Cost of Data Loss

The true cost of data loss

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 cost of protecting it) and the likelihood of loss. There has already been some work relating to the likelihood of loss and the cost of preserving data, but understanding the magnitude of any loss is significantly more difficult.

It is difficult to value data. It's even more challenging to put a figure on the cost of the knock-on effects of data loss (reputational damage or loss of productivity for instance). Often the sums involved only become apparent after a disaster has happened.

No one wants to be in the position of suffering their own data loss in order to drive a business case.

However, we know that destructive data disasters have already happened to some unfortunate organisations and individuals. Given the commercial sensitivity of any such loss, it is likely that the disasters we know about are just the tip of the iceberg - no one likes to share such information publicly.

Assuming that the organisations that have suffered data loss have survived the incident, the next logical step is for them to undertake a root cause analysis and possibly to quantify the magnitude of the loss. This means that there is almost certainly information in existence (but hidden away) that can shed some light on the true cost of data loss. The problem is, it is challenging to persuade those who have this information to share it, even for the good of the community.

And that's where this anonymous survey comes in.

The Digital Preservation Coalition and Jisc are asking organisations and individuals to share anonymously some key information that will benefit the whole of the digital preservation community (and beyond). This information will be used to undertake some quantitative and qualitative research around the cost of data loss and then produce an anonymised report for the benefit of the community. If the project is sufficiently successful, then the survey will be left open. The research will be repeated at appropriate intervals and additional reports will be published.

We're also keen to understand and share mitigation strategies, so please share those as well if you can.

Please help (and please pass on the link to this survey—
https://jisc.onlinesurveys.ac.uk/cost-of-data-loss—to any individual or organisation you feel could help as well).

You may respond to this survey more than once.

Simplified Privacy and Data Protection information

ALL INFORMATION WILL REMAIN CONFIDENTIAL AND WILL ONLY BE USED TO INFORM THE RESEARCHERS/AUTHORS OF THE PUBLICATION.

1. Information will be collected and stored within the EU/UK.
2. No Personally Identifiable Information will be collected or retained without the permission of the individual concerned.
3. Information obtained will be stored in an anonymised or encrypted format on appropriately secured infrastructure.
4. The source information will be retained for 12 months after the publication of the research report or for 24 months after submission, whichever is the shorter period unless permission is obtained from the provider of that information to retain it longer.
5. Anonymous aggregated results will be retained indefinitely.
6. Jisc will be the data controller of this information. The Digital Preservation Coalition may act as a Data processor. If that is the case, there will be an appropriate Data Protection agreement in place between Jisc and DPC.
7. **All questions are optional**, but please answer as many as you can. The more information that you supply the more informed and useful the resulting research will be.

A full data protection and privacy notice can be found at https://www.jisc.ac.uk/website/privacy-notice

Note: It is possible to participate in this survey without providing ANY personal information.

Questions and timing

It's impossible to give an indication of how long it will take to complete, but we (and the community) are very grateful for any and all effort you put into it.

Some of the responses in this survey will require thought/reflection. With this in mind, we have provided an overview of the questions for you to download to enable you to prepare your responses off-line before submitting them.

Page 2: Categorisation

About the organisation

These questions are intended to allow us to understand you and your organisation.

See SIC Hierarchy - https://onsdigital.github.io/dp-classification-tools/standard-industrial-classification/ONS_SIC_hierarchy_view.html for further information about these classifications

1. **Industry** Please provide an indicator of the industry this data loss relates to.

    [AGRICULTURE, FORESTRY AND FISHING; MINING AND QUARRYING; MANUFACTURING; ELECTRICITY, GAS, STEAM AND AIR CONDITIONING SUPPLY; WATER SUPPLY; SEWERAGE, WASTE MANAGEMENT AND REMEDIATION ACTIVITIES; CONSTRUCTION; WHOLESALE AND RETAIL TRADE; REPAIR OF MOTOR VEHICLES AND MOTORCYCLES; TRANSPORTATION AND STORAGE; ACCOMMODATION AND FOOD SERVICE ACTIVITIES; INFORMATION AND COMMUNICATION; FINANCIAL AND INSURANCE ACTIVITIES; REAL ESTATE ACTIVITIES; PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES; ADMINISTRATIVE AND SUPPORT SERVICE ACTIVITIES; PUBLIC ADMINISTRATION AND DEFENCE; COMPULSORY SOCIAL SECURITY; EDUCATION; HUMAN HEALTH AND SOCIAL WORK ACTIVITIES; ARTS, ENTERTAINMENT AND RECREATION; OTHER SERVICE ACTIVITIES; ACTIVITIES OF HOUSEHOLDS AS EMPLOYERS; UNDIFFERENTIATED GOODS- AND SERVICES-PRODUCING ACTIVITIES OF HOUSEHOLDS FOR OWN USE; ACTIVITIES OF EXTRATERRITORIAL ORGANISATIONS AND BODIES]

2. **Organisation type**   How would you describe your organisation? For example, "Bank", "University", "Local government", etc

3. **Size** How large is/was the organisation or business unit your response relates to? Please use this organisation / business unit as the basis for subsequent answers.

    [Single individual; Micro enterprise (between 2 and 9 employees); Small enterprise (between 10 and 49 employees); Medium sized enterprise (between 50 and 249 employees); Large enterprise 1 (250 to 999 employees); Large enterprise 2 (1,000 to 4,999 employees); Large enterprise 3 (5,000 to 9,999 employees); Large enterprise 4 (more than 10,000 employees); Other]

4. **Location** Where is/was the organisation your response relates to. If the data loss relates to a multinational, but affected only a single region of that multinational, please select that region.

    [Africa; Asia; Central America; Eastern Europe; European Union (including the UK); Middle East; North America; Oceania; South America; The Caribbean; Multinational; Other]

Page 3: The data loss event

The Issue

These questions relate to the nature, magnitude and timing of the data loss event.

5. **Nature of the issue** What was the nature of the data loss event. For example, "Total data loss", "Partial data loss", "Ransomware attack", etc.

6. **Magnitude of the data loss - 1** How much data was involved?

   [0 - 999 megabytes (under 1 gigabyte); 1 - 499 gigabytes; 500 - 999 gigabytes (under 1 terabyte); 1 - 499 terabytes; 500 - 999 terabytes (under 1 petabyte); 1 petabyte or more; Other]

7. **Magnitude of the data loss - 2** If you have a relatively precise figure (e.g. 77 terabytes) for the amount of data involved, please tell us.

8. **Date - 1** When did the issue occur?

   [2022; 2021; 2020; 2019; Between 2017 and 2018; Between 2015 and 2016; Between 2010 and 2014; Between 2005 and 2009; Between 2000 and 2004; 1999 and earlier; Other]

9. **Date - 2** If possible, please provide a more accurate date than the year bands above.

The Event

These questions relate to the data loss event (and events leading up to the data loss).

10. **Summary of the issue** What happened that led to the data loss? Please provide as much information as you can.

11. **Summary of the impact** What was the impact on your organisation of the data loss? Please include not only immediate effects, but also any knock-on effects (such as job losses, damage to reputation, increased costs). If possible, please also indicate how far after the event the knock-on effect was identified.

12. **Summary of the mitigation** What did you do to mitigate against any future data loss incident? Please provide as much details as you can including (if possible) the costs related to implementing your mitigation strategy.

Page 4: Costs

These questions are intended to allow us to understand the financial and human impact of the data loss event.

13. **Currency** If you are telling us about any costs, please choose a currency so we can normalise the values.

     [€ - Euro; £ - GBP; $ - USD; Other]

14. **Overall cost/value of the loss - 1**
     Please provide a range for the overall cost/value of the loss in the currency selected above. Include, if possible, the value of the data, what it cost you to recover the data and recover from the loss (if indeed you did recover) and any secondary knock-on costs you have been able to quantify.

     [1-999; 1,000-4,999; 5,000-9,999; 10,000-49,999; 50,000-99,999; 100,000-499,999; 500,000-999,999; 1,000,000-4,999,999; 5,000,000-9,999,999; 10,000,000-49,999,999; 50,000,000-99,999,000; 100,000,000-499,999,999; 500,000,000-999,999,999; 1,000,000,000 or more; Other]

15. **Overall cost/value of the loss - 2** If you have an actual figure (or figures) for the overall cost/value of the loss please provide it here. Include, if possible, the value of the data, what it cost you to recover the data and recover from the loss (if indeed you did recover) and any secondary knock-on costs you have been able to quantify. Please indicate what your figure covers (the list above might help). If you can itemise the costs that would be fantastic.

16. **Value information provenance - 1** Is the figure(s) above an estimate or is it backed by any form of standard financial / cost analysis / accounting process?

     [The figure is an informed estimate; The figure is an estimate based upon some forms of standard accounting practices; The figure is based upon relatively rigorous standard accounting practices; Other]

17. **Value information provenance - 2** Please try and quantify the rigorousness of the figure(s) above using the following scale.

     [Guess, low degree of confidence; Estimate; Informed estimate and some standard techniques. Reasonable degree of confidence; High degree of confidence; Rigorous accounting techniques applied throughout]

18. **Value in relation to turnover** What was the magnitude of the loss in relation to your normal organisation or business unit turnover?

     [Less than 1%; 1% to 3%; 4% to 5%; 6% to 9%; 10% to 14%; 15% or more; Other]

19. **Were any jobs lost?** Were any jobs lost as a direct or indirect consequence of this loss?

     [None; 1-2; 3-5; 6-10; 11-49; 50-100; 101-499; 500-1,000; More than 1,000; Other]

Page 5: Other information sources

Access to any additional sources of information that relate to this data loss will help us to understand it better.

20. **Uublished information - 1** Is there any publicly available information about your data loss you could provide us with links to? For example, news stories, corporate website information, and so on.

21. **Unpublished information - 1** Is there any private information about your data loss or information that is public, but not available on the open internet that you could provide in confidence for our researchers to look at? For instance, internal reports, intranet stories, root cause analysis, and so on.

22. **Additional Information** Is there ANY additional information that you feel may be of use when it comes to understanding the magnitude of the effect on your organisation? This could include things such as loss of contracts, difficulties in obtaining insurance, delays in progressing plans for expansion, etc. If so, please enter it here using as much detail as possible.

Page 6: (Possibly) Taking things further

A confidential follow-up conversation with one of the researchers could be very helpful as regards understanding the impact of your information loss. If you are willing to take part in such a discussion please provide us with contact information.

You may wish to use an anonymous remailer if you have concerns regarding your privacy. This page provides links to such services: https://nordvpn.com/blog/free-anonymous-email-account/

Should you decide to provide us with contact information you agree that you have asked us to process it as described in our standard privacy notice (at https://www.jisc.ac.uk/website/privacy-notice). You may instruct us to stop processing it at any time by e-mailing paul.stokes@jisc.ac.uk. This will remove your data from future statistics and reports. Until then, we'll use it only to produce the statistical analyses and reports described in the introduction.

23. **Name/pseudonym**

24. **Email address**

25. **Phone number**

26. **Alternative contact information**

27. **Case study** Your data loss story may be suitable for use as a a case study (either as a story in its own right, or as a part of a case study with multiple use cases / subjects). If it is, may we publish it in a suitably anonymised fashion as a case study within the report and possible subsequent reports?

28. **Publication - Organisation** Information that identifies individuals or organisations will not ever be published without permission. If you have provided any information that identifies the organisation associated with the data loss (or you are willing to do so) are you willing for us to use the name of that organisation in the research report and possible subsequent reports?