

# Transforming Lost Bits to Cultural Assets: Using Data Forensics in Digital Humanities

Leo Konstantelos<sup>1</sup>, Anna Shadbolt<sup>1</sup>

<sup>1</sup>The University Of Melbourne, Victoria, Australia, {leo.konstantelos, annams}@unimelb.edu.au

## INTRODUCTION

The aim of digital forensics within a research data curation context is to examine digital media in a manner that ensures intrinsic integrity and captures contextual information (such as file creation, versioning, software dependencies, user logs and IPR information) for use, re-use and preservation of the data underpinning research activity. The forensic analysis can further aid in securing sensitive data, auditing research processes and recovering data from damaged, failed, corrupted, or otherwise inaccessible storage media.

As the field of research that represents “a global, trans-historical, and transmedia approach to knowledge and meaning-making” [1], Digital Humanities are a natural interdisciplinary space to study theoretical and practical models for employing digital forensics in the management, curation and preservation of digital data that are the product of research activity. And yet, the interplay between digital humanities and digital forensics remains largely underexplored, particularly within the context of information management and cultural heritage institutions. Opportunities exist for potential collaboration and sharing of expertise with benefits for both sides. This is suitably argued in [2] whereby:

*“Digital Humanities, for its part, can benefit from applications of digital forensics to improve its strategies for data handling, curation, and preservation. LAMs, meanwhile, could use the new tools and techniques that are commonplace in Digital Humanities—ranging from large-scale text analysis with techniques such as topic modeling to GIS modeling and visualization—to analyze and explore collections of born-digital materials. If digital forensics is indeed the modern-day incarnation of such centuries-old techniques as diplomatics, philology, and bibliography, then such connections ought to be within reach.”* [2]

With this in mind, the University of Melbourne Library undertook the curation and preservation planning for the Australian Sound Design Project (ASDP)<sup>1</sup> using the dedicated Data Forensics Lab, which was established in 2013 as part of the University Library’s research data curation service. In this paper we briefly discuss the ASDP, then present the workflow employed to forensically curate the project’s research data. We conclude with the lessons we learnt from this process and some thoughts on future directions.

## THE AUSTRALIAN SOUND ARCHIVE PROJECT

The Australian Sound Design Project was the brain child of Dr Ros Bandt, an internationally acclaimed sound artist, composer, researcher and scholar. The project – instantiated as an online resource – ran from 2001-2007 at the Australian Centre, School of Culture and Communication at the University of Melbourne and was funded by a grant from the Australian Research Centre. The ASDP was the first national website and database dedicated to researching and publishing original works and discourse relating to sound and its design in public space. It recorded and catalogued the original works of over 200 sound designers, artists and composers, cross-referenced with contextual and provenance information in order to create both a comprehensive research tool for analytical study and an archival collection.<sup>2</sup>

From the onset of the project, Ros was concerned with issues of sustainability and long-term preservation of the ASDP. In a 2007 interview for the Resonate Journal, she argued that:

*“the future of the website is uncertain: for the last seven years I have worked on the site as an honorary fellow. Digital publishing is expensive and skilled people such as my excellent research assistants are needed for complex coding. Unlike books, the need for funding doesn’t go away with the event of the item being published. How do we sustain digital publishing?”* [3]

<sup>1</sup> <http://www.sounddesign.unimelb.edu.au/site/index1.html>

<sup>2</sup> <http://www.sounddesign.unimelb.edu.au/site/about.html>

Some initial steps were taken towards creating a more persistent record of the ASDP database and website. The entire ASDP website was burned into a CD-ROM [3] while the online resource has been in the custody of the eScholarship Research Centre (ESRC)<sup>3</sup>, stored and maintained in the University's data centre. As part of the project's "exit strategy" the computers used by the research team were also passed on to the ESRC for safe keeping. The significance of the project as a cultural and research record was a given, but what became obvious was the need for a formalised, structured process to curate the data for archiving so that we could adhere to the fundamental principles of provenance, original order and chain of custody.

## DATA FORENSICS WORKFLOW FOR THE ASDP

The workflow that we followed for forensic processing of the ASDP digital materials consisted of six steps or phases, presented in Figure 1. At the point of acquisition, we tried to ascertain that these were packaged and transported to the lab in a secure manner. We then accessioned the materials, i.e. recorded in the order of acquisition and assigned an identifier, before attaching the media to the forensic machine and creating a forensic image. We preferred to work from the forensic images for analysis, storing the original media in forensically approved conditions within the lab. A detailed workflow is shown in Figure 2.

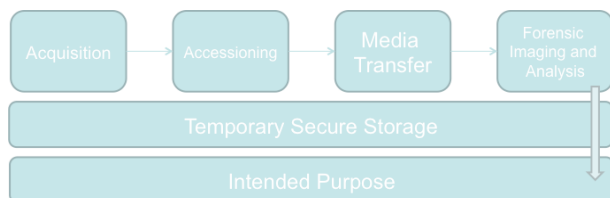


Figure 1: Detailed forensic workflow for the ASDP

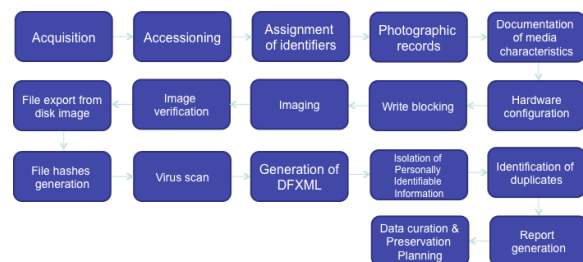


Figure 2: The generic forensic workflow for the ASDP

## LESSONS LEARNT

Working with obsolete storage media to curate the research data generated by the ASDP provided insight in a number of areas. For one, modern computer forensics systems are designed to work with contemporary technology. The necessary interfaces to connect legacy hardware are either not present or difficult to acquire and set up without relevant expertise. Moreover, forensic software is not designed to read or even identify legacy file systems.

The products of forensic imaging and analysis are – for the most part – large, complex file structures that in turn require specialised tools for access, processing and archiving. Our experience with the forensic curation of the ASDP data highlighted the lack of readiness and expertise - both within research and practice – to work with these outputs. Integration of digital forensics metadata (e.g. DFXML) is not evident in current repository environments.

Lastly, we came upon a huge arena of legal, copyright and licensing issues relating not only to the data itself (e.g. third party material), but also the operating environments (operating systems, software) of the projects computer systems. Forensic analysis opens up new paths for retrieving and analyzing data, but these opportunities come with distinct challenges. What became clear at the end of the ASDP curatorial process is that it is inexpedient for digital forensics to exist in isolation, as a monolithic process that remains ad hoc. A clear challenge is how to incorporate forensics use in existing business processes and workflows.

In his foreword to the excellent "Digital Forensics and Born-Digital Content in Cultural Heritage Collections" [4] report by Matthew Kirschenbaum et al., Charles Henry notes that "There is a palpable urgency to better understanding digital forensics as an important resource for the humanities". Closing this paper, we argue that we also need to explore new avenues, where digital forensics become an intrinsic part of research methodologies in digital humanities rather than just a "tool".

<sup>3</sup> <http://www.esrc.unimelb.edu.au/>

## REFERENCES

1. Burdick, A., Drucker, J., Lunenfeld, P., et al. (2012). *Digital humanities*. Cambridge, Mass: MIT Press.
2. Lee, C. A., Woods, K., Kirschenbaum, M., and Chassanoff, A. (2013). From Bitstreams to Heritage: Putting Digital Forensics into Practice in Collecting Institutions. BitCurator Project report. Available at, <http://www.bitcurator.net/docs/bitstreams-to-heritage.pdf>
3. Bandt, R. (2007). The Australian Sound Design Project. Resonate Magazine. Retrieved June 2 2014, from: <http://www.australianmusiccentre.com.au/article/the-australian-sound-design-project>
4. Kirschenbaum, M. G., Ovenden, R., Redwine, G., & Donahue, R. (2010). *Digital forensics and born-digital content in cultural heritage collections*. Washington, D.C: Council on Library and Information Resources.

## ABOUT THE AUTHOR(S)

**Dr. Leo Konstantelos** is Research Data Curator at the University of Melbourne. Before moving to Australia in 2013, Leo was a researcher at the University of Glasgow; and a research fellow at the University of Portsmouth, UK. He has extensive experience in digital preservation and curation, having conducted research in a number of high-profile collaborative EC-funded projects. He holds an MSc in Information Technology (2005) and a PhD in Digital Humanities, both from the University of Glasgow.

**Anna Shadbolt** has had a variety of Information Consultancy roles within the Library at the University of Melbourne over the past 10 years. With an academic background in community psychology, Anna's focus has been building bridges and partnerships (communities) between researchers and support personnel to enable effective and efficient capture, management and sustainability of research outputs. Anna has played a key role in the development of research data management policy and a variety of tools and services to support its implementation. She is currently the Manager of the Digital Scholarship team within the Library Research Program: the home of both the data curation and digital forensics services and the digitisation services.