What does data loss really cost?

A lot more than you think

|  |  |  |
| --- | --- | --- |
| **Paul Stokes** | **Tamsin Burland** | **Sarah Middleton** |
| *Jisc*  *UK*  *Paul.stokes@jisc.ac.uk*  *0000-0002-7333-4998* | *Jisc*  *UK*  *Tamsin.burland@jisc.ac.uk*  *0000-0002-5129-979X* | *Digital Preservation Coalition*  *UK*  *sarah.middleton@dpconline.org*  *0000-0002-7671-403X* |
|  |  |  |

**Abstract – Jisc and the Digital Preservation Coalition (DPC) have undertaken an anonymous surveying exercise in order to unearth the true cost of catastrophic data loss—not only in terms of the value of the data, but also the cost of the knock-on effects that may only become apparent some considerable time after the event. This poster is intended to present the findings from that survey and introduce a final report which will help organisations make a stronger case for robust and effective digital preservation practice.**

**Keywords – Sustainability, Cost, Value, Risk, Data loss.**

**Conference Topics – Resilience, Exchange.**

# Introduction

Digital Preservation is about mitigating risk. Mitigations cost money. It is hard to justify spending that money without a firm grasp of the magnitude of the sums of money involved (the value of what's at risk) and the likelihood of loss.

There is already a growing body of work relating to quantifying the likelihood of loss occurring (for example, The Digital Archiving Graphical Risk Assessment Model [DiAGRAM] from the National Archives [1]). Unfortunately, it's not so easy to value the data at risk. The knock-on effects of data loss (reputation loss for instance) are even more challenging to quantify in monetary terms. Often the sums involved only become apparent a long time *after* a disaster has happened.

We know that destructive data disasters have already happened (inevitably one might argue). Some recent headlines illustrate this:

* *Server crash takes out rich digital archive at Memorial University* [2]
* *Victoria University of Wellington accidentally nukes files on all desktop PCs* [3]
* *PASIG 2017: “Sharing my loss to protect your data” University of the Balearic Islands* [4]
* *University loses 77TB of research data due to backup error* [5]

This means that there is (potentially) data extant that would give an insight into the problem. Such data, if suitably anonymised/redacted, could form the foundation of a "Cost of failure" publication showing how devastating the impact and cost of real-world data loss can be. Regrettably (and quite understandably), those who have suffered this type of loss are rarely willing to acknowledge the fact, let alone talk about the numbers involved. So we have little insight into the true extent of their losses.

With this in mind, the Digital Preservation Coalition (DPC) and Jisc set out to provide a means whereby individuals and organisations could with confidence and anonymously provide information about the extent and cost of any significant data loss events to a partnership of two trusted organisations (namely Jisc and the DPC). The intention is also to provide a mechanism to collect lessons learned and mitigation strategies.

The aim is to collect examples, from both the UK and overseas, from a range of sectors to represent the customer / membership bases of both Jisc and DPC—for example Higher Education, Research, Public Sector, GLAM, private sector—in order to highlight why organisations should invest in digital preservation. This will help them make a sustainable business case with credible exemplar data.

To achieve this end, an anonymous survey was created using the Jisc On-Line Survey Tool [6] and published in February 2022. Both Jisc and the DPC publicised the survey in the following weeks/months. At the time of writing the survey is still open and collecting data. It is intended that it will close at the end of April. The survey results are to be used to create a publication for launch at iPres 2022.

# The Poster

This poster is intended present the methodology used, the anonymous aggregated findings and to highlight key headline findings from the survey. It is also intended to introduce the final report.

# REFERENCES

1. The Digital Archiving Graphical Risk Assessment Model (DiAGRAM) from the National Archives— <https://nationalarchives.shinyapps.io/DiAGRAM-dev/>
2. Server crash takes out rich digital archive at Memorial University—<https://www.cbc.ca/news/canada/newfoundland-labrador/mun-digital-archives-wiped-out-1.4787960>
3. Victoria University of Wellington accidentally nukes files on all desktop PCs—https://arstechnica.com/gadgets/2021/03/university-of-wellington-accidentally-deletes-files-on-all-desktop-pcs/
4. PASIG 2017: “Sharing my loss to protect your data” University of the Balearic Islands—https://blogs.bodleian.ox.ac.uk/archivesandmanuscripts/2017/09/27/pasig-2017-sharing-my-loss-to-protect-your-data-eduardo-del-valle-university-of-the-balearic-islands/
5. University loses 77TB of research data due to backup error—https://www.bleepingcomputer.com/news/security/university-loses-77tb-of-research-data-due-to-backup-error/
6. Cost of Data Loss Survey— https://jisc.onlinesurveys.ac.uk/cost-of-data-loss