ETD2005 Conference
Quakes, Quivers and Quacks
How the first online ETD has fared over 10 years

Simon Charles Nepean Pockley
Manager, Digital Object Management System, Deakin University

ABSTRACT

Accessible since early 1995, the *Flight of Ducks* was the first fully online Ph.D. The difficulties encountered in keeping this work live have included institutional dysfunction, redundant technologies, and citation erosion. Strategies to overcome these problems have required a significant re-conceptualisation, not just of the work or who should take responsibility for maintaining it in the digital environment but of the form and function of ETD itself.

1. INTRODUCTION

The title and content of this paper follows on from, *Killing the Duck to Keep the Quack* (Pockley 2001) presented at the Fourth International Symposium on Electronic Theses and Dissertations, Pasadena.

The *Flight of Ducks* (Pockley 1995) was the first fully online doctorate in that it was conceived, researched, submitted, examined, and stored – online. It is easy to forget that in 1995, an understanding of a web enabled poetic was drawn more from the vision of such prescient thinkers as Vannevar Bush and Ted Nelson than from any practical experience. The World Wide Web was a novelty: there were very few personal digital repositories, neither semantic nor syntactic consensus for metadata, and most 'interactive' writing took little account of the implications of having two-way communications over scholarly material. The idea of the ETD was mainly understood as a paper document converted into PDF.

That the first networked ETD is still seen as too radical a departure from the cautious implementations that characterise this community of interest, may suggest that it was simply an early aberration. Nevertheless, after a decade, it continues to engage a diverse academic audience as well as provoke encounters with a wide range of non-academic participants.

By way of general description: the *Flight of Ducks* reflects its content by taking the poetics, conventions, and literary forms of exploration through landscape and applying them to a datascape. The topography of the datascape is shaped by the structural displays of a personal digital repository of texts, images, and sounds. Ideas about digital preservation, aboriginal representation, and narrative structure are explored through scholarly texts. These texts are embedded in the datascape as cross referenced hypertexts. By integrating external contributions, this ETD extends beyond the margins of a fixed document to act as an expanding container; a participatory space where ideas can be exposed to the rigours of peer review. As a live, evolving work, it has been necessary to throw out the idea of the finished work and to re-define the notion of the published work.

Note: this paper is available within the *Flight of Ducks,* embedded in the content it describes.

2. QUAKES: major shocks that have threatened the project's survival

Over 10 years there has been consistent growth in the rate of individual machine access (not hits) to the site. In the last 12 months more than 2 million individuals have visited. For most of its life, annual visitation has fluctuated around 1 million. There have been 3 significant shocks that have made the site unstable and twice threatened its continuity. Contrary to expectations, these shocks have arisen from 'political' events rather than from technical redundancy. Access statistics have only been available intermittently. Consequently, Figure 1 is an approximate schematic (no statistics available for PANDORA). The graph charts the impact on visitation figures for the periods when the site was removed from servers by its hosts.
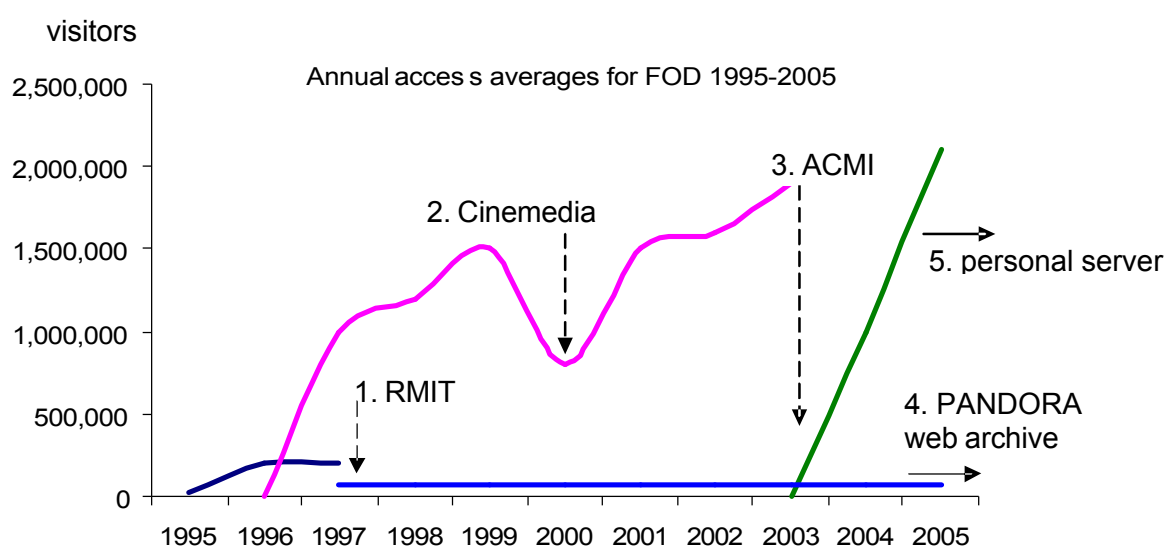

Figure 1. Impact of server changes on annual access statistics

Censorship: The first of these shocks came from RMIT University when, at the point of submission (for examination in 1997), a politicised Human Research Ethics Committee ordered that the site be removed from the University server. The Ethics Committee was concerned that the project's preservation strategy; that of global proliferation, was at odds with the values associated with material collected from Australian Aborigines in 1933. Such was the controversy that the Australian examiners withdrew. Nine months later, when the Ethics Committee conceded their mistake, the project was examined in the USA.

Concurrently, the site had been mirrored on another server through a research agreement with the newly aggregated cultural organisation known as Cinemedia. In addition, the *Flight of Ducks* had been selected as a pilot project for the National Library of Australia's PANDORA web archive of work of national significance. Although examiners at UCLA and MIT were able to access the live work from the Cinemedia server, as well as from a temporary CD-ROM of the files, the 1997 PANDORA capture provided a snapshot in time that satisfied the University's Higher Degrees Committee that a durable record of the work (at the time of examination) had been provided. However, the *Flight of Ducks* was never reinstated on the RMIT University server.

Name change: Having found a host server at Cinemedia, the second shock, in 1999, was a domain name change. The change was due to the formation of a new cultural organisation, the Australian Centre for the Moving Image (ACMI). ACMI was building a digital collection.

The *Flight of Ducks*, along with a more popular ETD, *Adventures in Cybersound* (Naughton 1995), provided a range of curatorial challenges. During the managed transfer there was a significant drop in access.

Death in custody: The third shock, in 2003, was neither managed nor expected. ACMI found itself in financial difficulty. During a restructure, the ACMI Board cut the size and scope of its digital collections and dissolved both the ACMI Collections and Web units. A significant overlap of private, scholarly, and professional activity had occurred while employed at ACMI as the Collections Manager. Both the development files and the live files were stored on the ACMI server. Most were deleted without warning. The process of reconstructing the site from email and captured files continues today.

3. QUIVERS: technical changes - problems with solutions

Redundant markup: Reconstruction prompted a process of re-conceptualisation. The *Flight of Ducks* had tracked the development, maturation, and redundancy of HTML elements; HTML had been one of its principal organising and enabling engines. It is beyond the scope of this paper to detail the manifold changes that have been required to clean up the twin contaminants of embedded presentation and deprecated markup in HTML. The process of deploying style sheets and conforming to W3C recommendations for valid and accessible markup has been, and continues to be, a regular maintenance activity. The datascape of the *Flight of Ducks* has been formed from an arrangement of more than a 1000 separate HTML files. While some of the smaller fragmentary files can take only a few minutes to reconstruct, the larger files can take several hours.

Furthermore, changes to the syntax and semantics of the Dublin Core Metadata Standard have led to significant changes to the structure and content of the detailed metadata that has been embedded in the head of every file.

Re-conceptualisation and internal links: To compound the problem, the emergence of XML as a more flexible, robust, and conceptually sound form of markup language has created an entirely different kind of challenge. While HTML (in its purest form) provided a method for resolving document structure and presentation, XML provides a meta-language for containing and manipulating ideas and meaning.

A simple example of the transformative and flexible use of XML that could be applied to any form of textual scholarship can be seen in the way the original text of the 1933 journal (from the *Flight of Ducks*) can be distinguished from subsequent versions containing additions.

```
<1933journalFJAP>
<pencilsource url="FOD0724.html" />
…The women had a few stitches of clothing of very little use unless held together.
<journaladdition>
They were carrying wooden pitchis on their heads with skill and grace…
</journaladdition>
</1933journalFJAP>
```

Figure 2. Truncated fragment of XML version of journal entry showing how original text is distinguished from material that has been added later.

Chunks of meaning can be identified and marked-up by the extent to which they can be detached. Determining what constitutes a chunk of meaning that can be contained by XML is first and foremost a function of its useful purpose or perspective. The fragmentation of

content in the *Flight of Ducks* was initially driven as much by theatrical notions of screen space as it was by the structural elements of text that could be contained and linked separately in HTML. HTML could express structural relationships and presentation but not content values. Migration into XML has provided an opportunity to take an entirely a different approach where content values can be transformed into RSS or quite a different HTML output.

```html
<p class="1933journalFJAP">
<a href="FOD0724.html" tabindex="2">
<img src="images/source.gif" width="45" height="48" alt="source icon" border="0"><!--/img--></a>
…The women had a few stitches of clothing of very little use unless held together.
<span class="journaladdition">They were carrying wooden pitchis on their heads with skill and grace… </span><!--class=" journaladdition"-->
</p><!-- class="1933journalFJAP"-->
```

Figure 3. Fragment of HTML output from XML entry showing how original text and added material can be displayed. Markup includes future proofing comments.

Replacing structural markup (HTML) with a more content centric approach to identifying textual components led to an extensive re-conceptualisation that made hundreds of file fragments redundant when their content was absorbed into larger chunks. As a result, there has had to be a massive reworking of the internal links because the narrative structure of the work, as parallel groups of flat files, was navigated and annotated by thousands of cross references.

External links: The absence of persistent URIs for citing references is an ongoing problem that leads to citation erosion or 'linkrot.'
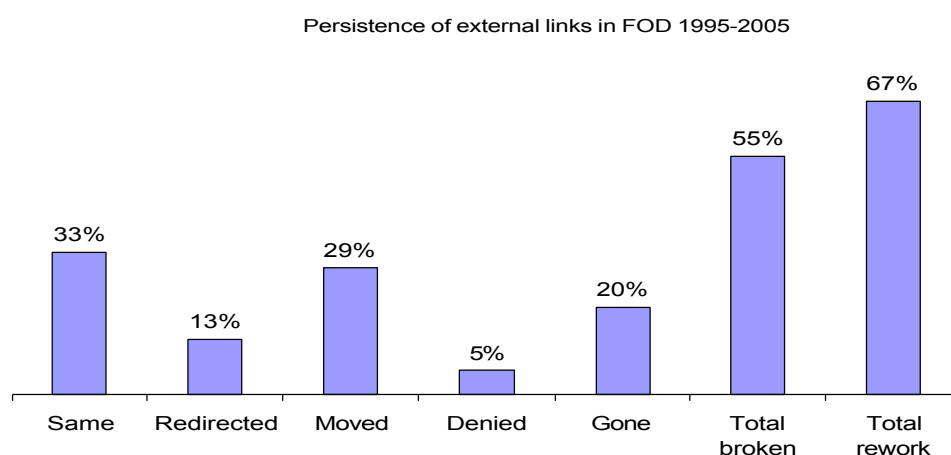
Persistence of external links in FOD 1995-2005

Figure 4. Analysis of external links in electronic thesis by % of total

Analysis of the external links used in the thesis hypertexts shows that while a third of the links were able to be resolved after 10 years, more than half were broken. Automated redirects accounted for 13% of all links and 29% had been moved to a different location without an automatic redirect. 67% of all external links required rework. 20% of the references could not be found anywhere. Password protection accounted for 5% of the links that returned an 'access denied' message. Of the 42% of links that had moved, one quarter

remained on a server with the same domain name. Three quarters were found on a different server (not shown in the graph).

4. QUACKS: vital signs of life

After a decade of maintaining the *Flight of Ducks,* one of the most interesting and consistent questions from contributors is - why? The foremost answer is that the thoughtful human interaction that occurs almost every day is not only a constant source of delight but inevitably leads to what one contributor described as, 'the flesh meeting.' A tendency towards physical meeting as a consequence of networked communication is a phenomenon at odds with notions of disconnection that are invariably associated with an intellectual life supported by the web.

There are fascinating examples, revealed through email conversations, of academics whose research interests have been incorporated into their daily lives, from the production of handmade holograms of aboriginal designs to the replication of motifs in snow. Online readers of this paper can join these conversations. Cultural sensitivity continues to provide the greatest challenges for balancing the implicit need to archive all communications with the pressing insistence of privacy. For example, significant, but very private contributions have been made about the appalling conditions in remote Aboriginal communities.

Following the unexpected contributions of the RMIT Human Research Ethics Committee, a new area of the ETD began to grow and expand. The Committee's interest in secret/sacred objects prompted research into the online trade and exposure (eBay) of culturally sensitive material. Every ceremonial artefact traded is now documented and archived into the *Flight of Ducks*. This has revealed some surprising insights.

Regardless of the institutional dysfunction, or good fortune, that has kept the *Flight of Ducks* 'in the wild,' this ETD remains outside the custodial influence of the university that approved and administered it. As a consequence, the rationale behind *proliferation,* as a demonstrably successful digital preservation strategy, has extended to values that can be characterised as open, transparent, unprotected, and responsive.

That the *Flight of Ducks* is 'live' and growing is not necessarily at odds with existing academic and administrative requirements. The imperative to freeze the doctoral thesis in time, for the purpose of examination, is easily accommodated. But why should an ETD end there? The more time passes, the more likely it is that there will be more to say and more to understand about any chosen area of research.

That an ETD should be open to annotation and responsive to change is not only dependent on its architecture, but on its real connections to human activity and thought. If we acknowledge the passion and endurance that is required to complete a higher degree of substance, there is a discontinuity in the expectation that the momentum of scholarly discourse should stop at the point of submission for examination.

Doctoral research inevitably leads to professional and academic practice. The *Flight of Ducks* demonstrates that the ETD can also serve as a personal repository for containing the revisions, expansions and contractions of post-doctoral thought – as well as locating articles and conference papers within a contextual framework.

Who should be responsible? Most university libraries are simply not resourced to undertake even the mundane activities that have been necessary to maintain the *Flight of Ducks* as a living resource.

By extending the concept of the datascape to a larger info-sphere, there are a number of converging conditions and activities that suggest a trend towards a common information space - c*ommon*, in the sense of public space, the 'common' of the 'creative commons.'

The first condition is a pervasive, ambient network. The second is that storage capacity is moving to the point where it exceeds our ability to fill it. Storage is becoming so inexpensive that it is already offered free (like email). The third is that our creation tools are both GPS and time aware. The fourth, that our creation tools allow us to upload and annotate content into the network in XML as RSS feeds or similar.

In this unbounded public space, discovery could be improved through support for browsing. The location of any digital object is defined by its position in time and/or in space. The Wikipedia provides a prototype of how this can be achieved in time. The emergence of locative media and geo-annotation and cartographic projects linking GPS co-ordinates to maps and data provide ready examples of the advantages of spatial navigation.

Just as a new model of scholarly publishing (reducing the monopoly power of academic journals) is emerging from the development of institutional repositories, so access to an electronic common could profoundly alter the role of the university as the custodian of the ETD. Within peer to peer (blog) communities we can already see the breaking down of borders that define university ownership of scholarly outputs as well as new, more fluid and vital aggregations of scholarly discourse (Farmer 2004). It is important to recognise that these communities are emerging from the ground-up through a need to configure personal information space rather than being imposed from the top-down by institutions asserting administrative control.

There is an opportunity for universities to take a more service oriented approach and to encourage the development of communities where the notion of an ETD is more of an inclusive electronic practice than an electronic object. Could we accept the ETD as an ECS - an electronic communicating space?

5. CONCLUSION

Both academic and cultural institutions can present difficult political environments that have the potential to severely compromise the durability of ETDs. There is clearly a need for persistent URIs if citation erosion is to be controlled. While format obsolescence can be managed, it is likely to be time consuming and costly for institutions. If continuity of access is to be maintained through migration into more robust formats such as XML, re-conceptualisation will be required. Consequently, responsibility for long-term access shifts towards the ETD creator. There is a need for a personal archiving infrastructure where the ETD is not simply a digital object but a self archiving space capable of accommodating ongoing scholarship.

6. REFERENCES

Farmer, J. (2004). Communication dynamics
http://www.ascilite.org.au/conferences/perth04/procs/farmer.html
Naughton, R. (1995) *Adventures in Cybersound* [persistent URL]
http://nla.gov.au/nla.arc-13071
PANDORA – National Library of Australia [persistent URL] http://nla.gov.au/nla.arc-10245
Pockley, S. (1995) *The Flight of Ducks* http://www.duckdigital.net/FOD/
Pockley, S. (2001) *Killing the Duck to Keep the Quack*
http://www.duckdigital.net/FOD/FOD0055.html