, or achievements and tragedies that have affected them. Eastwood (2002, p.61) sees the preservation of archives as playing a cultural role in the community to foster memory and the understanding of the past. He says there is a presupposition that archival appraisal has an underlying goal, and that goal depends on the kind of society in which it is performed (p.60). Brothman says that the values archivists bring to the task of appraisal embody society’s values (p.62). Hans Booms’ 1971 essay outlined the social context of appraisal and appraisers, demonstrating how it determined their decisions (Reed, 2003, p.71). Samuels says the size and scope of a collection should be judged by local needs and constraints, not national norms (Boles, 1987, p.364). Neumayer and Rawber (2007, p.2) believe that appraisal has now been made important by communities for a different application – “to forget rather than to keep”. Newton (2001, p.90) seems to agree with this, highlighting the ongoing plea by a variety of users for some sort of involvement in decision-making in records appraisal.

Society’s use of the internet and digital material has also had an effect on appraisal. Bailey (2008, p.12) believes “we now have a new generation emerging who would far rather *search for* than *manage* information, and who value quantity over quality”. This therefore means that “the whole concept of selective retention and the deletion of data is now alien, not only to the IT industry, but to popular culture and the society in which we live” (p.100). Evans (2007, p.395) describes “archives of the people” as products of customers deciding what to digitise, and Benkler sees the networked digital environment as an opportunity for intelligent people to satisfy their curiosity and contribute to society (p.397).

2.3 The influence of technology on appraisal

Cox notes that appraisal approaches developed for data and digital objects are likely to be different (Harvey, 2010, p.132). Digital storage has removed many of the obstacles to

universal retention that were unavoidable with physical records (Bailey, 2008, p.98). However, digital storage consumes vast amounts of electricity (p.105) which is expensive and impractical, and some records will still be rejected because they aren't of sufficient quality to meet record evaluative criteria (Boles, 1987, p.363). According to Bailey (2008, p.99), we now have the tools to enable us to find the required needle of information in a digital haystack of limitless storage. He believes IT departments would rather spend money on buying a new server than spend time and effort appraising and sorting the contents of the existing one (p.101). He therefore suggests that "perhaps we should be ready to admit that keeping everything may now be an option, rather than simply dismissing it as impossible or just plain undesirable". However, he also says that although it may be possible to store vast amounts of meaningless and ephemeral information, this does not necessarily mean that we should (p.105). He asserts that appraisal is still required in the Web 2.0 world, but on a scale we've never had to contemplate before (p.111). Lockwood (1990, p.396) says that keeping everything is not economically feasible and is undesirable, as such bulk is a hindrance to future research. Neumayer and Rawber (2007, p.2) believe that appraisal should be abandoned in its current form. They say "it is more than clear that it will be impossible to keep everything forever" (p.1). However, in the digital world, they suggest that random selection may be the best approach, as this would reduce the volume of data to be dealt with and should maintain an authentic, unbiased view of history. They believe the main advantages of this strategy are simplicity and cost effectiveness, privacy protection, preservation of a fair and unbiased view of contemporary life, and a future proof process (p.2). Bailey (2008, p.114) suggests that macro appraisal is a more scalable approach for Web 2.0, particularly as it places little value on the consideration of content (p.115). Pymm (2010, p. 23) believes that, without the teachings and interference of professions, the resultant holdings may be idiosyncratic and biased in many ways, but will portray a range of unorthodox viewpoints of potential value to future researchers. However, he warns that in taking this approach, individuals with strong opinions and a wish to promote their viewpoints and agendas could skew an archive's holdings (p.22). Harvey states that

Appraisal and selection are desirable because they improve management of resource limitations by reducing the quantity of data and digital objects maintained and, thus, increase the economic validity of long-term retention of data...This leads to better digital curation.

(Harvey, 2010, p.147)

However, Bailey (2008, p.102) suggests that, with the ability to store everything we produce digitally, “perhaps the rationale for appraisal has finally run its course”.

2.4 The use of policies in appraisal

Boles (1987, p.357) believes that the archival community might be better served if the traditional language archivists have used to discuss appraisal was replaced with a new vocabulary that emphasises policy. He says that decisions on collection policy etc. are closely linked to appraisal because they determine the basic character of what will be found in an archive (p.359). The collection policy “is the first level of appraisal, defining a basic interest in an area or topic to the exclusion of other areas and topics” (p.362). Boles asserts that “appraisal goals must be rationally established and described to create policy” (p.368). This is important, as most of what occurs in archival appraisal is the establishment and implementation of policy decisions, not value judgements (p.367). Harvey (2010, p.85) agrees with this, believing that, to ensure decisions are consistent, appraising and selection requires planning and the development of selection and retention policies. Boles (1987, pp.367-368) also says that considering appraisal as policy stops it being isolated from the other archival activities, and helps rationalise the entire appraisal process.

The importance of planning and policies seems to be amplified when dealing with digital and web-based records. Harvey (2010, p.86) believes that developing policies for all aspects of digital curation is vital for its effectiveness. He says having policies in place assists an organisation in developing a digital curation strategy and plan coherent digital curation programs (p.87). He also believes, there should be procedures for curation in place. Harvey says that collecting data needs planning to ensure they are collected with relevant description and representation information in ways that ensure their accuracy (p.85). He states that the preparing of digital objects to add to a digital archive and adding them to a digital archive requires planned procedures to be in place. He also believes the use of a data management plan template helps ensure the aims of curation are achieved through defining responsibilities, creating a high-quality archive, supporting data creators and users, and adhering to conditions associated with use and deposition of the data.

2.5 Future researchers and their desires

It is difficult to predict who the audience of a Web 2.0 project will be and what content they'll like best (Theimer, 2010, p.219). Researchers in the future will work differently (Harvey, 2010, p.4). For example, it wasn't until the late 1960s that the new field of social history was developed (Lockwood, 1990, p.395), and in recent decades there has been an increase in the popularity of genealogy and local history (Twigge, 2003, p.17). According to Lockwood (1990, p.395), new fields of history suggested new ways to use materials previously thought to be of minimal value, and prompted historians to look for new kinds of records. Jenkinson (1980, p.243) believed the possibilities of archives to be "inexhaustible". In the last two to three years, there has been considerable interest in concepts of open science and open research which can be achieved using Web 2.0 tools, and according to a Research Information Network Report (2010, pp.5-6), the majority of researchers are making at least occasional use of one or more Web 2.0 tools or services for purposes related to their research.

2.6 Creators and users of Web 2.0 records

There have been a number of studies in a variety of disciplines written about who uses and creates Web 2.0 records and for what purposes. Theimer (2010, p.16) states that from 2005 to 2009, the share of adult internet users who have a profile on an SNS has risen from 8% to 35%. She says that the use of Web 2.0 tools has steadily been increasing among older web users and this has brought a huge rise in user expectations (p.221). According to Hargittai's research (2008, p.279), young people are most likely to participate on some SNSs and are more wired than their older counterparts. Facebook is the most popular service among students (p.282). There are minor ethnical differences in the SNSs that students choose to use and there is a relationship between parental education and the use of some SNSs. The research shows women are more likely to use SNSs than men and older participants use Facebook less (p.283). 'Millenials' is a term found in some of the research. Originally coined by Stephen Abram and Judy Luther, it describes people born between 1982 and 2002 who were "born with the chip" (Black, 2007, p.5). Computers and devices are part of their daily life and they were early and heavy adopters of IM and blogs. They are a collaborative group, nomadic, multitaskers, and are concerned with finding information, not the format it is in. They are more affluent, better educated, more ethnically diverse and focused on teamwork

and achievement. For them, convenience and ease of access are of highest importance (Samouelian, 2009, p.69), as pointed out by Bailey (2008, p.99) who says end users appear to value the accessibility and currency of Wikipedia over the authoritativeness of *Encyclopedia Britannica* for example. Black (2007, p.5) says that generational issues play a big role in web use, but to Millennials

Web 2.0 is not a change. It is the expected way of being.

(Black, 2007, p.6)

Bailey (2008, p.9) sights a report on the use of Web 2.0 in UK Higher Education that showed examples of academics using Facebook as the preferred method of keeping in contact with their students. He believes that the ways staff create potential records may challenge our assumptions and methodologies “but this is our problem, not theirs” (p.65).

According to Vossen and Hagemann (2007, p.7), the world of today has become flat. People from opposite sides of the globe can now interact, play, do business and collaborate thanks to the world wide web. They “have gotten used to the Web as a communication medium, a socialization platform, a discussion forum, a business platform, a storage device for their diaries, and as a constantly growing and expanding encyclopedia” (p.49). Vossen and Hagemann say Web 2.0 tools now “enable the creation of communication and relationships among individuals as well as among groups” (p.57). They give six reasons why people choose to blog – to broadcast personal news to the world; to share a passion or a hobby; to find a new job; to write about a current job; to link with customers; and to get new customers (p.51). Considerable change in the perception and usage of the internet has opened the door for this present-day willingness of people to share all kinds of information, private or otherwise (p.59). Cahill (2009, p.40) goes further, noting the changes from Web 1.0 to 2.0, stating that “end users no longer see themselves as passive absorbers of the information they need, but as creators and shapers of that information”. Pymm (2010, p.17) agrees and says that users are now involved in all aspects of the digital archive. Web 2.0 is now being utilised for new tasks. According to a Research Information Network Report (2010, p.5), researchers see Web 2.0 tools and services as having their own distinctive role for specific purposes and at particular stages of research. 13% of respondents said they use social networking services for purposes related to their work (p.6), although only 5% of respondents openly publish data and work in progress using blogs and other tools (p.31). The report also states that researchers who use Web 2.0 services come from all age groups and levels of seniority (p.47), and the differences in use between various demographic groups are relatively small

(p.6). Web 2.0 records are being used in background information searches, which look at ‘non-traditional’ online sites to help shortlist suitable candidates for a job (Cohen, 2009, p.16). Users also now expect applications on websites, so they are able to comment on a blog or a picture rather than writing a letter for example (Samouelian, 2009, p.64).

2.7 Practical issues with digital and web-based records

As Harvey (2010, pp.3-4) states, “the increasingly digital world that we all inhabit is changing the ways we work and play” and this has led to the generation of massive quantities of data in all areas of our lives. Evans (2007, p.388) agrees, saying that the Information Age means there are many more records to inventory, appraise, accession and process, but that it also suggests to the rest of the world that all information will be easily and quickly available. In a 2009 study of researchers in life sciences, it was stated that

The sheer volume of data and information that is now being produced, and expected to be produced in the future, is a cause for concern. Researchers fear that there will be too much data to handle, process, or even look at.

(Harvey, 2010, p.134)

Harvey (p.93) warns that digital curation is a field that changes rapidly and has few common understandings so far.

There are various physical and technical problems with archiving digital records, which are well documented, but there is not space here to discuss them in depth. However, certain issues are particularly relevant to this study. Pymm (2010, p.22) suggests that the technological challenges of establishing and maintaining digital archives are perhaps of less concern today. Crook (2009, p.834) also believes that when gathering files, most of the file types’ problems have been overcome, but that multimedia content remains an issue. In particular, the delivery mechanism and embedding technologies that make videos useful for users are a challenge in archiving. Yakel (2006, p.159) states that “archives have been less experimental in recent years and slow to adopt some of the more interactive features that support social navigation”. Harvey (2010, p.4) says that data is at risk from many factors, including technology obsolescence, technology fragility, a lack of understanding about what constitutes good practice, inadequate resources, and uncertainties about the best organisational infrastructures to achieve effective digital curation. He lists 13 threats to digital continuity (p.9), and notes that responses to these based on traditional preservation

approaches don't work (p.10). He says digital data must be managed from the point it is created if survival is to be ensured. He also believes most data archiving and digital preservation practices have been developed for static documents, so don't transfer successfully to more complex data (p.8). Harvey notes that data sets and digital objects must also be accompanied by relevant description and representation info (p.134). Theimer (2010, p.222) agrees, but suggests that the nature of digital materials means they lend themselves to being accompanied by information about repository, collection, records series etc., much more than analogue materials do.

When considering archiving the web, there are many practical issues with deciding which parts and elements to preserve. Crook (2009, p.833) notes that the large amount of data collected makes any attempt at quality assessment of individual websites difficult. He observes that a significant amount of Australian web sites are being archived, but that we can't say to what extent this is a comprehensive or complete collection (p.835). He warns that the "selection of which web sites to crawl is an often misunderstood activity and can take up surprisingly large amounts of time" (p.833). Dionne (2012, p.1) notes that a repository's crawling policy may even differ on an individual website basis. *PoWR: The Preservation of Web Resources Handbook* gives 3 approaches to selecting websites for archiving – bulk or domain harvesting, criteria-based selection, and event-based selection (Harvey, 2010, p.142). But Harvey asks if we want to preserve the full experience if possible when preserving web resources? Upward, McKemmish and Reed (2011, p.204) also pose the question whether archiving and recordkeeping practices are needed to manage the complex transactionality of the web, or to preserve just a few elements from it? Theimer (2010, p.214-215) considers the variety of these elements that can be preserved. For example, for blogs, an archive may preserve individual posts as text, accompanying images and audiofiles, comments, and/or screenshots of the blogpage to capture the look and feel of the template. With a wiki, a record of all the changes made could be kept, or content captured at designated intervals. Theimer says that the web has transformed what is required to 'publish' information, as anyone with an interest in a topic can publish a blog, contribute to a wiki, start a webpage or share information (p.5). Archivists struggle to capture comments and uploads and integrate them into their systems (Samouelian, 2009, p.65). The rapid development and proliferation of Web 2.0 services means it is hard to keep track of them or assess their potential benefits (If you build it, will they come?, 2010, p.7). Crook (2009, p.835) believes that "Web archiving...is never, it seems, going to become an area whereby established practices or protocols in

collection development will ever be fully established or maintained”. However, Theimer (2010, p.215) points out that the discussion of methods for preserving Web 2.0 records has only just begun.

Pymm (2010, p.13) notes that the rapid move to digital objects as the primary form in which records are produced means the traditional concept of archives is changing. For example, digital archives are non-traditional by comprising only digital content, but traditional in that they’re created and ‘curated’ by professionals (p.16). Upward, McKemmish and Reed (2011, p.232) believe that digital technologies and social networking can support continuum-based frameworks for implementing participating recordkeeping and archival models. However, they highlight the professional challenge to create and sustain records in multiple contexts of creation, interpretation and reuse (p.225). They say that the past approaches of controlling what we do make no practical contact with ways of dealing with the massive expansion in our information and communication apparatuses (p.226). They believe new recordkeeping methods and techniques are evolving, building on a growing knowledge and understanding of recordkeeping metadata (p.227). Yakel (2006, p.163) states that “reconceptualizing the role of the archivist and the researcher is hard, whether this is in terms of reference, description, or collection development”. Theimer (2010, p.222) believes the issue of preservation of provenance is challenging, if not impossible, in the digital world. Upward, McKemmish and Reed (2011, p.219) agree, saying that records now have multiple, simultaneous and parallel provenance, and the notion of the original becomes moot when parts can be replicated identically and rendered in various displaying devices (p.224). However, Themier (2010, p.222) asks “how much effort should we put into preserving provenance or providing context for our collections on the web if most users neither understand nor want it?”.

2.8 Legal issues with web archiving

Legal issues can particularly have an effect on archiving the web. Web 2.0 has brought challenges for archives and historical organisations, such as maintaining archival principles in the digital world and copyright concerns (Theimer, 2010, p.6). Theimer (2010, p.15) points out that information posted in any web forum can be shared, even if the forum itself is private. Neumayer and Rawber (2007, p.2) suggest that random selection may offer a high

level of privacy protection. However, Cahill (2009, p.22) says that in the Web 2.0 environment, our notion of privacy is changing. The Freedom of Information Act and Data Protection Act make no distinction between whether information forms part of a formal record or not (Bailey, 2008, p.63). Theimer (2010, p.210) believes that the main legal issue to consider with Web 2.0 projects is copyright. Harvey (2010, p.134) gives this and intellectual property rights (IPR) as examples of legal restrictions that prevent the curation of data. Australian copyright law says online publications are not included within legal deposit (Crook, 2009, p.833). Crook (p.835) gives two main reasons for not archiving virtual worlds & social networks in particular – content generally has copyright and privacy features that disallow it; and the nature of the resource places it outside public internet. Harvey (2010, p.146) says that disposal decisions might result from concerns like violation of copyright or other legal requirements, national security, confirmation that research has been falsified, or concerns over the confidentiality of the data. However, according to Dionne (2012, p.2), there are actually very few legal decisions that impede the progress of web archiving at this time.

2.9 Appraisal in different countries

According to Reed (2003, p.63), there are several different traditions of appraisal, most notably European, UK and USA. She says that “techniques and methods of appraisal are bound to the context of the theory or traditions of archival practice in which they take place”. Contents of policies differ according to an organisation’s mission and requirements (Harvey, 2010, p.87) and different traditions of practice determine who will undertake appraisal decision making (Reed, 2003, p.70).

In British and some European traditions, records are archives and archives are records – there is no conceptual difference (p.65). This derives from the teachings of Jenkinson. In the 1940s, the phased review approach was adopted in Britain (p.72). Then, following the recommendations of the Grigg Report of 1954, the Public Record Office implemented a system of regulated life cycle reviews which comprised the basis of their practices during the 1950s to 1990s (p.69). In 1999, they introduced a review of their selection policy that identified 8 themes, grouped under two headings – policy and administrative processes of the state, and interaction of the state with its citizens (Twigge, 2003, p.17). In the early 20th century, the Public Record Office embarked on a major consultation exercise to involve the

public more closely in the process of selecting records for permanent preservation (p.16). In regards to web-based and digital records, there are now prominent groups in the UK concerned with digital curation, such as the Digital Preservation Coalition (DPC) and the Digital Curation Centre (DCC), who carry out activities such as persistent lobbying of influential stakeholders to take digital curation seriously (Harvey, 2010, p.98).

Reed (2003, p.73) says that the practice of the National Archives of the USA is encapsulated and expounded in Schellenberg's *Modern Archives*, and according to Boles and Young (1985, p.122), his distinction between evidential and informational values still remain the cornerstone of appraisal in the States. Lockwood (1990, p.396) agrees, indicating that the National Archives doesn't have a unified retention or collection policy, the staff rely on Schellenberg's *Appraisal of Modern Public Records* for theoretical guidance. The American approach of explicit attention to records thought to be of interest to the future researcher was contrary to the European tradition embodied in Jenkinson (Reed, 2003, p.71). Duranti went as far as to say that American appraisal is in fundamental conflict with traditional archival theory and the very nature of archives (Boles & Greene, 1996, p.299). In American practice, responsibility for appraisal decisions was distributed, with agencies being responsible for determination of primary value and archivists responsible for determination of secondary values (Reed, 2003, p.69). The technique of disposal scheduling was derived from American practice in the late 1930s to 1940s (p.73). During the 1980s, collecting archives borrowed from the library tradition of collection development, and the notion of active and targeted documentary arose in USA, which allowed the development of the concept of documentation strategies (p.75). The development of records management as a separate but related professional group in USA also affected the implementation of appraisal, and one of its lasting legacies is the development of formalised criteria with which to measure records against (p.73). In regards to web archiving, the Wayback machine at the Internet Archive has been crawling the web on a daily basis since 1996 (Vossen & Hagemann, 2007, p.14). The US National Library of Medicine has developed a methodology for appraising its web documents and has developed four 'permanence levels' – permanent: unchanging content, permanent: stable content, permanent: dynamic content, and permanence guaranteed (Harvey 2010, p.143). The September 11 Digital Archive contains no physical materials, and was the first major digital object acquired by Library of Congress for 'permanent' preservation (Pymm, 2010, pp.19-20).

Australian practice has evolved within the context of inherited approaches from Europe, UK and USA and formed a hybrid model (Reed, 2003, p.63). According to Reed (p.64), Australian recordkeeping traditions are clearly based on the influence of Jenkinson, however Australian archivists deal with essentially 20th century records. Ian MacLean said Australia agrees with Jenkinson that records and archives are the same thing, whether eventually preserved permanently or not (p.66). Reed claims that the subjectivity of Schellenberg's approach requires models to frame practices for active documenting, which are absent in the Australian archival landscape (p.74). She says that in Australia, the two roles of administrative importance and cultural importance cannot seem to sit in harmony (p.66). The descriptions of practices of the Archives Office of New South Wales prior to 1977 shows a reactive process of appraisal based solely on the response to agencies proposing material they wished to dispose of (p.69). A very active involvement in regulating current recordkeeping was a feature of Commonwealth Archives Office from the mid-1950s. In the 1970s they made a shift in emphasis from the need to justify destruction of records to a requirement involving critical appraisal of material to justify its retention (pp.69-70). Australian Archives staff were asserting that "too much is made of the 'judgement of the past of the professional archivist'" (p.70). What followed was an era of relative resource affluence which has now largely ceased, where abundance of space and fuzzy appraisal decision making criteria led to the appraisal regime in Australia being generally ad hoc (p.72). Australian archivists entered the 1990s profoundly confused, which led to a period of introspection on the nature of records (p.76). A recent trend is for Australian archival institutions to now define their 'collecting policies' (p.75) and continuum archivists in Australia have argued for the integration of current and historical recordkeeping (Upward, McKemmish & Reed, 2011, p.198). In recent years, Australia has been instrumental in providing new ways to look at recordkeeping and approaches to appraisal, such as the DIRKS methodology (Newton, 2001, p.96). Australia has a different culture to many other countries, where indigenous communities are distrustful of archives due to their colonial and post-colonial culture (Upward, McKemmish & Reed, 2011, pp.206-207). Most indigenous interviewees in the Trust and Technology Project viewed all records that related to them as their own records (p.208). Crook (2009, p.835) says that Australia is not gathering a vast amount of individually produced creative content hosted on video, photograph and art hosting web sites, blogs, virtual worlds and social networking sites, but that small directed projects have attempted to gather some of this. The National Library of Australia undertakes web archiving using 3 methodologies – selective archiving within the PANDORA Archive, which

has been archiving Australian web publications since 1996; contracted whole domain harvesting; and utilisation of the Archive-IT service (p.831). Australian copyright law says online publications are not included within legal deposit (p.833), but the National Library of Australia does have arrangements with Flickr and archiving permission from MySpace and YouTube (p.835). However in 2009, the Australian Library and Information Association (ALIA) stated that “Australia lacks an adequately funded national research and cultural digital preservation program” (Pymm, 2010, p.15).

In Germany, a more function/structure based approach was developed to combat the subjectivity in appraisal decision making (Reed, 2003, p.72). In 1937, Prussian archivists rejected ‘finger in the wind’ methods of determining appraisal, which led to more defined appraisal criteria based on the analysis of relative functions and positioning of the creating body (p.71). Prussian archivists were concerned primarily with the ‘quality’ of the record for historical research (p.72). German practice argued that records couldn’t be appraised in isolation, but must be placed in their administrative context. East German archivists were actively experimenting with appraisal during the 1970s and beyond (p.74). During the 1980s, East Germany attempted to implement a focus on the content of records to reflect social processes resulted in the development of a list of 500 events which ought to be documented (p.75). In Iceland, rules on appraisal and disposal were ignored and all records were kept until 1985 when the National Archives Act was passed (Kristinsdottir, 2003, p.199). Due to the provisions of this act, there was an increase in records transferred to the National Archive and appraisal then became significant (p.201). The new system requires the National Archive to give permission to parties to allow them to dispose of any records in their archives and the Act assumes appraisal has already taken place before records are transferred to the National Archives of Iceland (p.202). The appraisal procedure puts emphasis on the content of records, however the Board of the National Archives have a basic rule to preserve all records dating from before 1950 (pp.202-203).

2.10 Collaboration

Eastwood (2002, p.71) states that “a pluralistic society needs pluralism in its archival system”. Schellenberg said that diverse judgements spread the burden of preserving the documentation of a country among its various archival institutions (p.71). According to

Harvey (2010, p.96), collaboration is key to effective curation. He believes collaboration ensures the best use of resources through sharing expertise and experience, and through developing and building technical resources and solutions that can be shared (p.97). He assures that “collaboration is...firmly embedded in digital curation practice” (p.96) and that ideally, both information professionals who curate data and the creators of the data should be involved in developing and applying appraisal and selection criteria (pp.139-140). All communities and stakeholders involved in curation should participate in discussions about the challenges posed (p.96). Upward, McKemmish and Reed (2011, p.235) agree, seeing the need for the archival profession to operate in a more coordinated fashion across the massive breadth of change and novelty involved in digital recordkeeping in online cultures. However, they ask whether archivists are well placed to harness the emerging trend towards ‘glocalism’ - a continuum-oriented term that covers localised diversity and global singularity (p.237)? Harvey (2010, p.93) says there is a need to be fully aware of other activities in the digital curation community on an on-going basis. He believes no single organisation could adequately archive, preserve and provide access to digital materials (p.97). Evans (2007, p.389) agrees, asserting that archivists cannot operate in isolation.

Some countries seem to have embraced the need for collaboration more than others. The implementation of the documentation strategy, which originated as a concept in Australia, completely hinges on the willingness and ability of individual, cooperating archives to accept designated records (Boles, 1987, p.364). The Indigenous population in Australia has encouraged collaboration. Upward, McKemmish and Reed (2011, p.218) believe that “decolonising archival functionality and professional recordkeeping practice can be envisaged as a collaborative, co-creative journey involving Indigenous and archival communities”. They assert that “postmodern and continuum ideas, coupled with new digital and social technologies, and Indigenous ways of knowing, open up exciting possibilities for pluralizing archival functionality, acknowledging parallel recordkeeping universes, or even realizing an archival multiverse, and building shared recordkeeping and archiving spaces that enable the co-existence of different and contested narratives” (p.221). Australia has also embraced web-based projects involving collaboration, such as the PANDORA Archive, which forged a network of nine Australian archiving participants and achieved a collaborative national approach to archiving (Crook, 2009, p.832). American archives have also shown collaboration in regards to web archiving. For the 9/11 Archive, the Library of Congress took on the preservation role while the archive managed access and acquisition (Pymm, 2010,

pp.20-21). The Internet Archive collects records of sites based anywhere in the world, including annual directed crawls to gather what it can of the .au domain (Crook, 2009, p.832). Archive-It is a hosted web archive service provided by them that is used to gather collections of overseas web sites recording particular social and political events (p.833). Archives involved with web archiving can be seen to be embracing collaboration and creating ties with other archives either locally or globally.

Chapter 3: Methodology

To address the aims and objectives outlined in the introduction, data was collected from two archives in three different countries using a qualitative methodological approach. For each of these archives, their policy documents were analysed and a semi-structured interview was undertaken via email with a suitable representative of the institution. When analysing the data, themes and concepts were identified and these were then collated and compared by country.

3.1 Methodological approach

A mostly qualitative methodological approach was selected for this study. It was considered to be more suitable than a quantitative method, as it would allow scope for discussion and unusual responses with open questions, which are useful for exploring new areas of research such as this (Bryman, 2012, p.247); the findings would better represent the point of view of the participants than that of the researcher (p.408); and as there were only six participants, there wouldn't be an unmanageable amount of data to process, reducing the need for coding that is associated with quantitative results. One question in the semi-structured interview did however employ a quantitative approach, in order to discover which types of Web 2.0 records were being preserved in each archive. This was considered to be more likely to gain accurate results than allowing a participant to list types from memory, and would make comparison simpler for this question.

3.2 Methods

3.2.1 Literature review

Before finalising the research questions, a review of existing literature on the subjects of appraisal, Web 2.0 resources as archival records, and archival practices in different countries was conducted. This was done by entering the following combinations of terms into the Library and Information Science Abstracts (LISA) and Library, Information Science and Technology Abstracts databases - 'Web 2.0' and 'archive'; 'social network' and 'archive'; 'Facebook' and 'archive'; 'Twitter' and 'archive'; 'Web 2.0' and 'appraisal'; 'appraisal' and

'traditions'. Suitable articles and published works were identified from these searches. After this initial process, further reading was ascertained from references and the bibliographies in these sources. As this dissertation focused on Web 2.0 resources, it was deemed suitable to include blog entries and online publications in the list of readings. These were mainly discovered through searching 'Web 2.0' and 'archive' on Google, and by relevant links circulated by JiscMail. Themes were identified within the reading, which then formed the focus of the literature review. These themes also provided the areas that the research should concentrate on, especially when designing the interview questions.

3.2.2 Selection of participants

The population for this research was archives that accept or preserve web-based records in their collections. Countries suitable to examine and compare were identified as those whose primary language is English, as this is the researcher's main and only fluent language. The three countries selected were the UK, USA and Australia, as these all have a strong archival tradition and regularly contribute to internationally recognised journals to further the field of archival science. This suggested their theory and practices have an established foundation but that they are also investigating the best way to deal with the relatively recent problem of digital and web-based records.

Purposive sampling was used to select the countries and the individual archives to contact, in order to ensure those sampled were relevant to the research questions (Bryman, 2012, p.418). A search on Google was conducted by typing in the term 'web archive' with either 'UK', 'USA', or 'Australia'; and the list of members of the International Internet Preservation Consortium (IIPC) was searched for members from these three countries. This identified archives whose focus was web-based records and who were nationally significant. An attempt was made to find the member of staff in each selected archive who was involved with acquisition and appraisal of records, by consulting their website. The subject heading for the initial email to request participation was 'FAO Web Archive Appraisal/Collection Manager', to ensure that the message reached the relevant staff member who would be sufficiently able to answer the interview questions. The email also stated that "if you feel someone else within your organisation would be more suited to answering the questions, please forward this message to them".

The initial email to request participation (see Appendix) clearly stated the aim and focus of the study, how the research was to be conducted, the ideal time-frame to return

responses by, and that participation was voluntary, confidential and anonymous. Of the seven archives contacted, six agreed to participate and were then sent the interview questions as an attachment with a request for a copy of or a link to their appraisal and/or collection policies. The seventh archive was non-responsive, despite a follow-up phone call three weeks after the initial request for participation.

3.2.3 Interviews

A semi-structured interview was used as the main method in this research. This was conducted via email, mainly due to the difficulty of geography, with participants being located in different countries far apart. The use of email allowed participants to answer the questions at a time convenient to them, and for their replies to be more detailed and considered than if conducted face-to-face or by telephone (Bryman, 2012, p.669). It also meant that answers were recorded accurately, removing the potential human error in note taking and transcription. A semi-structured interview conducted by email allowed for initial questions to be standardised, but for participants to answer in the order they preferred with a great deal of leeway in how to reply. It also meant follow-up questions could then be asked after the researcher had properly absorbed the initial responses.

The majority of questions asked in the semi-structured interviews were open, allowing for respondents to answer in their own terms, encourage discussion and thought about the topic, and allow for unusual responses to be given. The advantages of open questions are that they don't suggest certain kinds of answers to participants, so that levels of knowledge and understanding of issues can be tapped into. Also, they are useful for exploring new areas of research such as this (Bryman, 2012, p.247). When creating the questions, reference was made to the literature review so that the questions asked focused on the themes it had identified, as well as keeping the research objectives in mind at all times.

The first question asked was a closed, quantitative one, designed to discover what types of Web 2.0 resources were being archived by each participant. The set list of answers was created by giving choices of general Web 2.0 records that had most often been referred to in the literature or that it had suggested were most regularly created and used by online users. These were blogs, wikis, social network profiles, Twitter feeds ('tweets') and YouTube videos. An 'other' option was given, as there are a huge variety of Web 2.0 records that could also be archived, even if they weren't the common types identified. A definition of Web 2.0

was given at the beginning of the interview form, to ensure the respondent understood the types of resources they were being asked about.

Although a pilot study was not conducted, the interview form was emailed to three archive professionals and two professionals not connected to the archives and records management field. This was to establish whether the form itself was simple and easy to fill out, and whether the questions were clear. Their feedback was taken on board and two questions were reworded, before sending the finalised interview schedule to the participants.

3.2.4 Data analysis

A thematic analysis was the approach adopted for analysing the data collected. For both the policies and the interviews, themes and sub-themes were searched for in the text or answers. This was done by looking for repetitions, both within an individual policy/interview and across all the policies/interviews; missing data, i.e. what was not highlighted as important by the participant; similarities and differences; and the themes identified in the literature review. These themes were then entered into a framework for comparison across the three countries (see Appendix).

3.3 Limitations and lessons learned

3.3.1 Literature review

The main limitation with the literature review was the lack of suitable material. The majority of the reading available about Web 2.0 and archives was concerned with how institutions used Web 2.0 tools to provide access to their holdings, add information to their records, or for outreach. There was very little material about the preservation of Web 2.0 resources as archival records. Although this was limiting in the amount of information available, it did highlight the gap in research of this topic area, and therefore solidify the need for this study.

Another limitation was that only material written in English could be used. Two potentially relevant articles were identified, however they were in other languages and access to the foreign publications could not be found. This limitation reduced the amount of potential sources, and given the global slant of this study could be seen as especially relevant.

3.3.2 Selection of participants

A potential limitation in the selection of participants involved the choosing of which countries to focus on. They are all Western countries, whose main language is English, generally have a high rate of regular internet usage among their citizens, businesses and government, have a strong archival tradition, and regularly contribute archival journals that will be read around the world. This could suggest that the findings therefore cannot be generalised to other countries whose societies don't fit these criteria. However, in regards to the internet usage, for any country where this is not prolific, the investigation and findings of how Web 2.0 records are preserved will not hold any relevance for them.

A major concern of this study is the number of participants. For the UK and Australia, there were only two institutions that accepted and preserved web-based records, therefore they could only have two participants each. In both countries, the institutions archived the web on behalf of organisations in their nation, so the need for other archives to preserve web-based material was non-existent and all the online records from businesses, archives, universities, government, communities etc. were centralised into these two national web archives. This made the number of participants interviewed very small, however as they were dealing with web-based records from across their country, their answers could be seen as highly representative of their nation.

3.3.3 Interviews

The risk of sending all the questions at once was that respondents could have read all the questions and replied only to those they felt interested in or that they could contribute to (Bryman, 2012, p.668). Asking one question at a time, or sending the questions in small batches was considered, however it was decided that this was more likely to allow the participant to drop-out of the exchange as it would cause more inconvenience to them and make the interview process seem longer. As was shown with the responses received, all questions were answered in every interview.

One limitation of conducting interviews online is that answers tend to more considered and grammatically correct, and may lose some of their spontaneity. Also, although Bryman (2012) says we are in the middle of huge growth in the amount of research conducted on the internet, "e-research is very much a work in progress" (pp.679-681). This

means that an understanding of reliability and validity of this method and ideas of best practice are still being discussed.

3.3.4 Data analysis

A limitation of thematic analysis is that the approach is at risk of being affected by the researcher's own world view or interpretation of what the participant believes to be important. However, as the researcher was aware of this, they were able to minimise this bias by concentrating on the raw data and not their own expectations or beliefs.

3.4 Methods summary

For the purposes of this study, a qualitative methodological approach was undertaken. This involved the analysis of policy documents and the use of semi-structured interviews with open questions, conducted by email. Participants were identified through purposive sampling and although this might suggest that the findings aren't able to be generalised, the national coverage they have makes them not only suitable for this study, but also means their responses provide an enlightening understanding of the topic and a real contribution to completing the aim of this research. Throughout this study, every effort was made to consider all methodological options and limitations, to identify the best approaches for the topic and that would meet the objectives given in Chapter 1, as well as to reduce the possibilities of bias and unreliable or invalid results.

Chapter 4: Results

This chapter presents the findings from analysing the collection policies and answers given in the semi-structured interviews. For the policies, the results have been separated by the country of the institution. For the interviews, the results have been separated by theme. To make the findings simple to discuss and easy to distinguish by country while keeping anonymity, the participants have been labelled UK1 and UK2, USA1 and USA2, and AUS1 and AUS2.

4.1 Collection policies

4.1.1 UK policies

The two UK policies analysed were both the longest and the shortest ones in the study – the first being one page long and the second being 10 pages long. They were 1½ and 2½ years old. Both referred to the institution’s general collection aim and mission statement and indicated the importance of archiving the web. Both referred to relevant legislation, namely intellectual property rights (IPR), copyright, data privacy and the Public Records Act 1958. The criteria and the process for selection of suitable web-based records were outlined in both policies. As were references to collaboration such as “[this archive] operates within a broader collaborative framework” and “[this archive] seeks collaboration with other...national and international stakeholders in web archiving”. Both policies suggested a societal influence on their selection of websites. The first provided a link that allows people to suggest websites for archiving, and the second stated that “policies may be reviewed and revised in the light of feedback received from the users of the records”.

As the second policy was far more extensive than the first, it covered a variety of themes that the first did not. These included types of records that aren’t included in their collections; technological issues with collecting and preserving web-based records; the responsibilities of others; a recognition of other web archives; reasons for having collection policies; and a section specifically about Web 2.0 records. This last theme highlighted that the institution believed they “might be worthy of preservation” and that they are currently doing research into how to archive these types of records. The first policy did not specifically mention Web 2.0 records or separate them out in its policy.

4.1.2 USA policies

The two USA policies were a similar length, being about 3 pages long. The first was 4 years old and the second did not have a version date, but seemed to have last been edited in 2012. It was indicated to the researcher that the first policy was “a bit out of date” but that the institution hoped to revise and update it in the coming year. In regards to the second policy, it was intimated that this wasn’t a detailed collection policy, as they had three thematic web collections and adhered to those individual subject profiles for guidance in selection.

Both policies gave reasons for and the importance of archiving the web. They also both mentioned collaboration, the first stating that they “acknowledged the importance of international collaboration for preserving Internet content” and that they “develop partnerships and cooperative relationships”; the second stating that their commitment to collection development and preservation best practice “is informed by collaboration with other research libraries and the broader web archiving community”. The second also indicated their current project that is developing and testing models of collaboration with other research libraries, scholars, web content producers and web archiving programs.

Some themes were covered by one policy but not the other. The first provided a history of web archiving at the institution, stated their general aim and mission, described the current collections of web content that they hold, and raised a concern with the cost of storage and purchasing new tool sets to capture web-based records. It indicated this last point as a reason why “selection must be considered carefully”. The second policy described how web content is acquired and indicated an awareness of legislation, particularly IPR, when discussing permissions to archive an organisation or individual’s websites.

4.1.3 Australia policies

The first Australian document was just over 1½ pages long, 5 years old and was a collection plan rather than a policy. Its opening sentence also warned that it was “currently only in draft format”. The plan covered themes including how websites were captured, selection criteria, a variety of different selection and acquisition techniques, the technical restrictions of capturing and storing web-based records, and the value in archiving the Web. This last point included the need to capture the Web “from both a records management and risk management perspective”.

The second Australian archive didn't have one collection policy. Rather, it had separate selection guidelines for each of the partners whose collections made up its collaborative web archive. Each set of guidelines covered similar themes, including a history of their web archiving activities, references to relevant legislation, terminology, selection criteria, the concept of value, current web-based holdings, and what was included and excluded from their collections. Links to the 10 separate guidelines were prefaced with a general introduction, which provided the scope of the collaborative archive, an awareness of other web archives around the world, basic criteria for selection and its general mission.

4.1.4 General observations

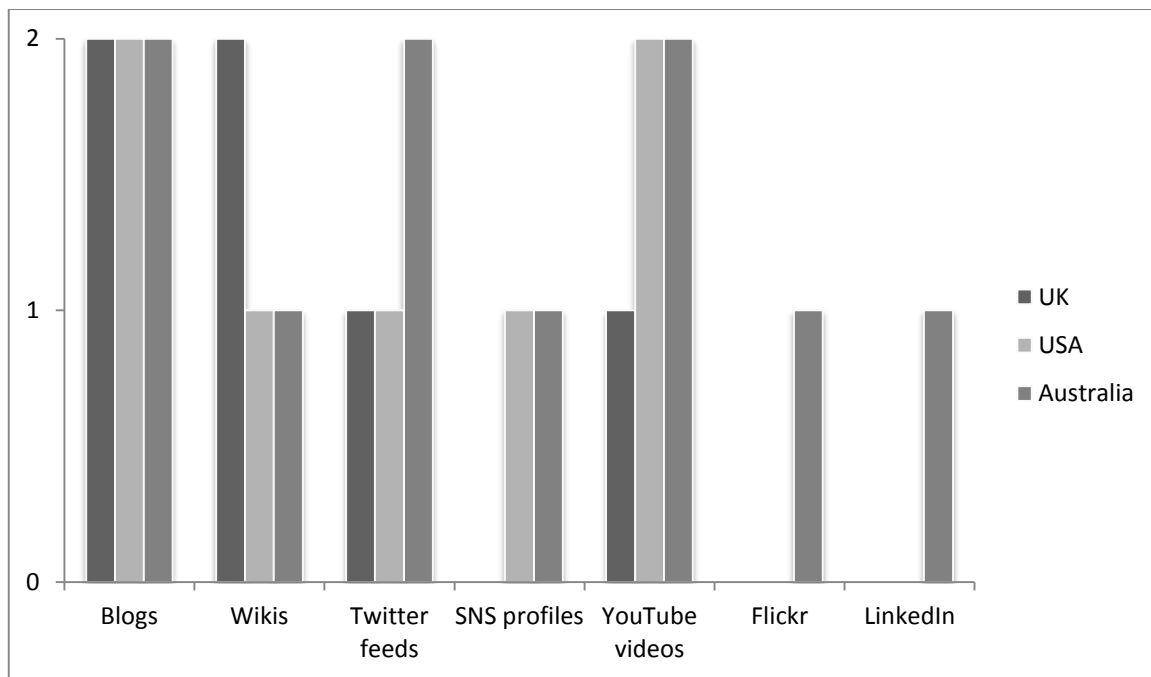
The archives all had separate policies or guidelines for dealing with web-based records, but Web 2.0 records were not distinguished from these in their own policy, or in real detail within the web collection policies. All the policies provided links to or referenced the institution's other collection policies, so that the web-based records would fulfil the same mission and compliment their current holdings and collection activities. All the policies also mentioned collaboration, either in the capturing of records, the setting of standards, or for the future of successful web archiving. In addition to this, the majority of the policies stated that they were a member of the International Internet Preservation Consortium (IIPC).

4.2 Interviews

4.2.1 Types of Web 2.0 records

Question 1 was designed to discover which types of Web 2.0 records are collected and preserved by the participating archives. The chart below shows the results. Flickr and LinkedIn were answers given in the 'other' option.

As can be seen, blogs are collected by all the archives; and YouTube videos, Wikis and Twitter feeds are preserved by the majority of participants. The UK does not collect social network sites (SNS) profiles at all, but prefers to preserve blogs and wikis. Australia collects the widest range of Web 2.0 records.



In addition to the above results, one USA participant said that for one of their themed collections, they would scope their crawls to include whatever was linked from the main sites they archived, including content on social media sites such as Pinterest, Vimeo and MySpace.

4.2.2 How Web 2.0 records are selected

All participants indicated that Web 2.0 records were selected in line with their general mission, as per their remit, or as part of their wider collections. Web 2.0 records were not generally selected on a different basis than other web-based records, or other types of records in general. However, occasionally specific projects meant that a collection could purely be comprised of Web 2.0 records - “sometimes things like blogs are the entire focus of collection activity” (USA1). UK2 was the only participant with an automated process for selection, stating that

because we have a defined remit, we are somewhat different to other institutions. We do not “select” resources as many would, but rather capture on a large, largel[y]-automated scale, and perform quality assurance after capture.

(UK2)

AUS1 was unique in saying that they archived Web 2.0 records from a risk-management perspective rather than an archival one, protecting their organisation’s reputation “by keeping and maintaining a record of what was on the web”.

4.2.3 The use of policies

All participants agreed that collection policies were useful in practice, but USA1 and UK2 noted the importance of keeping these up-to-date “to reflect reality and to ensure that important information is not overlooked” (UK2). Both Australian participants suggested that selection guidelines were also useful in defining what wasn’t included or was excluded from their collections. In addition, AUS2 stated that although they use their selection policy, “it doesn’t rule our daily work.”

4.2.4 Value

All participants regarded Web 2.0 records and other web-based records as having archival value, usually in the same way as physical and digital records. In addition, most participants extended their value as evidence of life online, methods of communication, individuals’ beliefs and identifying social trends. However, the Australian responses suggested that the value of Web 2.0 records is yet to be realised, stating for example that “the value of these records is yet to be seen in a proper context” (AUS1). USA1 put conditions on the value of these types of records, suggesting they should be studied as a body of material rather than individually, and that other sources may be more likely to have useful content “but it depends on what one is looking for and where it might best be found”.

4.2.5 The influence of society

When asked about society’s impact on appraisal decisions, half of the participants specifically said that society didn’t have much direct influence. USA2 said their institution had public nomination forms for users and creators to nominate sites for archiving, but that these had not been heavily used and sites suggested were usually already in their collection. However, half of the participants said that society may have an indirect influence, either by capturing records that “reflect what society is and does” (UK1) or by motivating institutions to improve their ability to capture new modes of communication as they are created and used. AUS1 sees a future influence from society, referring to the increase in open access social movements that “are placing different values and emphasis on archiving records and information than professional archives and records managers”. They believe that

There may come a time when there is more community led/non-professional appraisal decisions...particularly with regard to web records. We may find we have to give a greater weight to societal reasons when determining retention periods than we had to previously when appraising records. (AUS1)

4.2.6 Technological issues

Technology was an issue brought up by all participants in their interviews, particularly in relation to Web 2.0 records. The difficulties suggested included formats, size, complex coding, and the frequency that Web 2.0 software and tools are updated. “YouTube, for example, is forever altering their url structure and robots.txt settings, making it a lot of work to achieve success capturing their content.” (USA2). Twitter, Facebook and YouTube were specifically named by UK1 and USA2 as being particularly problematic.

4.2.7 Access

Two participants drew attention to the difficulties of providing access to Web 2.0 records. USA2 stated that the size and complexity of collections like the Twitter archive made providing useful access very challenging. UK2 highlighted the problem of “how to present or provide access to the material in a meaningful and user-friendly manner”. They said that, “as with all archival resources, providing meaningful access is a challenge, as is encouraging potential user groups to exploit the records”.

4.2.8 Amount of material

Three participants were concerned with the amount of web-based material that is created and needs to be captured, managed and preserved. USA1 stated that “at the scale we are doing archiving (often thousands of sites being crawled at any given time), and the rapid nature that the tools used to generate Web 2.0 content, it can be a challenge to archive and preserve such content, and to keep a handle on the amount of content being collecting[sic] and how we might store it once archived”. USA2 drew attention to the risk of duplication, saying that content in Web 2.0 records was often replicated on an organisation’s website, and that there was a “danger of capturing unnecessary extra content from third-party sites.” They pointed to the use of collection policies as being helpful in keeping the size of collections manageable. AUS1 said that the sheer number of records makes sentencing them for destruction an impossible task, and that trying to find all the Web 2.0 records within its remit would be very time consuming “and we’d probably never reach the end of finding them all”. However, they dealt with the amount of material by not appraising it – “it is much simpler and easier at this point in time as storage is so economical and retrieval so easy, to just keep everything until the technology develops to the extent where web records...can be sentenced at the point of

creation or at the conclusion of a business transaction and their disposal be automated when the time comes.”

4.2.9 Legal issues

One USA participant was the only respondent to directly mention legislation or legal issues as a restriction and a consideration when archiving Web 2.0 records. This was in relation to rights issues (IPR) and the need to seek permissions where necessary. However, both Australian participants seemed to have an awareness of legal issues, as they discussed the difficulty of archiving password protected sites and contacting creators to obtain permission to archive and provide access to their web-based material.

4.2.10 Future researchers

All participants saw Web 2.0 records as being useful to the types of researchers that currently use other digital and physical archival records, such as historians. Several respondents also suggested other types of researchers that may particularly engage with Web 2.0 records. These were digital humanities specialists, journalists, private investigators, employers seeking information on prospective and current employees, possible litigants, and those interested in trending analysis. USA2 said that virtually all types of researchers would find Web 2.0 records useful, citing for example Twitter, which “is already being mined by epidemiologists tracking flu outbreaks”. AUS2 believed that “social historians are likely to find the content of most interest”. They warn however that “we do not know what researchers will find interesting or require” (AUS2).

4.2.11 Collaboration

The majority of participants made reference to collaboration and interoperability when archiving web-based records. This was in relation to three areas. The first was in the selection of suitable records for preservation, such as working with other departments within their institution or external partner organisations. The second was working with other archives to provide the best web archiving service and tools to capture and preserve web-based material, ie. Archive-It provided by the Internet Archive. The third area was in relation to the possibility of international standards for archiving Web 2.0 records. In regards to this last question, AUS2 believed that “standards are important in webarchiving[sic] to ensure in the integrity of the material collected and to allow some interoperability between different collecting agencies.” However, USA1 said that standards for archiving Web 2.0 records

“would likely be helpful, but not by themselves, disconnected from other web archiving” and UK1 stated that “standards for capture, preservation, description and access are something to work towards; but those standards are only part of wider questions about interoperability of public archives of web resources”.

4.2.12 Issues with creators

Participants from the USA and Australia all provided a theme not encountered in the literature review – the problems with creators of Web 2.0 tools, software and records. Three respondents stated that the frequency that creators updated or changed their technologies caused problems for crawlers and harvesting sites. “The people creating the publishing platforms aren’t thinking about preservation” (USA1). AUS2 said that material being unattributed was a problem, particularly as this meant archives may be unable to contact the author to gain permission to preserve the record. USA2 sums up the issue by saying “to date none of these sites have adopted practices to make their content easy to archive successfully”. They believe the challenge is to get the companies who provide these sites to adhere to standards. AUS1 provided some suggestions for how to deal with the issues with creators. They believe that “if organisations were able to control the creation of Web 2.0 sites or at least maintain a central list of such sites, archiving them would be a lot easier and you could be confident that you were archiving them all”. In regards to creating international standards for the collection of Web 2.0 records, AUS1 stated that “it might be more beneficial to ensure or encourage the providers and creators of the Web 2.0 software...to adhere to current web and accessibility standards to make the overall preservation of Web 2.0 records easier.”

4.3 Results summary

The results of the policy analysis and answers given in the interviews show that there are not any significant differences in the approaches and attitudes of different countries when archiving Web 2.0 records. Participants agreed on the lack of direct influence of society on appraisal decisions, the potential of Web 2.0 records for new types of researchers, and the need to engage with creators of Web 2.0 software and records to enable these resources to be archived successfully. All participants expressed concerns with various technical and practical issues, but didn’t see these as complete barriers to archiving web-based records and were attempting to find processes that could best deal with these problems. Legislation and

legal issues were referenced in the collection policies of the participants' institutions, but not by the respondents themselves. Finally, the most striking result was the proliferation of collaboration as a major theme and consideration for all participants, their institutions and in their activities.

Chapter 5: Discussion

In conducting a literature review and interviews with archives based in different countries, as well as analysing their collection policies, similar themes and areas of significance have been established. It is important to further analyse and relate the results from the previous chapter to the reading discussed in Chapter 2, in order to learn more about these themes, their significance and how this study fits in with the rest of the literature. In turn, this should provide an indication of how we can advance the field of Web 2.0 and archives, what we need to know in order to move forward, and attempt to meet the aims and objectives that were given at the start of this study.

5.1 Discussion of the results

When looking at the results collected, it seems clear that there are no significant differences in the approaches and attitudes of the UK, USA and Australia when capturing and preserving Web 2.0 records. They generally collect the same types of Web 2.0 material; they understand the importance of having selection policies; they see the value in archiving web-based records and are thinking about who may find them useful in the future; they don't believe society has a direct influence on their appraisal practices; they are concerned with similar technological and practical issues of capturing and dealing with web-based material; and they are all involved in and advocate collaboration and cooperative approaches to archiving society's online presence. The reading suggested that different countries based their archival theory and practices on different traditions, such as Schellenberg's ideas of value in USA, Jenkinsonian teachings in UK, and an amalgamation of these traditions in Australia (Reed, 2003). This is due to countries having different governmental, military, cultural and social histories which naturally affect the records that survived and the beliefs a nation would have about preserving its social memory. Practices have also been affected simply by the types and amount of records produced by a country, as was shown in the literature by Kristinsdottir (2003) when discussing the ability of Iceland to ignore the idea of appraisal and simply keep everything until the 1985 National Archives Act was passed. However, as the internet has enabled the 'flat' world we now operate in (Vossen & Hagemann, 2007, p.7), the production of records and the tools used to create them have also become universal. The results of this study show that archives are selecting web-based material to complement their

other collections, and using their general mission or remit to identify suitable records to capture. This suggests that although the records being selected may be based on different traditions, the general approach and attitude to archiving the web are not affected by which country is doing the collecting. This in turn shows that the possibility of finding a global approach to archiving Web 2.0 records looks promising, as the traditions of different countries should not be a barrier.

Participants' attitudes to technological and practical issues of archiving the web were also encouraging in this study. The problems and restrictions that the respondents highlighted were the same as the issues discussed in the literature, such as the volume of data (Harvey, 2010 and Crook, 2009), the problems with multimedia content (Crook, 2009), the ability to potentially store everything (Bailey, 2008), which parts or elements to preserve (Upward, McKemmish & Reed, 2011 and Theimer, 2010), and the rapid development of Web 2.0 services (If you build it, will they come?, 2010). But although these challenges were a concern discussed in all the interviews, the results showed that archives were attempting to deal with these issues and continue to find better practices and solutions to overcome these types of problems. There was also an awareness that inevitably they would not be able to capture everything, but that they were still able to archive a significant and representative portion of relevant web-based material for their collections. Technological and practical issues were therefore a challenge, but not a barrier to preserving Web 2.0 records and the wider web. It is important that literature is written about these issues in order to find best practice solutions and ensure practitioners and creators are aware of potential problems. However, their obstacle to archiving the web should be kept in proportion.

The reading indicated that archives are aware of their role in preserving social memory. Writers such as Eastwood (2002) and Pymm (2010) suggest that this responsibility in turn causes society to have a direct influence on appraisal, by defining what they expect to be found in archives and expressing their desire for preserved records to focus on events that have particular relevance for them. Upward, McKemmish and Reed (2011) also discussed the attitude of the indigenous population in Australia to archives. The need to build relationships with these communities in order to preserve their past would suggest that, in that country in particular, the influence of society would be especially apparent. However, when asked about society's influence on appraisal decisions, there was a general consensus among all three countries that there was no direct influence to speak of. Even where an institution had public nomination forms for users and creators to nominate sites for archiving, or to provide

feedback that could change their selection policy, these had not been heavily used and any sites suggested were usually already being preserved. Archives were trying to allow society to influence their capture and appraisal activities, but the public don't seem to have a huge desire to wield that kind of power. This disagrees with Newton (2001, p.90) who stated there was an ongoing plea by users for some sort of involvement in decision-making in records appraisal. But, as one participant put it, "this is not an area where society seems to have concerns, much less any expression of those concerns that affect what we are trying to do" (USA1). The only real influence society has is indirectly, as with all types of archival holdings. They do this simply by creating records, as these are the only ones an archive has the choice of accepting or not – "our collections reflect what society is and does" (UK1). An interesting point of consideration was provided by one participant who looked ahead to a time when there might be more community led or non-professional appraisal decisions. They seemed to believe this was more likely with web-based records. This is perhaps an area that needs further investigation, as the idea of open access social movements was not one that appeared in the reading when researching Web 2.0 and archives. However, this does tie in with articles by Pymm (2010) and Yakel (2006) for example, who suggest that the traditional concept of archives is changing and the role of the archivist is being reconceptualised.

An interesting point that was uncovered by this study which didn't appear in the literature, was the problems web archives have with the creators of Web 2.0 records and software. Up till now, this issue seems to have been largely overlooked by writers on this topic. However, practitioners are certainly aware of the need to engage with creators of Web 2.0 tools, services, software and records to enable these resources to be archived successfully. There were even suggestions made that the focus should not be on creating standards for archivists and records professionals, but that best practice standards should be drawn up for creators to adhere to; as well as encourage them to control the creation of Web 2.0 material and consider how they can help preserve these resources. Perhaps records professionals have enough awareness, knowledge and experience to archive the web in the best way with the current approaches at their disposal, while trying to improve on these practices. Therefore we should instead be concentrating on working with creators more directly, increasing awareness, and engaging with software and records at the point of their creation, as was advocated by Harvey (2010, p.8). This could ensure that these resources are able to be archived, or at least aim to be.

One of the most striking results of this study was the awareness of, but lack of concern with legal issues when archiving the web. The selection policies made reference to legislation and legal obligations that were relevant to the material archives were capturing and storing. However, this seems to be done as a formality or an obligation in itself, as the results of the interviews demonstrated that the practitioners, although aware of these legal issues, were not particularly concerned with them. This would suggest that although legislation is needed to regulate and protect creators, their records and archiving activities, this is not proving to be a barrier or major difficulty when archiving web-based material. This is contrary to what the literature may have suggested. Various authors highlighted copyright, privacy concerns and intellectual property rights (IPR) as being particularly hindering to the preservation of web-based records. The findings of this study however, seem to agree with Dionne (2012, p.2), who believes that in practice there are actually very few legal decisions that impede the progress of web archiving. Legislation and legal issues do not seem to play a large role in the day-to-day archiving of Web 2.0 records. Perhaps this could suggest that adhering to legal stipulations and ethical practices have become natural to records professionals, so that complying with legislation is not an issue but an instinctive way of working. It could also suggest that Cahill (2009, p.22) was right and our notion of privacy is changing, so that the bare minimum of legal protection for creators is sufficient. However, recent stories in the national UK press about the use of data on Facebook and Twitter for example, show that the general public are unhappy with the records they are creating being used for commercial gain or in ways they are unaware of. Although they are willing to 'publish' more private information more publicly than ever before, they may not be aware of the potential for that information to be archived and available for generations to come. This could prove to become a major issue in the distant future, when web archives are more utilised for studying the past. The question of 'ownership' is at the route of these concerns, therefore perhaps clearer legislation should be being created to ensure creators are aware of where they stand with this issue and how their records may be stored and used. There is a danger however, that such legislation would be too restrictive and would not fit in with the free and creative sentiment of the internet, and particularly of Web 2.0 services. Consequently, perhaps archivists should actually be more concerned with legislation in order to prepare for future issues, even if it doesn't currently greatly affect or restrict their web archiving activities.

The most important point to be taken away from this study is the significance and already largely embraced concept of collaboration and cooperative practices. The results show that in the selection of suitable records for preservation and finding the best web archiving service and tools, collaboration is essential. Not only are archives aware of this, but practices such as engaging with other departments and organisations, using other archives' services and carrying out collaborative projects show that this cooperative attitude is already established in regards to archiving the web. These findings agree with the literature, which includes many writers who advocate coordination and collaboration. Harvey (2010, p.97) clearly expresses the benefits of collaboration as being access to a wider range of expertise; sharing the costs of developing software and systems; access to tools and systems of other organisations; sharing of learning opportunities; encouragement of influential stakeholders to take digital curation seriously; increased ability to influence data producers and system developers; joint research and development of standards and practices; and an enhanced ability to attract resources and other support for well-coordinated curation programs at regional, national or sectoral levels. All of these advantages seem to be being utilised by the participants, or they are acutely aware of them at least.

From the reading, the policies studied and the interview answers provided, it is clear that although the Internet Archive is based in the USA, it is involved with most national web archiving initiatives around the world. Their expertise and experience are obviously proving useful to countries who want to capture their online presence. As well as their widely used tool Archive-It, this suggests that working with the Internet Archive could be central to creating a global approach to archiving Web 2.0 records, and that they may be able to coordinate these efforts. This is also true of the International Internet Preservation Consortium (IIPC). The surveyed archives not only state that they are members of the IIPC in their selection policies, but it also has members from a variety of countries around the world. Being part of this group is perhaps the reason why some archives have shown a good awareness of other web archiving initiatives across the globe in their collection policies. With the traditions of different countries having been shown not to be a barrier, and their approach and attitudes to archiving Web 2.0 records proven to be very similar, the IIPC seem best placed to coordinate these and other countries in creating a global approach to preserving web-based materials.

In regards to what this global approach should cover, it seems the themes identified in the literature and the findings of this study should be the main focus. International standards

or guidelines could be drawn up to ensure a level of quality, consistency and long-term preservation and access to these records, which would need to include advice surrounding the technical and practical issues of archiving web-based material. These standards would need to enable interoperability of archives from around the world, but also be “broad enough to allow institutions to address their specific needs” (UK2). Initiatives to engage with creators of Web 2.0 software could be launched to encourage the consideration of the effect on preservation that system changes and updates could have; as well as working with records professionals to make web-based materials suitable for archiving. Relevant international legislation could be created that all countries who wish to archive Web 2.0 records could adhere to or encourage to be made official by their government. Above all, a global approach should be shaped through collaboration and cooperation of all nations, centred around the help and advice of the Internet Archive and the forum provided by the IIPC.

5.2 Discussion summary

In conclusion, it seems that different countries approach archiving Web 2.0 records in the same way, albeit that the actual records selected are based on their own traditions and established missions or goals. Concerns such as technology and practical issues are shared, but are not proving to be barriers to preserving web-based material. The influence of society and legislation, contrary to the literature, are not seen to have a huge effect on the day-to-day working of web archives. Collaboration is not only important but is also seen as essential to ensure that potential records are correctly identified, captured and successfully preserved, as well as being aware of and interacting with other web archiving initiatives around the world. Archives need to coordinate their efforts to create international standards and a global approach to archiving Web 2.0 records, and this would seemingly be embraced as collaboration is already widely utilised by institutions.

Chapter 6: Conclusion

6.1 Review of this study

The aim of this study was to explore how different countries are approaching archiving Web 2.0 records and ascertain if there is a global approach which can be identified. The objectives, as stated in Chapter 1, were:

- To discover which kinds of Web 2.0 records are being preserved by archives.
- To establish if there are similarities or differences between countries' approaches and why.
- To observe how cultural, regulatory and technical frameworks impact the curation of Web 2.0 records.
- To ascertain if a global approach to preserving Web 2.0 records would be appropriate or practical.
- To make recommendations for appropriate approaches for archives.
- To identify areas for future research.

From the literature review, the main themes in regards to archiving Web 2.0 records were found to be the concept of value; the influences of society and technology on appraisal decisions; the use of policies; the creators, users and future researchers of Web 2.0 resources; the practical and legal issues relating to archiving the web; the different archive traditions and approaches of countries; and the idea of collaboration in the archive field.

To meet the objectives of this study and investigate the themes discovered in the literature review, a qualitative methodological approach was undertaken to allow scope for discussion, more descriptive answers and unusual responses. This was considered more suitable than a quantitative approach, as the area of research was relatively new. The method involved the analysis of policy documents and the use of semi-structured interviews with open questions, conducted by email. This allowed respondents to consider their answers and provide as much detail as they wished, without being steered towards any particular (expected) answer. Participants were identified through purposive sampling, which was suitable for this study to ensure the data collected would be relevant to the research topic. This data was then analysed using a thematic approach. Overall, the methods used in this study were relevant to the research topic and the most useful in investigating a wide range of

issues surrounding the archiving of Web 2.0 records. They allowed for discursive and unusual responses to be given, and provided enlightening results. The findings can be seen as reliable and valid, as participants were not lead or directed to giving any particular answers and the results show a relation with the literature already written about this topic. There is an issue of generalizability, as the countries chosen to survey were all Western nations, whose main language is English, generally have a high rate of regular internet usage among their citizens, businesses and government, and have a strong archival tradition. This could suggest the findings cannot be generalised to countries that don't fit this brief.

In regards to the aims and objectives given above, this study has achieved everything it set out to. It was discovered that blogs are the most popular type of Web 2.0 records to be preserved, with YouTube videos, Wikis and Twitter feeds also being common among web archive collections. The UK does not collect SNS profiles and Australia collects the widest range of Web 2.0 records. Generally, there were no significant differences between countries' approaches to archiving these materials, perhaps because the nature of these resources are not confined to location and encourage use and interaction from across the globe. It was established that cultural, regulatory and technical frameworks, although of concern, did not have a huge regular impact on the curation of Web 2.0 records. It has been ascertained that a global approach to archiving Web 2.0 records would not only be possible, but would seemingly be actively encouraged and embraced by institutions.

6.2 Lessons learned

The main lesson learnt from conducting this study is concerned with engaging with participants. The initial request for participation received a successful response rate, showing that the level of detail and information provided was suitable for the candidates it was sent to. A brief phone call to two potential participants who hadn't replied confirmed that this was due to the number of emails and enquiries they regularly received, and prompted a then quick reply to the request from one of them. This level of agreement to participate (all except one said yes) was considered successful. It was proven by the researcher that keeping in regular contact with participants was essential in order to receive answers to the interview questions. All respondents received a follow-up request for their replies after a month, which prompted almost all answers to be received over the next few days. A gentle reminder and reassurance

that their participation was valuable was all that was needed. A ‘thank you’ message was sent on receipt of the answers, which also made it easier to ask follow-up questions without respondents feeling over worked.

One other lesson learnt was the advantage of conducting a pilot study, something which this research did not do. Although a handful of people were used to check that the question form was simple to fill out and that the questions could be understood, a full pilot of the interview would have been useful. This is especially true in regards to one particular question, which was designed to be fairly open to interpretation to allow for a range of answers and the opportunity for unusual responses not previously considered. However, two participants in particular were unsure of what the question was actually asking and so couldn’t provide a proper response without a follow-up explanation or question being sent. Feedback from pilot interviews would have highlighted this issue so that the question could have been better phrased.

6.3 Future research

With the issue of generalizability being the main limitation to this study, it would be advisable for this research to be replicated for other countries dissimilar to the ones already surveyed. This would either further prove that there are no significant differences in the attitudes and approaches to archiving Web 2.0 records; or show that a country’s archival traditions and culture can have an effect when comparing Western and non-Western practices. This replication of research would also show whether there were collaborative efforts happening elsewhere in the world and whether a global approach would truly be possible.

Other areas of future research have been revealed by the literature review. The most imperative of these is the major gap in publications on Web 2.0 as a resource to be archived, rather than as a tool for access or outreach. In addition to this, the ‘problem with creators’ issue that was exposed by this study could prompt further investigation. This could take the form of a survey of creators of Web 2.0 records to see how aware they are of their rights, the potential for their material to be archived, and what uses of these records they disapprove of.

Bibliography

All citations and this bibliography have used Harvard APA as their style guide.

Bailey, S. (2008). *Managing the crowd: Rethinking records management for the Web 2.0 world*. London: Facet.

Black, E. (2007). Web 2.0 and Library 2.0: What librarians need to know. In N. Courtney (Ed.), *Library 2.0 and beyond: Innovative technologies and tomorrow's user* (pp.1-14). Connecticut: Greenwood.

Boles, F. (1987). Mix two parts interest to one part information and appraise until done: Understanding contemporary record selection processes. *American Archivist*, 50(3), 356-368.

Boles, F. & Greene, M. (1996). Et tu Schellenberg?: Thoughts on the dagger of American appraisal theory. *American Archivist*, 59(3), 298-310.

Boles, F. & Young, J. (1985). Exploring the black box: The appraisal of university administrative records. *American Archivist*, 48(2), 121-140.

Bryman, A. (2012). *Social research methods* (4th ed.). New York: Oxford University Press.

Cahill, K. (2009). *User-generated content and its impact on web-based library services*. Oxford: Chandos.

Cohen, S. (2009). People finding 2.0. *Information Today*, 26(5), 16.

Cook, T. (2005). Book reviews: No innocent deposits: Forming archives by rethinking appraisal. *The American Archivist*, 68(1), 164-168.

Crook, E. (2009). Web archiving in a Web 2.0 world. *The Electronic Library*, 27(5), 831-836.

Dionne, M. (2012). Web archiving: Selection, capture, preservation, marketing #saa12 [Web log message]. Retrieved from <http://www.cmswire.com/cms/information-management/web-archiving-selection-capture-preservation-marketing-saa12-016864.php>

Eastwood, T. (2002). Reflections on the goal of archival appraisal in democratic societies. *Archivaria*, 54, 59-71.

Evans, M. (2007). Archives of the people, by the people, for the people. *The American Archivist*, 70(2), 387-400.

- Greenfield, R. (2012). What do people do with their Twitter archives? So far not much [Web log message]. Retrieved <http://www.theatlanticwire.com/technology/2012/07/what-do-people-do-their-twitter-archives-so-far-not-much/55019/>
- Hargittai, E. (2008). Whose space?: Difference among users and non-users of social networking sites. *Journal of Computer-Mediated Communication*, 13(1), 276-279.
- Harvey, R. (2010). Automating the appraisal of digital materials. *Library Hi Tech*, 28(2), 313-322.
- Harvey, R. (2010). *Digital curation: A how-to-do-it manual*. London: Facet.
- Jenkinson, H. (1980). *Selected writings of Sir Hilary Jenkinson*. Gloucester: Alan Sutton.
- Kristinsdóttir, K. (2003). The turning point in 1985: The history of appraisal and disposal of records in Iceland. *Archival Science*, 3(2), 199-204.
- Lockwood, E. (1990). "Imponderable matters": The influence of new trends in history on appraisal at the National Archives. *American Archivist*, 53(3), 394-405.
- Maclean, I. (1995). An analysis of Jenkinson's 'Manual of Archive Administration' in the light of Australian experience. In P. Biskup, K. Dan, C. McEwen, G. O'Shea & G. Powell (Eds.), *Debates and discourses: Selected Australian writings on archival theory 1951-1990* (pp. 218-224). Canberra: Australian Society of Archivists.
- Neumayer, R. & Rawber, A. (2007). Why appraisal is not 'utterly' useless and why it's not the way to go either: A Provocative Position Paper (PPP). *Digital Preservation Europe*. Retrieved from http://www.digitalpreservationeurope.eu/publications/appraisal_final.pdf
- Newton, T. (2001). Will the tension ever end?: Some observations and suggestions from an appraisal archivist. *Archives and Manuscripts: The Journal of the Australian Society of Archivists*, 29(1), 86-97.
- Pymm, B. (2010). Archives and Web 2.0: The example of the September 11 digital archive. *Archives and Manuscripts*, 38(1), 13-26.
- Reed, B. (2003). Diverse influence: An exploration of Australian appraisal practice: Part One. *Archives and Manuscripts*, 31(1), 63-82.
- Research Information Network. (2010). *If you build it, will they come?: How researchers perceive and use Web 2.0*.

Richardson, K. & Hessey, S. (2009). Archiving the self?: Facebook as biography of social and relational memory. *Journal of Information, Communication & Ethics in Society*, 7(1), 25-38.

Samouelian, M. (2009). Embracing Web 2.0: Archives and the newest generation of web applications. *The American Archivist*, 72(1), 42-71.

Schellenberg, T. (1956). *Modern Archives: Principles and Techniques*. Chicago: University of Chicago.

Theimer, K. (2010). *Web 2.0 tools and strategies for archives and local history collections*. London: Facet.

Twigge, S. (2003). Public records, public consultation: The involvement of stakeholders in the selection of records at the Public Record Office, the United Kingdom National Archive. *Archives and Manuscripts: The Journal of the Australian Society of Archivists*, 31(1), 15-23.

Upward, F., McKemmish, S., & Reed, B. (2011). Archivists and changing social and information spaces: A continuum approach to recordkeeping and archiving in online cultures. *Archivaria*, 72, 197-237.

Vossen, G., & Hagemann, S. (2007). *Unleashing Web 2.0: From concepts to creativity*. Massachusetts: Elsevier.

Yakel, E. (2006). Inviting the user into the virtual archives. *OCLC Systems & Services: International Digital Library Perspectives*, 22(3), 159-163.

Appendix A: Participation request

Dear [name or sir/madam],

I understand how busy you and your archive are, and I'm sure you frequently receive requests from students conducting studies. However, your participation will not take very long and should require no research before providing a response. I therefore hope you will consider the following request for your help.

I am a UK postgraduate student studying MSc Econ Archive Administration at Aberystwyth University. I am currently researching the preservation of Web 2.0 records as part of my dissertation and to complete this qualification. The main aim of my study is to explore how different countries are approaching archiving Web 2.0 records and to ascertain if there is a global approach which can be identified. I wish to investigate how archives choose, collect and appraise Web 2.0 records. This research is important as Web 2.0 tools are regularly used by an ever increasing number of people around the world, and the records they create can offer new types of insight into many areas of research in the future, such as social connections, public opinion, societal change, political history, technological advances, and many more.

I am writing to you as a member of an archive that accepts or aims to collect web-based records, and because you are involved with the collection and/or appraisal of these types of documents. I would therefore like to ask if you would answer a few set questions via email about the topic outlined above? I would ideally like a reply to the questions to be returned by the end of January 2013, if it would be appropriate for your work schedule.

Please note that participation is voluntary, confidential and anonymous, as you and your institution will not be personally identified in the research. The data collected will be held on a password protected computer and destroyed once my dissertation has been submitted. If you feel someone else within your organisation would be more suited to answering the questions, please forward this message to them.

If you require any more information or have any questions, I can be contacted via this email address, or by phone on +44 (0)1276 503325. I look forward to hearing whether you are able to participate in this study.

Yours Sincerely,

Ruth Kusionowicz

Appendix B: Interview schedule

Thank you for agreeing to participate in this study. Please answer the questions below and send your replies back at your earliest convenience to rek09@aber.ac.uk. If you have any problems or queries, please contact me on this email address, or phone +44 (0)1276 503325.

Please note that participation is voluntary, confidential and anonymous, as you and your institution will not be personally identified in the research. The data collected will be held on a password protected computer and destroyed once my dissertation has been submitted.

Ruth Kusionowicz

Questions

For the purpose of this study, Web 2.0 has been defined as:

a perceived second generation of web-based services that emphasise online collaboration and sharing among users; is the network as a platform; and is interactive and encourages participation, where end users upload as well as download.

1) Does your institution collect and preserve Web 2.0 records?

Yes - what types of Web 2.0 records does your institution preserve? (select all that apply)

- Blogs
- Wikis
- Twitter feeds (tweets)
- Social network profiles
- YouTube videos
- Other (please state) [Click to enter text.](#)

No - why not? Ignore question 2.

[Click to enter text.](#)

2) How does your institution identify suitable Web 2.0 records for preservation?

[Click to enter text.](#)

3) How useful do you find your appraisal/collection policy to be in practice?

[Click to enter text.](#)

4) What would you say are the major considerations when archiving Web 2.0 records?

Click to enter text.

5) What types of restrictions are there (if any) to the preservation of Web 2.0 records?

Click to enter text.

6) In your experience, what influence have you found society to have on appraisal decisions?

Click to enter text.

7) What value do you believe Web 2.0 records have/could have as archival records?

Click to enter text.

8) Who do you think could find preserved Web 2.0 records useful in the future?

Click to enter text.

9) Do you believe international standards for the collection of Web 2.0 records would be useful? Why?

Click to enter text.

10) If you have any additional comments about archiving Web 2.0 records, archiving in different countries, your experiences in appraisal, or anything else related to this study, please feel free to add them here.

Click to enter text.

Please indicate whether you would be willing to answer any follow up questions at a later date? Yes No

Appendix C: Example of thematic framework

Theme	Country		
	UK	USA	Australia
Influence of society	<p>Q2 "...supplemented by suggestions resources from external partner organisations and from users" (UK1)</p> <p>Q6 "to the extent that our collections reflect what society is and does." (UK1)</p> <p>Q6 "None...." (UK2)</p>	<p>Q6 "we think the answer at this point is 'not much'" (USA1)</p> <p>Q6 "This is not an area where society seems to have concerns...." (USA1)</p> <p>Q6 "More indirectly, by its actions and use of new modea of communication and information exchange...." (USA1)</p> <p>Q6 "We have public nomination forms... These have not been heavily used...." (USA2)</p>	<p>Q6 "We have found society to have little influence...." (AUS1)</p> <p>Q6 "The influence of open access social movements is increasing...." (AUS1)</p> <p>Q6 "We try to capture titles of interest at the time that society is concerned about them." (AUS2)</p> <p>Q6 "We try our best to collect material that represents our society at the moment we are collecting." (AUS2)</p>
Access	<p>Q4 "how to present and provide access to the material in a meaningful and user-friendly manner." (UK2)</p> <p>Q10 "providing meaningful access is a challenge...." (UK2)</p>	<p>Q7 "...their size and complexity is very challenging to manage and provide useful access to." (USA2)</p>	