The Value of Catastrophic Data Loss

Data loss as a community benefit.

|  |  |  |
| --- | --- | --- |
| **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 information can be used to help organisations make a stronger case for robust and effective digital preservation practice and to inform those trying to take steps to avoid their own data loss disaster. This workshop, intended to bring together those who have lost data with those who wish to avoid losses, explores the usefulness of such data loss events to the community.**

**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 to others (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. To 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 the first iteration will close at the end of April. The survey results are to be used to create a publication for launch at iPres 2022.

The survey results and the subsequent publication are, however, only part of the story. Knowing how and why disasters happened and the magnitude of the problem in monetary terms doesn't necessarily mean that the information is useful. It needs to be coupled with strategies to use that information, strategies to mitigate, and strategies to prevent. Above all, the individuals that make up the community need to know about these disasters and how to bring them to the attention of the appropriate people in their organisations in such a way as to ensure that they are acted upon.

And that's where the proposed workshop comes in.

# The Workshop

This half day workshop is intended to be a forum where those who have lost data can exchange information with those who would very much like to avoid having their own data loss disaster. The workshop will be run under Chatham House Rules allowing participants to share information freely.

There will be three strands of discussion:

* Past events with invited speakers to set the scene. Up to four speakers each offering a short insight into:
  + The cause of their disaster
  + The magnitude of their disaster—how much data was lost, how much it cost them.
  + Mitigations—what they wish they'd had in place, what they've put in place since.

There will also be an opportunity for ad-hoc contributions from the floor (attendees will be encouraged to come prepared to share their insights)

* An introduction to the Jisc/DPC survey
  + the preliminary results from the survey
  + how to submit information to the survey
  + using information from the survey report
* Discussion exploring the usefulness of such data loss events to the community. The direction of the dialogue will be dictated by those in attendance, but possible areas for discussion include:
  + Are they useful? If not, why not?
  + What could the community do to maximize their benefit / avoid losing the insights
  + Strategies for using disaster case studies to drive policy and business cases

The discussions from the session will be written up (in a suitably anonymised form) and fed back into the Jisc/DPC cost of data loss outputs.

# 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