Digital**Preservation**Coalition

THE GLOBAL BIT LIST OF ENDANGERED MATERIALS

November 2025

Table of Contents

1. Message from the Executive Director

The Global *Bit List* of Endangered Digital Materials reports the threats to digital materials and the impact of efforts to reduce those threats. It has the potential to offer positive and negative insights, tempering overstated fears while challenging unfounded complacency. It is the voice of a global expert community, drawing impartially on their decades of insight and experience. It places their concerns before the world. It challenges those in authority to reflect deeply and act promptly to ensure a sustainable future for our digital assets.

As the report makes clear, digital preservation is a global challenge and a generational one. The threats faced by our shared digital legacy are not hypothetical: they are real, they are accelerating, and they arise in all sorts of new and unforeseen ways.

The challenges we face in digital preservation are in part a measure of the success and scope of the digital transformation. Very few areas of our lives have not been altered by technology, consequently very many areas of our lives are affected by deletion, obsolescence and loss. There is no single cause. Our digital lives are enmeshed in complex supply chains that provide personalised access to information and services. Every point in that supply chain is vulnerable and any one of them can be disrupted so that access and interactions are lost. In most cases those supply chains can be re-established with time and effort but, as the *Bit List* tells us, we cannot take that for granted.

We live in times of conflict, such that the persistence of digital resources may seem of no concern. Ending land wars, cyber-wars, information wars; healing displacement, disaffection and disinformation; solving the climate crises, ending human suffering and building sustainable economies: surely these are the preoccupations of our time? Only if we under-estimate the necessity and the opportunity of robust digital infrastructure.

We may seek healthier, wealthier, safer and smarter countries and communities; we may push for more transparent, more sustainable and more responsive processes. It is hard to envision these positive outcomes without long term thinking; it is hard to imagine how these can be progressed without digital technologies. Put that together and without digital content, preserved and accessible, these long-term goals will remain elusive. And the converse: it is hard to imagine long term goals when the basic tools of sustainable progress face such risk.

Born digital materials are 'born vulnerable'. So are the plans that are built on them.

The risks are well documented, and they are only increasing, but there are reasons to be hopeful in this report too. The frequently announced digital dark age is offset, not for any inherent reasons, but because by observing the risks, action is made possible. That is the response of the digital preservation community. Digital materials do not preserve themselves. They require people, organisations and resources. They require strategies, infrastructures and collaboration. Without urgent action, we risk losing the digital record of our time—our stories, our governance and our culture.

The *Bit List* makes this much clear: digital preservation is not a hypothetical anxiety for the few. It is an imperative for us all.

2. Executive Summary and Reiteration of Recommendations

The DPC's *Global 'Bit List' of Endangered Digital Materials*—or the *Bit List*—is a powerful call to action from the global digital preservation community. It highlights the digital content most at risk of being lost, based on real-world conditions and expert insight. By clearly identifying both the urgency of the threat and the value of the content, the *Bit List* shines a spotlight on the digital materials that demand immediate attention to ensure their survival. The *Bit List* is reviewed and updated every two years, with the last edition in 2023.

Upon publication of the *Bit List* in 2023, the *Bit List* Council embarked upon a major review of its structure and process with a view to making it simpler to understand, easier to use and easier to maintain—helping the *Bit List* better serve its main purpose as a tool for advocacy. The result is this report, and a new accompanying online version, updated and published for World Digital Preservation Day in November 2025.

The first update users will notice is a change to the name of the resource. Previously, the *Global 'Bit List' of Endangered Digital Species*, the reference to 'species' has been exchanged for 'materials.' The early iterations of the *Bit List* were inspired by the [IUCN Red List of Threatened Species](#) and this served as a useful model when the *Bit List* contained fewer entries, but feedback suggested that the reference was no longer required and in fact a more straightforward terminology would be more effective. So, we have a subtle change to the name, and in the list where previously we referred to 'species' we use 'entries;' in place of 'species groups' we have 'categories.'

An addition to the terminology is 'threats.' A major outcome of the *Bit List* Council's review of structure was the recommendation to frame the entries using a threat-based structure. A threat refers to a specific source of identifiable danger, with the potential to cause of harm or damage. The *Bit List* Council felt that being able to identify these sources, structure the *Bit List* report around them and target efforts in their direction would enable this resource to more effectively be deployed to support advocacy endeavours. In this report therefore, users will find entries arranged first by 'threat' then by 'risk classification.'

Arranging the *Bit List* entries in this way also highlighted imbalances within the entries and prompted the consolidation of some entries which had previously been expanded. Again, the aim of the consolidation was to simplify the *Bit List* and to make it easier to use.

Users of the new online version of the *Bit List* may navigate the resource as best serves their purpose. A key word search offers the ability to directly discover individual entries, while categories and risk classifications enable the identification of groups of related entries.

Therefore, while no review of the risk to individual entries has taken place, there are many new features of this 2025 report, and we retain and reiterate the recommendations from the previous report.

The DPC calls on our members, partners and colleagues globally to take four steps:

- Enable and support the preservation of digital materials emerging as a form of protest in the context of political upheaval, and to do so urgently as the threats to the loss of such materials are pernicious and immediate.
- To address the gaps in treaty provision that would enable the preservation of digital cultural heritage at a time of conflict, such as envisaged in the 1954 UNESCO Convention on the

Protection of Cultural Property in the Event of Armed Conflict and related. Recognizing that cyberwarfare can make a battlefield of any connected device, digital resources are inevitably and already enmeshed in numerous conflicts, hot and cold. Digital cultural heritage should be granted special legal protection from electronic warfare such that attacks on the memory of the world may be treated not simply as cybercrimes, but in relevant conditions also as war crimes and in cases of cultural genocide as crimes against humanity.

- To remove and reduce barriers to the preservation of social media, enabling reasonable preservation actions by trusted and legitimate actors.
- Continue and expand the research and provision of digital preservation capability.

We call on auditors, regulators, legislators to formulate plans that:

- Demand a higher standard of competence and attention to digital preservation in the context of regulated industries and public authorities to prevent data loss, recognizing the reputational and real harms to stakeholders, to themselves and to future generations that arise, and which are entirely avoidable.
- Ensure higher standards of data management are achieved through the introduction of systematic benchmarking of the digital preservation competence of regulated and statutory agencies that rely on data.
- Challenge short-termism within the sectors and agencies they regulate so that minimum retention periods are not misinterpreted as maximum, and ensuring value can be obtained from data gathered at great expense.

We call on data controllers, chief technology officers and corporate audit committees to:

- Use the *Bit List* to understand the risks they face, taking steps to remove, reduce or mitigate them.
- Fold digital preservation policies into cyber-resilience strategies.
- Recognize that long term commitments cannot be met solely on a project basis, and therefore to fold short term and exploratory digital preservation projects into longer term strategic plans.
- Prioritize the deployment technologies, the development of policy and the employment of staff to address digital preservation risks and the harms that result from data loss.
- Benchmark their digital preservation maturity as an organization and ensure that the capacity of staff and paid for services match their stated goals and aspirations.
- Assess and adopt new technologies based on the total cost of ownership, recognizing that longer term solutions may have larger up-front costs but will save money in the long term, avoid the accrual of technical debt and enable longer term exploitation of data and services.

We call on courts and law enforcement to:

- Use the full extent of the law to prosecute data losses that arises from criminal negligence or malfeasance, especially those deletions that expose professional misconduct or defeat public accountability.

We call on the global digital preservation community, especially our members, supporters and partners to:

- Help promulgate the *Bit List*, and report ways in which it has been used to advocate for and prioritize digital preservation within their institutions.
- Advise about omissions and updates from the *Bit List* to enable a continuous and informed cycle of revision to help maintain its currency and authority.
- Use the *Bit List* as a tool to roadmap the development of tools and solutions that will tackle some of the more intractable technical challenges reported here.
- Use the *Bit List* to guide and support students and new entrants into the digital preservation workforce so that they are suitably prepared for the material challenges of their chosen profession.

By its nature, the *Bit List* remains a provisional statement. It is published with the understanding that risks and experiences in digital preservation are emergent. The extent of the digital domain, the complexity of the threats, and the sophistication of emerging solutions mean that no process could ever fully capture the risks and challenges faced by digital content around the world.

The *Bit List* Council also recognizes that differences in emphasis and subtleties of local context may well have been overlooked, and that material changes may have occurred during the process of compilation, which should be taken into consideration for the next revision. We welcome corrections and suggestions on how the list could be improved, sharing of your own case studies or examples, and nominations of at-risk materials, and invite you to participate in the global community which is developing good practice around digital preservation in the enhancement of the list for its next comprehensive review scheduled for November 2027.

3. Acknowledgements and thanks

The production of the 2025 Global *Bit List* of Endangered Digital Materials is made possible through the generous support of the *Bit List* Council and their respective networks.

All entries to the *Bit List* are reviewed by these dedicated organizations which represent global expertise in the preservation of the listed digital species. The DPC gratefully acknowledges all of their support, commitment and contributions towards the publication of the *Bit List*.

- Tom Shaw
- Trevor Owens, AIP
- Kieron Niven, Archaeology Data Service
- Jo Fleming, Art Gallery of New South Wales
- Neil Jefferies, Bodleian Libraries/ Open Preservation Foundation
- Michael Day, British Library
- Leontien Talboom, Cambridge University Libraries
- Alicia Wise, CLOCKSS
- Kevin Long, Digital Repository of Ireland
- Amanda Tomé, Digital Research Alliance Canada
- Ellie O'Leary, DPC
- Andy Jackson, DPC

- Alex Chan, Flickr Foundation
- Hannah Smith, Historic Environment Scotland
- Mark Reynolds, Houses of the Oireachtas
- Pedro Gonzalez-Fernandez, Library of Congress
- Kate Murray, Library of Congress
- Meghan Lyon, Library of Congress
- Abbie Grotke, Library of Congress
- Scott McCoy, National Film and Sound Archive
- Maria Ryan, National Library of Ireland
- Trond Teigen, National Library of Norway
- Dan Eady, Nga Taonga
- Kate Roberts, Nga Taonga
- Robin Friend, PRO Victoria
- Somaya Langley, Science Museums Group
- Sofie Niemi, Swedish Tax Agency
- Paul Young, The National Archives (UK)
- Tom Storrar, The National Archives (UK)
- Angeline Takawira, UN IRMCT
- Matt Barr, University of Glasgow
- Zoe Bartliff, University of Glasgow
- Elizabeth Thompson - Macrae, University of St Andrews
- Sharon Webb, University of Sussex
- Alexander Roberts, University of Swansea
- Amanda May, US NARA

Additional subject matter expertise was provided by the following members of the digital preservation community, without which the *Bit List* would have been incomplete and for which we thank them unreservedly.

- Marcus Baw, Baw Medical

As the DPC Subcommittee responsible for the resource, members of the Advocacy and Community Engagement (ACE) Subcommittee also provided valuable advice and an editorial review of the *Bit List* before its publication for which we are very grateful.

- Paul Stokes, Jisc
- Joanna Fleming, Art Gallery of New South Wales
- Ray Moore, The City of Edinburgh Council
- Josie Partridge, University of Birmingham
- Kathryn Rowswell-Faust, EUMETSAT
- Rachel Talbot, British Geological Survey
- Niamh Murphy, University College Dublin
- Mia Willis, Historic England
- Laura Drake Davis, Library of Congress
- Veronica Berry, Library and Archives Canada
- Lori Stirland-Legret, University of Bristol
- Eren Delaney, University of Warwick
- Leanne Harrington, Guinness Archive (Diageo)
- Delaney Sweep, University of Calgary
- Kari May, University of Pittsburgh Library System
- Leona Fearon, National Library of Ireland

4. Aims, Objectives and Limitations of the *Bit List*

The *Bit List* is a free, open and community-sourced advocacy tool developed by the Digital Preservation Coalition (DPC) to highlight digital materials at risk of loss. Drawing on the experience of the global digital preservation community, the *Bit List* was first released in 2017 and is reviewed every two years with updates in between. It identifies a wide range of endangered digital content across diverse organizational settings. These risks may stem from technological obsolescence, media decay, or broader issues such as accountability, policy, and business processes.

The list is based on real-world experience from professionals who work to keep digital materials accessible over time. These experts come from various countries and sectors. The *Bit List* is not a top-down or theoretical project, and it does not serve political or commercial interests. Entries to the list are included because experienced professionals have faced challenges preserving them and believe they should be highlighted.

While the *Bit List* covers all types of digital materials, it does not include everything which is at risk— only those that the community has identified. If a digital material is not listed, this does not mean it is safe—just that it has not yet been reported. On the other hand, if a digital material is listed, it shows that some effort has been made to protect it. Most entries are broad and may overlap, which can affect how urgent the need for action is.

The *Bit List* serves three main purposes for three key audiences:

- **As an advocacy tool**, it aims to influence the technology sector and senior decision-makers by highlighting the urgent risks facing digital materials. It encourages informed, long-term investment in digital preservation.

- **As a practical guide**, it offers a simple roadmap for tackling preservation challenges. The *Bit List* invites all contributors to the digital preservation community to address the issues raised and recognizes progress toward solutions.

- **As a current overview**, it provides up-to-date guidance for the digital preservation community. As a resource, it aims to be especially helpful for newcomers and educators, while also supporting experienced professionals facing new challenges and those making the case for digital preservation within their organizations.

The *Bit List* offers a provisional commentary with the recognition that the extent of the digital domain, the complexity of the threats, and the sophistication of emerging solutions mean that no process could ever fully capture the risks and challenges faced by digital content around the world. It is published and reviewed with the understanding that new risks are continuously arising; every day and (inevitably) between editions of the *Bit List*. Members of the *Bit List* Council recognize that differences in emphasis and subtleties of local context may well have been overlooked, and that material changes may have occurred during the process.

The *Bit List* is designed to be collaborative, iterative and provisional. Thus, if readers are aware of significant digital collections that do not match up with any of the broad examples given but are at material risk, they are encouraged to draw these to the attention of the *Bit List* Council through the DPC's Chief Community Officer. These will be reviewed in time for publication of the next scheduled comprehensive review of the *Bit List*. Where digital materials face an imminent extinction event before that, their evaluation may be accelerated and an addendum published to the *Bit List* in order

to provide the timely, impartial and expert advocacy that may be required. Corrections, comments and nominations are welcome.

5. Using the *Bit List*

a. Accessing the List

In order to best support the advocacy endeavours of the digital preservation community, the *Bit List* is offered both as an online resource and a downloadable report (this document), to facilitate the various ways in which its information is used.

Online, the *Bit List* provides a searchable set of entries and a means to browse the listing by category and risk classification.

The downloadable *Bit List* report contains observations about the context in which the entries have been reviewed, as well as recommendations and key messages for particular sets of actors and stakeholder groups. Within the report, entries are arranged first by *threat* then by *risk classification*. Individual entries are then presented in alphabetical order. A full index of entries is provided as an appendix to this report.

b. Interpreting the entries

In both formats, each entry contains the following information:

Non-current Portable Magnetic Media (1)   (3)

Materials saved to floppy disks, tape, portable hard disks or other numerous magnetic storage devices where the media is out of warranty and reader devices may no longer be supported or integrated easily into hardware infrastructure: typically, more than five years old. (2)

Portable Media (4)

Examples

Floppy disks; tape; certain kinds of portable hard disks, zip drives. (5)

Imminence (6)	Effort (7)
●●●●○	●●●○○
4/5	3/5
Action is recommended within twelve months. Detailed assessment is a priority.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.
Hazards (8)	**Mitigations (9)**
Poor storage; inability to access readers; no replication; encryption; aggressive compression; uncertainty over IPR or the presence of orphaned works. Hardware/Media	Active management; dependable access to readers; strong documentation; documentation independent from the media.

Bit List History (10)

Added to list: 2019 (11)

2024: No change. →	2023: No change. →	(12)

Last Review (13)

The 2019 Jury introduced this entry to ensure that the range of media storage is properly assessed and presented. Portable magnetic media is ubiquitous but is fragile not just to physical wear and tear but also to magnetic interference and bit-rot. The substrates of the disks can prove unstable, and in some cases, proprietary reader technology means that the disk becomes obsolete before it degrades. Storage at scale also means the percentage likelihood of failure increases. The 2021 Jury and 2022 Taskforce agreed with the entry's assigned risk classification with no noted trend towards increased or reduced risk.

The 2023 Council agreed with the risk classification of *Critically Endangered* with the overall risks remaining on the same basis as before ('No change' to trend). Additionally, a new entry 'Non-current Rare Portable Magnetic Media' was created as a split, related standalone entry to highlight the increased risk.

The 2024 Interim Review Council agreed These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information (14)

There is no 'active management' of data found on these media items. The data should be copied off of the media and into a digital preservation system that allows for active management. Data found on these media should be considered a backup, at best.

We know what to do with this type of material, it is the scale that makes it a problem.

There is really no excuse for using floppy disks for storage these days. Tape is a different proposition since it allows high-density back up offline and nearline. But there are challenges with the backward compatibility of popular and even relatively recent LTO versions.

Case Studies & Examples (15)

- The Magnetic Tape Alert Project, Information for All Programme (IFAP) of UNESCO, in cooperation with IASA, the International Association of Sound and Audiovisual Archives, to alert stakeholders of the imminent threat of losing access to their audiovisual documents. The project included a survey of existing audiovisual documents on magnetic tape not yet digitally preserved. See **The Magnetic Tape Alert Project**, Pace, A. (2020), *IASA & UNESCO Information for All Programme* [accessed at 2023-10-24].
- **8" Disk Recovery: Kryoflux and Catweasel**, de Vries, D. (2015), *OPF Blog* [accessed at 2023-10-24].
- The British Library's Flashback Project, a proof-of-concept that explored the practical challenges of preserving digital content stored on physical media (magnetic and optical disks) using a sample of content from hybrid collection items dating from between 1980 and 2010. See **The Flashback Project: Rescuing disk-based content from the 1980s to the present day**, Pennock, M., May, P., Day, M., Davies, K. and Whibley, S. (2016), *11th International Digital Curation Conference, Amsterdam, Netherlands, 22-25 February*.

Keep Me Informed (16)

(1) Title – Each item on the list is given a short title to represent the digital material represented.

(2) Description – Each entry is described in general terms.

(3) Risk Classification – based on an evaluation by the *Bit List* Council, each entry is assigned a risk classification. This takes into consideration the imminence of action required to safeguard the material, the effort required to preserve the material and the inevitability of loss. The classifications assigned start at 'Lower Risk' and escalate to 'Practically Extinct' (see page 8 for the full list of classifications and their definitions). In each case, the context in which the digital material exists, as well as the presence of aggravating factors, are used to guide the evaluation.

(4) Category – entries belong to broad categories to aid users with discovery and identification. Entries overlap; any given entry may appear under multiple categories depending on technology, resourcing or organizational contexts. These overlaps mean items may also indicate greater risk given the that actions to tackle the risks are potentially more complex.

(5) Examples – examples are illustrative not exhaustive and in many cases these examples are also broadly defined, representing many specific instances. The examples typically include specific submissions made in the open nomination process and examples arising from Council discussions.

(6) Imminence – One of the evaluation criteria, it invites nominators and the *Bit List* Council to consider how quickly you need to act to save this digital material? This indicator is rated 1 – 5, where 5 = Act immediately within 6 months; 4 = Around 12 months; 3 = Less than 3 years; 2 = Within 5 years; 1 = More than 5 years.

(7) Effort to preserve - One of the evaluation criteria, it invites nominators and the *Bit List* Council to consider whether loss of this digital material could be avoided, and what the effort to preserve might be? This indicator is rated 1 – 5, where 5 = Inevitable; 4 = Loss seems likely; 3 = Major effort to fix; 2 = Small effort to fix; 1 = Entirely avoidable.

(8) Hazards - Each entry includes examples of **Hazards** which amplify the risks to a digital object and relate to the broader threats which are faced by the digital object. Entries are linked by the threats they face. Recognizing that entries are very broadly defined, digital materials may be at more or less risk depending on local circumstances. Where associated threats have been identified, these appear at the foot of the Hazards section.

(9) Mitigations – each entry offers a set of **Mitigations** or good practice that would reduce the risk imposed by the **Hazards** and threats. These are also implied recommendations for addressing and reducing risks to be followed in the timescale indicated. In most cases, a fuller assessment is also suggested. By implication, the actions that would arise from such an assessment are not likely to be trivial.

(10) *Bit List* History – this section tracks the trajectory of the entry on the *Bit List* since its first nomination and includes commentary from the *Bit List* Council to help users understand the rationale behind evaluations and decisions made.

(11) Added to list – every two years, the *Bit List* Council calls for new nominations to the *Bit List*. The year that each entry was nominated and added to the list is displayed here.

(12) Trends - Each item notes the trend from the preceding and current year with commentary. There are three classes of trend relating to the risks to clarify their meaning, which are summarized below:

- *Towards greater risk* - where the Council has reason to believe preservation is becoming significantly harder than the previous year. This may not be sufficient to cause the entry to change risk classification but indicates movement within the risk classification.
- *Material improvement* - where the Council has reason to believe **Trends** towards reduced risk have accelerated. This may not be sufficient to cause the entry to change risk classification but indicates movement within the risk classification.
- *No change* - In every other case there is no change to the trend. This does not mean the trend has stopped but that the Council believe the trend has continued as reported last year.

(13) Last Review – detailed comments from the *Bit List* Council pertaining to the most recent review cycle have been included where available. Often these follow and incorporate details from previous review cycles where relevant, providing a historical account of the entry on the *Bit List*.

(14) Additional Information – where available and relevant, any further useful information pertaining to each entry is provided.

(15) Case studies and examples – where these have been offered by nominators or the *Bit List* Council, case studies and examples in support of the *Bit List*'s assessment of each entry are included. Where the community is aware of better or more current examples or case studies, they are invited to share these for inclusion in the *Bit List*.

(16) Keep me informed (online only) - this link offers a way to express your interest in a particular entry.

c. Definitions

To assist with interpretation and reading of the *Bit List*, the terms used are defined as follows:

i. Threats

a) Organisational threats

Organisational threats to digital materials are risks that come from within an organisation. These can be political, human, or legal in nature. For example, if a company goes through major changes or shuts down, its digital assets may be lost or become inaccessible. Other risks include a lack of awareness, interest, funding, or the necessary skills to properly preserve digital materials. Complex systems or unclear responsibilities can also overwhelm staff or lead to important files being overlooked. Additionally, if there are no consistent policies or procedures, some materials might be wrongly excluded from preservation. Even when an organisation has the resources and willingness to preserve digital content, legal restrictions can sometimes prevent it.

- **Complexity Limitations -** many factors conspiring to overwhelm resources
- **Lack of Standardisation** - the absence of consistent, widely adopted practices, formats, protocols or frameworks for preserving digital content over time e.g. file formats, metadata schemas, workflows, storage, infrastructure or policy. This can create significant challenges for organizations trying to ensure long-term access to digital materials.
- **Legal Constraints** - laws, regulations and contractual obligations that affect how digital materials can be preserved, accessed, shared or stored over time. These constraints can vary widely depending on jurisdiction, sector, and the nature of the

digital content but may include copyright or data protection laws, access and use restrictions, jurisdictional conflicts and retention schedules.

- **Organisational Change** - Business changes or an organisation no longer exists.
- **Organisational Limits** - Lack of awareness, business interest, financial resources, and/or skills.
- **Out of Band** - Material ignored because it does not come from a 'recognised source for digital preservation.'
- **Sensitive Data** - digital information that requires special handling due to its potential to cause harm, violate privacy or breach legal or ethical obligations if improperly accessed, disclosed or preserved e.g. personally identifiable information, confidential business information, classified or government-sensitive data, culturally or ethically sensitive data, legal data, health or medical records.

b) Complexity threats

Complexity threats arise when it's unclear who owns the data or who has the right to access it. This can happen with data stored on specific platforms, data that is proprietary, or data meant only for short-term use—like messages on chat apps that aren't designed to be saved. Another issue is data that lacks proper metadata or has inconsistent information, making it difficult to find, interpret, or preserve.

- **Complex Platforms** - digital systems or environments that are technically intricate, multifaceted or difficult to manage and maintain over time e.g. proprietary or custom-built systems, interactive or dynamic content, software dependent environments, cloud-based platforms, multimedia systems, distributed, federated or legacy systems.
- **Ephemeral Services** - Services designed not to retain information
- **Lack of Metadata** - e.g. Provenance/chain of custody, Versioning, Code-book (e.g. Database structure, spreadsheet column contents), IPR
- **Proprietary Data** - digital content or information that is owned by an individual, organization, or vendor and is subject to restrictions on access, use, modification, or distribution. This type of data is often tied to specific software, formats, or platforms that are not openly documented or freely available.

c) Content threats

Digital content encompasses the files, metadata and storage media essential for preservation. Content threats arise when integrity is compromised, metadata is inaccurate or incomprehensible, and the appropriate storage media is no longer available to ensure retrievability – thus many Content threats are likely to overlap with Organisational and Technical threats. Any one of these Content threats may lead to loss of accessibility, authenticity and integrity.

- **Data Types/File-Formats** - Software Obsolescence, lack of standards but also Digital Rights Management (DRM)

d) Technological threats

Obsolescence and a lack of standards are central to technological threats to digital preservation. As older hardware, software and file formats become outdated or unsupported, this makes it difficult or even impossible to access or read digital materials

created with them. For example, files stored on floppy disks or created with now-defunct software may no longer be accessible without specialized tools or emulators. At the same time, the lack of widely adopted standards for file formats, metadata, and preservation practices may lead to inconsistencies in how digital materials are stored and managed. Without standardization, it becomes harder to ensure long-term accessibility, interoperability, and authenticity of digital content across different systems and over time.

- **Hardware/Media** - physical components and storage devices used to store digital content.
- **Proprietary and Custom-built Hardware** - physical devices or systems that are either owned and controlled by a specific vendor or manufacturer using non-standard components or interfaces (proprietary) or uniquely designed for a specific organization or purpose often without broad documentation or support (custom-built).
- **Software** - the programs, applications, and operating systems that are used to create, manage, access, and preserve digital content e.g. commercial or embedded software

ii. Risk Classifications

Practically Extinct

Digital materials are listed as *Practically Extinct* when examples cannot be identified or are inaccessible by most practical means and methods. It does not assume that the material is lost, but rather that loss is imminent and immediate action is required to avoid loss. It includes material where recovery is possible in very small samples but is impractical or has not been demonstrated at scale.

Critically Endangered

Digital materials are listed as *Critically Endangered* when they face material technical challenges to preservation, there are no agencies responsible for them or those agencies are unwilling or unable to meet preservation needs. This classification includes *Endangered* materials in the presence of aggravating conditions.

Endangered

Digital materials are listed as *Endangered* when they face material technical challenges to preservation or responsibility for care is poorly understood, or where the responsible agencies are poorly equipped to meet preservation needs. This classification includes *Vulnerable* materials in the presence of aggravating conditions.

Vulnerable

Digital materials are listed as *Vulnerable* when the technical challenges to preservation are modest but responsibility for care is poorly understood, or where the responsible agencies are not meeting preservation needs.

Lower Risk

Digital materials are listed as *Lower Risk* when they do not meet the requirements for other risk classifications but where there is a distinct preservation requirement. Failure or removal of the preservation function would result in assignment to one of the more threatened classifications.

iii. Categories

Categories are used to group entries of a similar nature together, to aid discovery and identification. The categories employed within the *Bit List* are:

- **Apps:** apps or applications are types of software designed to perform specific tasks or functions for a user. In The category includes apps which run on various devices, including smartphones and tablets, computers and web browsers.
- **Cloud:** the cloud refers to a network of remote servers hosted on the internet that store, manage, and process data, rather than using a local server or personal computer.
- **Community Archives:** community archives are collections of materials that document the history, culture, and experiences of a specific community, often created, maintained, and controlled by members of that community themselves.
- **Databases**: A database is an organized collection of data that is stored and accessed electronically. Databases are designed to efficiently store, retrieve, and manage data, making them essential for everything from websites and apps to enterprise systems and research projects.
- **Engineering Data**: engineering data refers to the technical information and measurements used in the design, development, analysis, and maintenance of engineering systems, products, or processes.
- **Formats:** file formats are standard ways that information is encoded for storage in a computer file. It defines how data is organized and what kind of content the file contains—such as text, images, audio, video, or structured data.
- **Gaming:** This category includes not only video games themselves but also the environments in which they have been played, user-generated content, and the social interactions they have facilitated.
- **Integrated Storage:** integrated storage refers to a unified system that combines multiple types of storage technologies—such as hard disk drives (HDDs), solid-state drives (SSDs), and cloud storage—into a single, cohesive framework.
- **Legal Data:** legal data refers to digital information that has legal significance and must be preserved to meet regulatory, compliance, or evidentiary requirements. This could include contracts and agreements, client records in law firms, email communications relevant to litigation, Government records subject to public access laws, audit trails and logs for compliance as well as metadata that proves provenance and authenticity.
- **Media Art:** media art is a form of artistic expression that uses digital, electronic, or multimedia technologies as a central part of its creation, presentation, or experience. It often blends traditional art forms with new media tools to explore innovative ways of storytelling, interaction, and aesthetics.

- **Museum and Gallery:** museum and gallery data includes all digital information generated or maintained by museums and galleries about their collections, operations, and exhibitions. This encompasses both digitized materials (converted from physical formats) and born-digital content (created in digital form).
- **Orphaned Works:** orphaned works are creative works—such as books, films, photographs, music, or artworks—for which the copyright holder cannot be identified or located, even after a diligent search. This makes it difficult or impossible to obtain permission to preserve this work.
- **Personal Archives:** personal archives are collections of materials created, received, or accumulated by individuals that document their personal, professional, or creative lives. These could include emails and correspondence, photographs and videos, social media posts, blogs, academic or professional work, digital art or creative projects as well as legal and financial documents.
- **Political Data:** political data refers to digital content that is either created for political purposes (e.g. election data, legislative records, political speeches, campaign materials) or is material which has been influenced by political decisions and sensitivities.
- **Portable Media:** portable media refers to removable storage devices that are used to store, transfer, or back up digital data. These devices are designed to be easily transported and connected to different computers or systems.
- **Public Records:** public records are any digital or digitized documents created or received by government bodies or public institutions in the course of their official duties.
- **Research Outputs:** research outputs are the results or products of scholarly, scientific or creative research activities. These outputs are created during or after a research project and are intended to communicate findings, contribute to knowledge or demonstrate impact.
- **Sensitive Data:** sensitive data refers to digital information that requires special handling due to its confidential, personal, classified, or culturally restricted nature. It may include Personally Identifiable Information (PII) like names, addresses, national IDs, medical records, financial details; Confidential Business or Institutional Data like Internal reports, proprietary research, strategic plans; Classified Government Data like national security documents, intelligence reports; Culturally Sensitive Materials like indigenous knowledge, sacred texts, or images with access restrictions as well as Legal and Regulatory Data like data protected under laws like GDPR, HIPAA, or FOIA.
- **Social media:** social media refers to digital platforms and technologies that enable users to create, share, and interact with content and with each other in real time. These platforms are designed for communication, community building, content sharing and networking and may include Facebook, Twitter/X, Instagram, LinkedIn, TikTok, YouTube, Reddit etc.
- **Software:** Software refers to a set of instructions, programs, or data used to operate computers and perform specific tasks. Unlike hardware, which is the physical component of a computer, software is intangible and consists of code written in programming languages.
- **Sound and vision:** 'sound and vision' refers to a category of digital materials that includes audio and visual content—either separately or combined.
- **Web**: web refers to the vast and dynamic collection of digital content published online, including websites, blogs, social media, multimedia, and interactive platforms.

6. The Bit List Report

The Bit List Report organises the entries first by Threat (Organizational, Complexity, Content and Technological), threat type, and then risk classification.

i. Organizational Threats

    a) Complexity Limitations

Digital Archives of Music Production

Digital materials created by musicians and fans as a by-product of performance or recording, not otherwise published or shared. The use of 'archives' in this context refers to music production data that is in an archive.

Categories

- Sound and Vision

Examples

Pre-production notes; demo recordings; photography; correspondence

Imminence	Effort
●●●●○	●●●○○
Action is recommended within twelve months. Detailed assessment is a priority.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Fragile or obsolete media for offline content; service provider preservation capability for online content; dependence on proprietary formats or products; lack or loss of documentation; Uncertainty over IPR or the presence of orphaned works; lack of version control; lack of policy or mandate

Mitigations

Replication; clarity of intellectual property rights; preservation agency involved and capable of looking after content.

Bit List History

Added to list: 2014

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

In 2019, this entry was created as a subset of a previous 2017 entry, 'Digital Music Production and Sharing,' which was split to draw attention to the different challenges faced by the different forms. Although it overlaps with other entries, including 'Pre-production TV and Movie materials,' it is a separate entry to emphasize the value of the archival materials relating to the recording process over and above the recordings themselves. The 2020 and 2021 Juries, as well as the 2022 Taskforce, agreed with the entry's assigned risk classification, with risks remaining on the same basis as before (no change to the trend).

The 2023 Council also agreed with the Critically Endangered classification, with overall risks remaining on the same basis as before ('No change' to trend), though they noted that action needed to be taken more imminently.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

For the imminence of action, it will all depend on the format of the records. Correspondence or photographs may be left for longer, but recordings will need closer attention, especially if it is a bespoke recording format.

This may be less of a digital preservation challenge and more of an archive or collecting challenge. This type of material in the past, like most 'unpublished' archives, has survived through luck and is largely out of a GLAM or institution's control relying on individuals to assess and evaluate if what they have is of significance. When these types of things come to an institution, based on significance on a case-by-case basis, are these digital objects then considered 'worth' the effort to a) bring into the collection and b) care and preserve them? So yes, while *Critically Endangered*, these types of collections are enormous and quite often not things one would want to keep for the long term; however, sometimes there is the odd gem.

There is the recognized inevitable loss of existing data but reducing this loss would require major effort to fix in terms of identifying organizations who are preserving this content, and it is not clear that this is being done already.

Case Studies & Examples

- Music industry's 1990s hard drives, like all HDDs, are dying, Purdy, K. , (2024), *Ars Technica* *https://arstechnica.com/gadgets/2024/09/music-industrys-1990s-hard-drives-like-all-hdds-are-dying/* [accessed at 2024-09-13].

Exhibition Content

Born-digital or hybrid-digital content that was created and/or commissioned for exhibitions and has not been accessioned into the collection. This species is considered as an object or a 'work' in its own right, and not interpretive materials.

Categories

- Museum and Gallery

- Media Art

Examples

Digital experiences, immersive works, digital artworks, hybrid-digital artworks, media art, and other born-digital or hybrid-digital objects or materials which can take many forms

Imminence	**Effort**
●●●○○	●●●●○
Action is recommended within three years, detailed assessment within one year.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Lack of technical documentation; complex interdependencies related to specific hardware, software and/or operating systems; significant volumes or diversity of data; dependence on proprietary products or formats; lack of preservation capacity in museum or gallery; technology is seen as inherently fragile and therefore risky to collect and preserve; conflation of access with preservation; lifespan of hardware technologies used in legacy works with few/no replacements and/or alternatives; uncertainty over IPR or the presence of orphaned works

Mitigations

Strong technical documentation; preservation capability; preservation pathway; clarity of intellectual property rights that enable preservation; availability of replacement or alternative hardware technologies

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

Last Review

2023 Review

This entry was added in 2019 under 'Digital Materials in Museums and Galleries' and previously rescoped in 2021 to 'Supporting Digital Materials for Museums and Galleries'.

The 2023 Bit List Council superseded the entry, splitting it into six discrete entries as the scope of the single entry was too broad to provide the guidance needed. The recommendation to break this entry down was also made by the 2021 Jury, as the types of digital collections content in museums can be vast and offer particular risks in museum and gallery contexts. This entry draws attention to risks preserving born-digital or hybrid-digital materials, which can take many forms and may be complex digital objects. These works may be commissioned from an external third-party company, or created by staff internal to an organization, often with a considerable amount of financial investment involved. What is particularly critical is that these materials have not been accessioned into an organization's collection. These objects may inhabit exhibition spaces for many years, and while they may be a key component of an organization's business, have not been (and may not be) ever accessioned into the collection.

2024 Interim Review

The 2024 Council agreed These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

While scores should remain as is currently, a main change to note for 2024 emerging and broader risks that have come even more to the fore for the Museums and Galleries sector, and other Bit List species more broadly: Cybersecurity and ML/AI. Both these have various influences on the sector, and indeed anyone who is managing and preserving digital collections, content and/or data regardless of sector. Political risks are also on the rise globally and this cannot be ignored. For example, from Cybersecurity forum very briefly in the context of discussing the CLOCKSS Threats and **Mitigations** model.

With these in mind, they recommend a full review should be carried out to consider whether broader risks arising with Artificial Intelligence and Machine Learning should be addressed in this section or form a new sub-species (or sub-sub-species), and for considering political risk as an additional category of risk. They also recommend reviewing and scoping for a category of hybrid digital objects that aren't media art (or are found in museums that aren't art museums) including medical devices, technological developments, etc.

Case Studies & Examples

- The Preserving Immersive Media Knowledge Base is a resource created to help share information between members of the digital preservation community who are caring for virtual reality (VR), augmented reality (AR), mixed reality (MR), 360 video, real-time 3D software and other similar materials. This site was born out of Tate's Preserving Immersive Media Project with funding from the Netherlands Institute for Sound & Vision. See **Preserving Immersive Media Knowledge Base**, *https://pimkb.gitbook.io/pimkb/* [accessed at 2023-10-24].

- The Collaborative Infrastructure for sustainable access to digital art LIMA project, to prevent the loss of digital artworks and to commonly develop the knowledge to preserve these works in a sustainable way. The project 'Infrastructure sustainable accessibility digital art' invests in research, training, knowledge sharing and conservation to prevent the loss of both digital

artworks and the knowledge to preserve them. See **Collaborative infrastructure for sustainable access to digital art**, LIMA, *https://www.li-ma.nl/lima/article/collaborative-infrastructure-sustainable-access-digital-art* [accessed at 2023-10-24].

- **Virtual Reality is a big trend in museums, but what are the best examples of museums using VR?**, Richardson, J. , (2023), *Museum Next https://www.museumnext.com/article/how-museums-are-using-virtual-reality/* [accessed at 2023-10-24].

- **Learning Lessons from the Cyber-attack: British Library cyber incident review**, The British Library, (2024), *https://www.bl.uk/home/british-library-cyber-incident-review-8-march-2024.pdf/* [accessed at 2024-09-06].

- **CLOCKSS: Threats and Mitigations**, CLOCKS Trusted Digital Repository Documents, *http://documents.clockss.org/index.php/CLOCKSS:_Threats_and_Mitigations* [accessed at 2024-09-06].

Media Art by Deceased Artists or Defunct Workshops

Media art where the artists or creative technicians are either deceased or not able to provide guidance on authenticity and installation.

Categories

- Media Art

Examples

Works produced by media artists now deceased, such as: Jeremy Blake, Beatriz Da Costa, Heiko Daxl or Stanislaus Ostoja-Kotkowski

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of documentation to enable maintenance; uncertainty over IPR or the presence of orphaned works; complex interdependencies on specific hardware, software or operating systems; lack of capacity in the gallery or workshop; lack of strategic investment; complex external dependencies; loss of institutional memory resulting from staff churn; poor working relationship between the gallery and artist/workshop; lack of conservation assessment

Mitigations

Strong documentation; clarity of preservation path and ensuing responsibilities; proven preservation plan; capacity of workshop to support re-installation; capacity of gallery to conserve; capacity of gallery to re-install; retention of institutional memory including archives of correspondence between gallery and artist/workshop; strong and continuing working relationship between the gallery and artist/workshop; regular conservation assessment

Bit List History

Added to list: 2015

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 as a subset of the 2017 'Media Art,' which was first introduced with particular reference to historical media art, but split by the 2019 Jury to ensure greater specificity in its recommendation. This entry represents works held in galleries where the artist is deceased or the workshop has closed, and there is limited prospect to obtain new documentation. The 2020 Jury found a trend towards greater risk based on how galleries, which often rely on visitors for income, have been closed for extended periods and circumstances of economic dislocation. The 2021 Jury agreed on a continued trend towards greater risk based on the increasing risk of this loss happening with more time sensitivity for early media artworks.

The 2023 Council agreed with the Critically Endangered classification with overall risks remaining on the same basis as before ('No change' to trend).

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend). However, they add that it is important to bring attention to the key element and importance of providing guidance on authenticity and installation. Emulation tools are helping but the missing guidance on authenticity increases the risk.

Additional Information

This entry includes a point in the lifecycle of all media art, so good practice recommendations are likely to become more important over time. Preservation issues may not become visible until the piece is brought out of storage for loan or exhibition, underscoring the value of continuous or periodic conservation assessment. The range of data/formats/hardware/software etc. can be new and varied, providing organizations with an ongoing technical challenge that they are not initially equipped to deal with. Some loss seems inevitable.

Preservation of legacy media artworks is dependent on access to obsolete technology and also the knowledge of how to operate said technology. Documentation around the production process and artist intent can be limited and more critical without any access to artists or technicians. This creates risk around the preservation of a truly authentic artwork..

Case Studies & Examples

- **Preserving and Sharing Born Digital and Hybrid Objects**, V&A Research Projects, *https://www.vam.ac.uk/research/projects/preserving-and-sharing-born-digital-and-hybrid-objects* [accessed at 2023-10-24].

- **Preserving and sharing born-digital and hybrid objects from and across the National Collection**, Ensom, T, and McConnachie, S., (2022), *http://doi.org/10.5281/zenodo.7097489*.

- The DPC 'Preserving Digital Art' Technology Watch Guidance Note is aimed at institutions starting to collect digital art as part of a wider collecting remit. It offers basic guidance on the specificities of digital art and how it may differ from other digital content in an institution's care. See **Preserving Digital Art**, Falcão, P., (2024), *Digital Preservation Coalition* *http://doi.org/10.7207/twgn24-02*.

- NEW MEDIA MUSEUMS: Creating Framework for Preserving and Collecting Media Arts in V4, initiated by the Olomouc Museum of Art as a joint international platform for sharing experience with building and maintaining collections of new media artworks across different types of institutions. The aim of the project is to find workable methods for heritage institutions to build and maintain collections of media arts, which are necessary for safeguarding this area for the benefit of society. See **NEW MEDIA MUSEUMS: Creating Framework for Preserving and Collecting Media Arts in V4**, Central European Art Database, (2021), *http://cead.space/Detail/projects/3797* [accessed at 2023-10-24].

- The Collaborative Infrastructure for sustainable access to digital art LIMA project, to prevent the loss of digital artworks and to commonly develop the knowledge to preserve these works in a sustainable way. The project 'Infrastructure sustainable accessibility digital art' invests in research, training, knowledge sharing and conservation to prevent the loss of both digital artworks and the knowledge to preserve them. See **Collaborative infrastructure for sustainable access to digital art**, LIMA, *https://www.li-ma.nl/lima/article/collaborative-infrastructure-sustainable-access-digital-art* [accessed at 2023-10-24].

- **Saving Stan: Preserving the Digital Artwork of Joseph Stanislaus Ostoja-Kotkowsk**, Ellis, T., (2023), *https://www.ideals.illinois.edu/items/128314*.

Content Published on the Web Which Cannot Easily Be Captured Through Conventional Web Archiving Practices

Material that is not capturable via conventional web-archiving practices (i.e. remote capture with a non-browser-based crawler). The common characteristic of the material is not so much the type of content, or the context but rather the preservation risk posed to the material as a result of decisions made by the website creators, to use technologies and make design decisions that do not support the capture of the content, combined with the limitations of current web archiving processes.

Categories

- Web

Examples

Examples are wide-ranging but include: 1. Interactive content such as maps, charts, 3D models, multi-page forms etc.; 2. Content that is only accessible through search (and does not support a blank search with pagination URLs; 3. Content that is only accessible via POST or Ajax requests (i.e. 'Load More' issues); 4. Content hosted on sites which aggressively block crawlers; 5. Content hosted on proprietary platforms whose technical implementation makes web archiving difficult (Wix, Medium etc.)

Imminence	Effort
●●●●●	●●●●○
Immediate action necessary. Where detected, should be stabilized and reported as a matter of urgency.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Creation and design decisions that do not support the capture of the content; limitations of current web archiving processes; uncertainty over IPR or the presence of orphaned works

Mitigations

Continued development of tools that can capture some types of this material; continued resourcing of the web archiving community and developers working in this space; longer-term work towards a cultural change that prioritizes accessibility and achievability

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

Last Review

2023 Review

This was a new Bit List entry added in 2023 to draw attention to the particular challenges of web content that is not capturable via conventional web-archiving practices. While this entry can be considered very wide-ranging, covering different kinds of content in and across different entries represented in the Bit List, it was added to draw attention to risks and issues not fully addressed by these entries: those relating to material created and designed with technologies that do not support the capture of the content combined with the limitations of current web archiving processes. Given the ephemeral nature of the web, any material that is not preserved in a web archive is at risk; However, this entry focuses on material that is not accessible to conventional crawlers that is considerably less likely to be preserved because: 1. While the development of alternative capture tools (i.e. forms of browser-based capture) provide potential options to mitigate some of these issues, they remain imperfect and using such tools requires significant time and resources, meaning they are unlikely to be applied to very large-scale crawls; and 2. While website creators can be

influenced to avoid such problematic technologies, even in the context of current work (e.g. within the UK Government), this can be difficult.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

The online experience is becoming increasingly 'mediated' and, due to the prevalence of personalized experiences, there is little that can be seen as 'generic' which can be meaningfully captured. Add-ons such as ad, tracker and script blockers also fundamentally change how users experience the web compared to others.

There is no specific time constraint but given the relatively ephemeral nature of the web, it is likely that there is an ongoing and constant loss of material.

Given the broad scope of the entry, it can be difficult to assign an overall significance level, with some examples being trivial and others being highly important. Quantifying the impact of loss of this entry is also difficult, but it would be fair to say that it would have a significant impact on the ability of citizens to hold their governments to account and on the completeness of the historical record. Given that these issues are common across the web archiving community, this thus becomes a global problem.

Some important examples from this entry includes highly significant content of national interest that is currently difficult to capture, for instance maps showing proposed changes to electoral boundaries, government blogs and published datasets which can only be accessed through search or via 'Load more', and whole sites of national importance that aggressively block crawlers. Other pertinent examples include PowerBI and Tableau which are both increasingly widely adopted visualization tools and are very difficult to capture and also to replay. They are used to disseminate data about all sorts of things but particularly government transparency information. The mitigating action of publishing the underlying data (for example as CSV or XLS(S)) is not often observed on the web.

Case Studies & Examples

- **Fixity When Nothing is "Fixed" - Reflections on Upstream Engagement and Digital Preservation**, Beking, A. , (2016), *Digital Preservation Coalition* https://www.dpconline.org/blog/wdpd/abeking-wdpd21 [accessed at 2023-10-24].

- **Adaptability in the face of adversity - Archiving the Web to Help Persons Forced to FleWeb Archiving the 2021 Federal Election at Library & Archives Canada – A Look Behind the Curtain**, Monks-Leeson, E., White, R. and Palendat, K., (2017), *Digital Preservation Coalition* https://www.dpconline.org/blog/wdpd/lac-wdpd21 [accessed at 2023-10-24].

- **Adaptability in the face of adversity - Archiving the Web to Help Persons Forced to Flee**, Wilson, T. , (2022), *Digital Preservation Coalition* https://www.dpconline.org/blog/blog-tom-wilson-22 [accessed at 2023-10-24].

- **Archiving the architecture of Saydnaya: Using Webrecorder to capture 3D-spaces and ambient audio**, Hawes, A., (2020), *https://youtu.be/mXLSwrEHyNU* [accessed at 2023-10-24].

Legacy Media Art

Media art in storage or not otherwise displayed but where the artists or technicians are available to support installation.

Categories

- Media Art

Examples

Media art in storage

Imminence	Effort
●●●○○	●●○○○
Action is recommended within three years, detailed assessment within one year.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Lack of documentation to enable maintenance; uncertainty over IPR or the presence of orphaned works; complex interdependencies on specific hardware, software or operating systems; lack of capacity in the gallery or workshop; lack of strategic investment; complex external dependencies; loss of institutional memory resulting from staff churn; poor working relationship between the gallery and artist/workshop; lack of conservation assessment

Mitigations

Strong documentation; clarity of preservation path and ensuing responsibilities; proven preservation plan; capacity of workshop to support re-installation; capacity of gallery to conserve; capacity of gallery to re-install; retention of institutional memory including archives of correspondence between gallery and artist/workshop; strong and continuing working relationship between the gallery and artist/workshop; regular conservation assessment

Bit List History

Added to list: 2018

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2017 as 'Media Art,' which was first introduced with particular reference to historical media art. The 2019 Jury rescoped this entry to ensure greater specificity in its recommendation to represent works held in galleries but no longer displayed, but where there is a

continuing working relationship between the gallery and the artist or workshop and reasonable expectation that support for preservation could still be obtained when required. The 2020 Jury identified a trend towards greater risk, given that many museums and galleries, which often rely on visitors for income, had been closed for extended periods. Moreover, any form of digital materials that rely on an individual's knowledge is at particular risk by a pandemic. For similar reasons, the 2021 Jury also identified a 2021 trend towards greater risk, noting that digital materials in museums and galleries records are likely to be at greater risk in these circumstances.

The 2023 Council agreed with the Endangered classification with overall risks remaining on the same basis as before ('No change' to trend), while also noting a decrease in imminence of action as well as the required effort to preserve.

2024 Interim Review

The 2024 Council agreed These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

This entry attempts to capture a point in the lifecycle of media art where preservation risks are increasing but not yet critical. There is a risk that preservation issues will not become apparent until the piece is brought out of storage when considered for loan or exhibition – often on timescales that make it too late to address preservation concerns effectively. Galleries should be aware that the range of data/formats/hardware/software embedded in media art can be wide and vary at different speeds.

Sooner action is needed to prevent the material from becoming *Critically Endangered* once the artist has died or relationships break down. Where the artist is still around, there is a major reduction in the inevitability of loss and its potential to be a potentially newsworthy subject. The loss of it would be just as impactful and significant though.

Preservation of legacy media artworks is dependent on access to obsolete technology and also the knowledge of how to operate said technology. Documentation around the production process and artist intent can be limited. This is a risk in terms of preserving a truly authentic artwork.

Case Studies & Examples

- The DPC 'Preserving Digital Art' Technology Watch Guidance Note is aimed at institutions starting to collect digital art as part of a wider collecting remit. It offers basic guidance on the specificities of digital art and how it may differ from other digital content in an institution's care. See **Preserving Digital Art**, Falcão, P. , (2024), *Digital Preservation Coalition* http://doi.org/10.7207/twgn24-02.

- Archiving Australian Media Arts Project is a research project funded by the Australian Research Council involving collaboration between university researchers and cultural institutions. The aim was to develop a good practice method for stabilising digital media artworks, providing emulated access to the artworks to researchers in reading rooms, and to investigate the contemporary exhibition and re-display of historical media artworks. This has also led to Aus Easi project. See **Archiving Australian Media Arts Project**, https://aama.net.au/ [accessed at 2024-09-06].

- NEW MEDIA MUSEUMS: Creating Framework for Preserving and Collecting Media Arts in V4, initiated by the Olomouc Museum of Art as a joint international platform for sharing experience with building and maintaining collections of new media artworks across different types of institutions. The aim of the project is to find workable methods for heritage institutions to build and maintain collections of media arts, which are necessary for safeguarding this area for the benefit of society. See **NEW MEDIA MUSEUMS: Creating Framework for Preserving and Collecting Media Arts in V4**, Central European Art Database, (2021), *http://cead.space/Detail/projects/3797* [accessed at 2023-10-24].

- The Collaborative Infrastructure for sustainable access to digital art LIMA project, to prevent the loss of digital artworks and to commonly develop the knowledge to preserve these works in a sustainable way. The project 'Infrastructure sustainable accessibility digital art' invests in research, training, knowledge sharing and conservation to prevent the loss of both digital artworks and the knowledge to preserve them. See **Collaborative infrastructure for sustainable access to digital art**, LIMA, *https://www.li-ma.nl/lima/article/collaborative-infrastructure-sustainable-access-digital-art* [accessed at 2023-10-24].

Original Digital Music and Sound Recordings

Original recordings of music and other performance from which retail products are derived, typically in multiple tracks and uncompressed high-resolution sound quality.

Categories

- Sound and Vision

Examples

Original official recordings of a song, sound or performance owned by music industry

Imminence	Effort
●●●○○	●●○○○
Action is recommended within three years, detailed assessment within one year.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Single point of failure; storage on old or degrading media; lack of ongoing investment in changing preservation requirements; lack of capability; poor documentation; dependence on small staff; Uncertainty over IPR or the presence of orphaned works

Mitigations

High-quality storage; meticulous and consistent replication; trusted repository; preservation requirement understood at the executive level and funded accordingly; leadership in preservation community; expert staff

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

In 2019, this entry was added as a subset of a previous 2017 entry, 'Digital Music Production and Sharing,' which was split to draw attention to the different challenges faced by the different forms. Though it overlaps with other entries, including 'Digital Archives of Music Production,' it is a separate entry to emphasize the inherent and great value of original recordings over and above those distributed and the concomitant need for active preservation. Both the 2021 Jury and 2022 Taskforce agreed with the Endangered classification with no noted changes towards increased or reduced risk. The 2022 Taskforce additionally recommended use of the term original or primary rather than the term master, except for where it is part of a format's formal name or an industry-standard use, which is now reflected in this entry and other relevant areas of the Bit List.

The 2023 Council agreed with the Endangered with the overall risks remaining on the same basis as before ('No change' to trend).

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

The imminence of action will depend on format and age, and the significance of loss may be more largely felt if they are recordings of a major recording star.

This is interesting as the recording houses should be seeing the value of these – so why are they not taking responsibility for looking after them? Do they not feel it is in their financial interests? The archival practices of the studios are typically based on value – the recordings are assumed to be worth keeping. However, this means relatively low-value originals may not be transferred to new media in a timely way and could be lost. There is no comprehensive deposit scheme to address the long tail of music production, and it is often unclear exactly where responsibility lies.

Case Studies & Examples

- **The Day the Music Burned**, Rosen, J. , (2019), *The New York Times*
  https://www.nytimes.com/2019/06/11/magazine/universal-fire-master-recordings.html
  [accessed at 2023-10-24].

- MQA (Master Quality Authenticated), a means of digitally capturing and storing original master recordings as files without the usual loss in fidelity usually experienced with compressed audio files, has gone into administration - if you have MQA files it should still work. That being said, MQA is not lossless and it's better to get true lossless FLAC files. See

MQA has gone into administration: what does this mean for Tidal and supported products?, McKerrell, H, (2023), *What HI-FI?* *https://www.whathifi.com/features/mqa-has-gone-into-administration-what-does-this-mean-for-tidal-and-supported-products*.

- **Music industry's 1990s hard drives, like all HDDs, are dying**, Purdy, K., (2024), *Ars Technica* *https://arstechnica.com/gadgets/2024/09/music-industrys-1990s-hard-drives-like-all-hdds-are-dying/* [accessed at 2024-09-13].

- **Starting with complexity: Archiving digital-born music compositions from Mac systems of the 80s/90s**, Mattmann, B. and Lindenmann, I., (2023), *Digital Preservation Coalition* *https://www.dpconline.org/blog/wdpd/starting-with-complexity* [accessed at 2023-10-24].

Pre-Production TV and Movie Materials

Digital records of the creative and production process for film and television, such as initial designs, screenplay and script, on set still photography, rushes or out-takes that are not included in the final production and therefore not available to on-air broadcast archives or film libraries.

Categories

- Sound and Vision

Examples

TV and Movie production archives in digital form; outputs of script management software; drafts of a screenplay; continuity photography; costume design; set design; lighting and sound design

Imminence	Effort
⬤⬤⬤⬤◯	⬤⬤⬤◯◯
Action is recommended within twelve months. Detailed assessment is a priority.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of custodial responsibility; Uncertainty over IPR or the presence of orphaned works; lack of appraisal; lack of recognition of preservation at executive level

Mitigations

Preservation responsibility understood and acted upon; preservation infrastructure and planning for key items; access and use of collections to inform subsequent productions

Bit List History

Added to list: 2017

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was first introduced in 2017 and noted as being 'of concern,' though the Jury did not have the capacity to assess the entry thoroughly. Additional expertise was recruited for the 2019 Jury, and the entry was added with the Endangered classification based on the materials not being collected in any coherent way, likely loss when not valued by production companies, costly space needed to hold them, and lack of expertise needed to catalogue and collect them. In other words, this is not primarily a technical problem and advocacy is needed urgently. The 2021 Jury added the trend towards greater risk in light of the rise of streaming services. The movie and film industry has been digitizing for a long time, but COVID has brought with it the rise of streaming services like Netflix, Disney+ etc., with productions tied to their organizations. These types of pre-production materials are great for promotion before and immediately after release; ongoing preservation after promotion finishes is unclear or uncertain. The 2022 Taskforce noted no change to the trend (they agreed these risks remain on the same basis as before).

The 2023 Council agreed with the Endangered with the overall risks remaining on the same basis as before ('No change' to trend) but noted an increased imminence for this entry.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

With the importance of advocacy in mind, it is important to raise awareness and educate directors and filmmakers about managing their digital archive so that it is still accessible if donated to a cultural institution in the future.

Recently Commissioned or Completed Media Art

Media art currently displayed in a gallery or in the process of being displayed.

Categories

- Media Art

Examples

Media art recently acquired by galleries that utilize specific hardware and software in order to be accessed or exhibited

Imminence	Effort
●●●●○	●●○○○
Action is recommended within twelve months. Detailed assessment is a priority.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Lack of documentation to enable maintenance; Uncertainty over IPR or the presence of orphaned works; complex interdependencies on specific hardware, software or operating systems; lack of capacity in the gallery or workshop; lack of strategic investment; complex external dependencies; lack of documentation about artist intent; lack of understanding of costs for display and preservation.

Mitigations

Strong documentation; clarity of preservation path and ensuing responsibilities; proven preservation plan; capacity of workshop to support artwork at de-installation; capacity of gallery to conserve after de-installation; capacity of gallery to re-install work; funding understood to re-install.

Bit List History

Added to list: 2014

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 as a separate entry, but it was previously introduced in 2017 under 'Media Art' with particular reference to historical media art. It was added for greater specificity for its recommendations, to represent works acquired and commissioned in the last five years where there is a reasonable expectation that documentation has been produced or could still be obtained. While the 2020 Jury found no change in trend, the 2021 Jury discussed how prospects for long-term preservation depend entirely on whether the artwork is collected post-commission and by an organization with the resources to care for it. They agreed that the classification remains Vulnerable but with a trend towards greater risk because the imminence of action is time-sensitive, requiring working with the artist to get the documentation from them about their work and what is needed before it is too late. Furthermore, there remains a vulnerability for the smaller museums or others that do not take the preservation of media art as seriously.

The 2023 Council agreed with the Vulnerable classification with overall risks remaining on the same basis as before ('No change' to trend), although noted a change in the imminence of action from 3 years to 12 months.

The 2024 Council agreed These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend). However, it was important to note that the 'Effort to Preserve | Inevitability' can vary. Some of the works take a huge effort to preserve and that perhaps this needs some middle ground in terms of communicating those aspects.

Additional Information

By the time digital art, time-based media, etc., has entered into the permanent care of a stewarding institution, many of its technologies are already end-of-life, unsupported, or the hardware components have deteriorated. Often the expertise to maintain these many interacting components sits outside the host organization, with a technical supplier to the gallery, and this is in itself vulnerable to business change. Although there are a few exceptions, there is a need for greater capacity within the museum and gallery sector to address the challenges.

There have been new initiatives for guidance and examples of institutions taking wider sectoral responsibility for standards, which have helped with the effort to preserve, such as Matters in Media Art information resource and guidance.

Media artworks are often made with a network of knowledge that can be precarious. Documentation around production processes can be minimal, and hence acting quickly with known processes can gather information before the knowledge and people networks start to disperse. This can mean preservation of production environments and associated workflows can be preserved alongside the media.

Some art works specifically leverage the limitations and characteristics of the systems that they incorporate, often in unusual ways. This can be hard to migrate or emulate accurately.

Case Studies & Examples

- **Preserving and Sharing Born Digital and Hybrid Objects**, V&A Research Projects, *https://www.vam.ac.uk/research/projects/preserving-and-sharing-born-digital-and-hybrid-objects* [accessed at 2023-10-24].

- **Preserving and sharing born-digital and hybrid objects from and across the National Collection**, Ensom, T, and McConnachie, S., (2022), *http://doi.org/10.5281/zenodo.7097489*.

- **Guidelines for the care of media artworks**, Matters in Media Art, *http://mattersinmediaart.org/* [accessed at 2023-10-24].

- The DPC 'Preserving Digital Art' Technology Watch Guidance Note is aimed at institutions starting to collect digital art as part of a wider collecting remit. It offers basic guidance on the specificities of digital art and how it may differ from other digital content in an institution's care. See **Preserving Digital Art**, Falcão, P. , (2024), *Digital Preservation Coalition http://doi.org/10.7207/twgn24-02*.

- NEW MEDIA MUSEUMS: Creating Framework for Preserving and Collecting Media Arts in V4, initiated by the Olomouc Museum of Art as a joint international platform for sharing experience with building and maintaining collections of new media artworks across different types of institutions. The aim of the project is to find workable methods for heritage institutions to build and maintain collections of media arts, which are necessary for safeguarding this area for the benefit of society. See **NEW MEDIA MUSEUMS: Creating**

Framework for Preserving and Collecting Media Arts in V4, Central European Art Database, (2021), *http://cead.space/Detail/projects/3797* [accessed at 2023-10-24].

- The Collaborative Infrastructure for sustainable access to digital art LIMA project, to prevent the loss of digital artworks and to commonly develop the knowledge to preserve these works in a sustainable way. The project 'Infrastructure sustainable accessibility digital art' invests in research, training, knowledge sharing and conservation to prevent the loss of both digital artworks and the knowledge to preserve them. See **Collaborative infrastructure for sustainable access to digital art**, LIMA, *https://www.li-ma.nl/lima/article/collaborative-infrastructure-sustainable-access-digital-art* [accessed at 2023-10-24].

---

b)   Legal Constraints

Digital Archives from Public Enquiries and Commissions

Data from public enquiries and reconciliation commissions which can be traumatic, politically uncomfortable and contested, typically comes in many different forms and formats. Data protection issues and cultural sensitivities only amplify the challenge to preservations.

Categories

- Political Data

- Legal Data

Examples

The Tunisian Truth and Dignity Commission to investigate human rights violations committed prior to 2012; enquiries into historical child abuse; Bloody Sunday Enquiry (Saville Inquiry); East Timor Tribunal

Imminence	Effort
●●●●○	●●●●○
Action is recommended within twelve months. Detailed assessment is a priority.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Risk of falsification; fragile or obsolete media; dependence on proprietary formats or products; lack or loss of documentation; inaccessible to web harvesting technologies; lack of version control; lack of integrity checks or integrity records; poor chain of custody; inability to identify an archival authority; loss of personal testimonies; Uncertainty over IPR or the presence of orphaned works

Mitigations

Strong sense of archival responsibility; carefully constructed rules around information privacy that retain robust and appropriate preservation capabilities; clear legislation on retention and permanency, an appraisal of perceived value with resources to undertake preservation actions

Bit List History

Added to list: 2011

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

In 2019, this entry became a subset of an entry introduced in 2017 for 'Digital Legal Records and Evidence,' which was split into four more discrete entries. This category includes evidence from public enquiries and commissions that have been presented in court. It recognizes that courts are not limited in the types of evidence that they can admit but that they have a responsibility to provide robust preservation that ensures the authenticity of their records and evidence.

The 2021 Jury noted that there is considerable evidence of good practice emerging from some of the examples where clear archival responsibility has been the key to progress but not to the extent of changing the risk profile or 2021 trend for the entry.

The 2023 Council agreed with the Critically Endangered classification with overall risks remaining on the same basis as before ('No change' to trend), while also noting that personal testimonies are particularly vulnerable, thus highlighting the importance of independent archives and advocacy for preserving these records.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

National or state recordkeeping regimes and legislation are often clear on the retention or permanency of these types of records. A major issue, however, is embargoes. When an embargo is lifted, will the file format or database continue to work, or will it longer work, making the data useless?

Case files and correspondence are one thing. Retention of these should be clear but may differ widely between jurisdictions and levels of government. If retention is not long-term or permanent, the risk of loss may not be so critical. Retention of 'unused' or 'potential' evidence is likely a different matter altogether. It may not even be considered a record, and certainly is not a record of the court. Should it be returned to the suspect or accused? Are their rights being considered here - not just in terms of preservation, but also simply disposition? There are legal and ethical issues around this that need to be fleshed out in conjunction with assessing its preservation risk.

Personal testimonies provided to commissions of inquiry but are not used as part of its legal investigations are in a particularly vulnerable state. The recent Mother and Baby Homes commission in Ireland collected a number of personal testimonies from survivors, but these were not used to inform the commission's final report as they were not presented to the commission's legal hearing iteration. After the commission's report, it announced that it intended to destroy these testimonies, and refused to allow survivors to access transcripts or recordings of their own testimony. A public outcry halted this planned destruction, but access to their testimonies is still difficult for survivors. Independent archives can provide a crucial role in cases such as these; whether state commissions would be willing to work with independent bodies in these cases remains to be seen.

Case Studies & Examples

- The Hillsborough disaster and inquiry as a case example of issues arising when inquiry records get split up. See **The report of the Hillsborough Independent Panel**, Hillsborough Independent Panel , (2012), *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/229038/0581.pdf* [accessed at 2023-10-24].

- The Hillsborough disaster and inquiry as a case example of issues arising when inquiry records get split up. See **Hillsborough: The Archive as a Vehicle for 'Truth Recovery.**, Sarsfield, D., (2021), *https://microform.digital/boa/posts/category/articles/423/hillsborough-the-archive-as-a-vehicle-for-truth-recovery* [accessed at 2023-10-24].

- **Tunisia: Truth Commission Outlines Decades of Abuse**, The Tunisian Truth and Dignity Commission, (2019), *https://www.hrw.org/news/2019/04/05/tunisia-truth-commission-outlines-decades-abuse* [accessed at 2023-10-24].

- **The Saville Inquiry: Q&A**, BBC, (2010), *https://www.bbc.com/news/10147362* [accessed at 2023-10-24].

- **Executive Summary of the IVD Comprehensive and Final Report**, Truth and Dignity Commission, (2019), *http://www.ivd.tn/rapport/doc/TDC_executive_summary_report.pdf* [accessed at 2023-10-24].

Digital Evidence and Records of Investigation Prior to Court

Digital materials assessed by police and other authorities in the course of investigation and retained as evidence of due process such as case files and correspondence, including materials not submitted to court.

Categories

- Legal Data

Examples

CCTV; Email; 3d scanning; social media interactions; police records; court records; text messages

Imminence	**Effort**
●●●●○	●●●○○
Action is recommended within twelve months. Detailed assessment is a priority.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Poor chain of custody; fragile or obsolete media; dependence on proprietary formats or products; lack or loss of documentation; inaccessible to web harvesting technologies; lack of version control; lack of integrity checks or integrity records; poor chain of custody; Uncertainty over IPR or the presence of orphaned works

Mitigations

Meticulous transfer and disclosure processes

Bit List History

Added to list: 2014

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 as an entry made in 2017 for 'Digital Legal Records and Evidence,' which the Jury split into four more discrete entries. This category includes evidence prior to court that may form part of an investigation or gathering of evidence but which are not formally submitted as evidence. It recognizes that police and other investigating authorities are not limited in the types of evidence that they need to administer, but that this creates an almost unbounded limit of preservation requirements to ensure authenticity and admissibility. A 2021 risk was identified based on examples bringing to question whether legal bodies have the skills and capabilities to preserve these materials should they need them if a case is reopened etc. The 2022 Taskforce found no significant trend towards greater or reduced risk.

The 2023 Council agreed with the Critically Endangered classification with the overall risks remaining on the same basis as before ('No change' to trend).

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

In the International organizations realm, more and more of these investigative missions are being set up. They are collecting huge volumes of data and the same issues around chain of custody, integrity records/checks continue to be aggravating especially with respect to authenticity and admissibility. Given the potentially huge volumes of data, and the drive to keep costs low, it is debatable whether there will be sustained funding for preservation.

Case files and correspondence are one thing: retention of these should be clear but may differ widely between jurisdictions and levels of government. If retention is not long-term or permanent, the risk of loss may not be so critical. Retention of 'unused' or 'potential' evidence is likely a different matter altogether. Is it even a record? Certainly, it is not a record of the court. Should it be returned to the suspect or accused? Are their rights being considered here - not just in terms of preservation, but also simply disposition? There may be legal and ethical issues around this that need to be fleshed out in conjunction with assessing its preservation risk.

Police forces tend only to have the resources to maintain forensic capability with relatively recent technology - for older technology, institutions and specialist companies are the only sources of expertise. This has an impact on cold cases.

There have been many examples of convictions being overturned when previously unused evidence was brought to light. Therefore the retention and preservation of unused evidence can have immense value.

Evidence in Court

Digital materials presented in court as evidence or documents such as rulings and proceedings generated through legal proceedings

Categories

- Legal Data

Examples

Evidence submitted to courts of all kinds, including text messages, photography, CCTV, email, 3d and 2d scanning, scientific reports and analyses, documents and websites

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Loss of context; loss of integrity; external dependencies; poor storage; lack of understanding; churn of staff; significant or diversity of data; poorly developed specifications; ill-informed records management; poorly developed transfer protocols; poorly developed migration or normalization; longstanding protocols or procedures that apply unsuitable paper processes to digital materials; Uncertainty over IPR or the presence of orphaned works

Mitigations

Well-managed data infrastructure; preservation enabled at ingest; carefully managed authenticity; use of persistent identifiers; finding aids; well-managed records management processes; recognition of preservation requirements at highest levels; strategic investment in digital preservation; preservation roadmap; participation in digital preservation community

Bit List History

Added to list: 2013

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry is a subset of an entry made in 2019 titled 'Proceedings and Evidence in Court,' which was itself created as a subset of entry in 2017 for 'Digital Legal Records and Evidence,' The 2021 Jury split 'Proceedings and Evidence in Court' into two more discrete entries to highlight their distinct preservation challenges and risk profiles. This entry includes evidence that has been presented as evidence in court. It was given a Critically Endangered classification to highlight its higher risk profile and additionally emphasize that courts are not limited in the types of evidence that they can admit but that they have a responsibility to provide robust preservation that ensures the authenticity of their records and evidence. The 2022 Taskforce found no significant trend towards greater or reduced risk

The 2023 Council agreed with the Critically Endangered classification with the overall risks remaining on the same basis as before ('No change' to trend). They emphasized the importance that organizations with these materials should have identified preservation actions established in their workplan–for digital evidence of investigation prior to court–to put into practice within the next three years.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

Temporary courts are continuing to gradually close and decisions about preservation and management of their archives are being made hurriedly and at the last minute. Some of the decisions are placing materials at high risk due to; materials being split all over the place - including to entities with no capacity or capability to preserve them, a seeming lack of understanding that preservation and management of the archives has no completion date, an unwillingness to invest in preservation or a drive to keep costs low which is resulting in negative implications for preservation, hurried choices on preservation measures which are not allowing for proper testing of approaches to safeguard authenticity and legal admissibility (e.g. extracting digital data from complex systems in formats that can then potentially not be restored).

Standard Records Management processes within designated agencies should be able to take care of the preservation of materials like this but given that it is likely to involve complex types of data, such agencies may not be equipped to deliver preservation effectively. It is surprising that courts are not more prominent in the digital preservation community, where solutions now exist.

Case Studies & Examples

- For example, the Special Tribunal of Lebanon 14th Annual Report (2022-2023) touches on the above comments concerning the planning and approaches developed and agreed between the United Nations and the Government of Lebanon to guide the Special Tribunal to ensure the completion of the Tribunal's residual functions, including the management and preservation of the records and archives of the Special Tribunal. See **Special Tribunal for Lebanon 14th Annual Report (2022 - 2023)**, Special Tribunal for Lebanon, (2023), *https://www.stl-tsl.org/sites/default/files/documents/annual-reports/STL_Annual_Report_2022-2023.pdf* [accessed at 2023-10-24].

- **ICA - Key Insights from the UN Side Event.**, Alméras, C., (2024), *http://www.ica.org/key-insights-from-the-un-side-event-on-preserving-and-accessing-records-of-temporary-international-criminal-courts/* [accessed at 2024-10-07].

Proceedings in Court

Digital materials generated through legal proceedings in court.

Categories

- Legal Data

Examples

Digital records of proceedings; digital records of rulings, and all manner of quasi-judicial proceedings and tribunals

Imminence	Effort
●●●○○	●●●○○

Imminence	Effort
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Loss of context; loss of integrity; external dependencies; poor storage; lack of understanding; churn of staff; significant or diversity of data; poorly developed specifications; ill-informed records management; poorly developed transfer protocols; poorly developed migration or normalization; longstanding protocols or procedures that apply unsuitable paper processes to digital materials; Uncertainty over IPR or the presence of orphaned works.

Mitigations

Well managed data infrastructure; preservation enabled at ingest; carefully managed authenticity; use of persistent identifiers; finding aids; well managed records management processes; recognition of preservation requirements at highest levels; strategic investment in digital preservation; preservation roadmap; participation in the digital preservation community.

Bit List History

Added to list: 2017

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry is a subset of a previous 2019 entry, 'Proceedings and Evidence in Court,' which was itself created as a subset of entry in 2017 for 'Digital Legal Records and Evidence.' The 2021 Jury split 'Proceedings and Evidence in Court' into two more discrete entries to highlight their distinct preservation challenges and risk profiles. This entry includes court proceedings and recognizes that courts have a responsibility to provide robust preservation that ensures the authenticity of these records. The 2022 Taskforce noted no change to the trend.

The 2023 Council agreed with the Endangered classification with the overall risks remaining on the same basis as before ('No change' to trend).

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

Recordings of proceedings in court may include the AV recording of the court session, which may pose particular preservation risks associated with the video files.

Temporary courts are continuing to gradually close and decisions about preservation and management of their archives are being made hurriedly and at the last minute. Some of the decisions are placing materials at high risk due to; materials being split all over the place.

including to entities with no capacity or capability to preserve them, a seeming lack of understanding that preservation and management of the archives has no completion date, an unwillingness to invest in preservation or a drive to keep costs low which is resulting in negative implications for preservation, hurried choices on preservation measures which are not allowing for proper testing of approaches to safeguard authenticity and legal admissibility (e.g. extracting digital data from complex systems in formats that can then potentially not be restored.

Standard Records Management processes within designated agencies should be able to take care of the preservation of materials like this but given that it is likely to involve complex types of data, such agencies may not be equipped to deliver preservation effectively. It is surprising that courts are not more prominent in the digital preservation community, where solutions now exist.

Case Studies & Examples

- The Special Tribunal of Lebanon 14th Annual Report (2022-2023), which touches on the above comments concerning the planning and approaches developed and agreed between the United Nations and the Government of Lebanon to guide the Special Tribunal to ensure the completion of the Tribunal's residual functions, including the management and preservation of the records and archives of the Special Tribunal. See **Special Tribunal for Lebanon 14th Annual Report (2022 - 2023)**, Special Tribunal for Lebanon, (2023), *https://www.stl-tsl.org/sites/default/files/documents/annual-reports/STL_Annual_Report_2022-2023.pdf* [accessed at 2023-10-24].

- **ICA - Key Insights from the UN Side Event**, Alméras, C., (2024), *http://www.ica.org/key-insights-from-the-un-side-event-on-preserving-and-accessing-records-of-temporary-international-criminal-courts/* [accessed at 2024-10-07].

Orphaned Works

Digital materials where copyright is uncertain, disputed or unknowable, meaning that preservation actions are constrained or prevented.

Categories

- Orphaned Works

Examples

Photographs, music recordings, literature.

Imminence	Effort
●●●●○	●●●●●
Action is recommended within twelve months. Detailed assessment is a priority.	Loss seems inevitable. Loss has already occurred or is expected to occur before tools or techniques develop.

Hazards

Lack of understanding of copyright; lack of documentation; dependencies resulting from hardware, software or media; lack of use resulting in lack of priority; lack of strategic investment in digital preservation; workflows that inhibit preservation of content that has not been licensed; encryption; poor storage.

Mitigations

Preservation pathway enabled; proven preservation plan applied; active effort to resolve IPR issues; institutional willingness to take risks for preservation.

Bit List History

Added to list: 2017

Trends

- **2024:** No change.

- **2023:** Towards greater risk.

Last Review

When this entry was added in 2017, there was little evidence of any renewed effort to address the issue of orphaned work. While there have been improvements to the baseline competence of the archival and library professions in their understanding of copyright and the skills to preserve contents, this alone provides a narrow basis for optimism and the scale of the challenge is likely to have grown just as quickly if not more so as aggravating conditions become more prevalent too. The 2021 Jury added that while content is preservable, the preservation of orphan works is a matter of process and risk appetite. Added to the complexity are changes to copyright legislation in and across different national and regional contexts, particularly for UK institutions post-Brexit, as noted in the additional comments below. For this reason, there was a 2021 trend towards greater risk.

The 2023 Council noted that this entry was trending towards greater risk as the longer orphaned works are left without preservation attention, the more vulnerable they become. The complexity of copyright legislation as well as works that cross over several domains/institutions adds to this increased risk. They additionally recommended that, given Orphaned Works is both a category and an entry, that the next major review for the Bit List includes a rescoping and splitting of this entry to cover the broad spectrum of Orphaned Works and their various preservation issues and efforts

The 2024 Council agreed with a 2023 recommended rescoping of Orphaned Works as both a category and an entry. Intellectual property rights impact every category, so it is better to embed considerations of it throughout relevant entries than to keep it as a single entry and standalone

category. These considerations have been incorporated into key elements and examples to relevant entries throughout the current Bit List, and the Orphan Works category and entry will be removed from the next 2025 Bit List.

Additional Information

A risk-based approach is recommended for organizations. Copyright infringements are only likely to become a significant issue in the context of access, and in most cases, the likelihood of any specific action is small. Preservation needs to be presented as a social good, one without which copyright holders would simply be unable to benefit from the property rights they seek to protect.

For UK institutions, the 2021 Jury recommended commentary by Naomi Korn on the status of orphan works and the impact of Brexit – that UK institutions are no longer able to make use of the EU Orphan Works Directive and the alternative Orphan Works Licensing Scheme is costly. A list of resources is available at *https://naomikorn.com/resources/*. For those in the UK, there is also the UK Copyright and Creative Economy Centre (CREATe) for resources on orphan works and copyright more broadly at *https://www.create.ac.uk/resources/*.

Case Studies & Examples

- The National Disc of the BBC Domesday Project offers an example of loss outside of the legal deposit mandate where the copyright owner cannot be traced. See **Domesday** , Finnay, A., *BBC https://www.atsf.co.uk/dottext/domesday.html* [accessed at 2023-10-24].

- Archive of Our Own is a non-commercial and nonprofit central hosting place for fanworks using open-source archiving software. As part of their system, there is a way for users to orphan works to keep the work active but delete any identifying data and transfer the work to a special account. See **Orphaning**, *Archive of Our Own https://archiveofourown.org/faq/orphaning?language_id=en* [accessed at 2023-10-24].

---

c)   Organizational Change

Always Online Games

Video games that are required to be continuously online. Gameplay is referenced here particularly as means of participation, along with social media and in-game interaction between players. This can include Massively Multiplayer Online games and single player games with always-on DRM.

Categories

- Gaming

Examples

Fortnite, World of Warcraft, Neverwinter, League of Legends

Imminence	Effort
●●●○○	●●●●○
Action is recommended within three years, detailed assessment within one year.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Lack of skills, commitment or policy from corporate owners; uncertainty over IPR or the presence of orphaned works; controversies around IPR; lack of offline backup; changing business model of providers; limited recognition of value of game play; limited recognition of value of game preservation; over dependence on goodwill of ad-hoc community; lack of preservation know-how at service providers; dependency on bespoke hardware or interfaces; increased reliance on always-on DRM for single player games

Mitigations

Well documented code; IPR supportive of preservation; large and committed user community; removal of always-on DRM for single player games

Bit List History

Added to list: 2023

Trends

- **2024:** Towards greater risk.

- **2023:** Towards greater risk.

Last Review

2023 Review

This entry was added in 2019 as a subset of the 2017 entry for 'gaming'. The 2020 and 2021 Juries noted a trend towards greater risk, due to the increased significance of these games during the COVID Pandemic as well as the evolving nature of MMOs, to the extent that the 2021 Jury changed the risk classification from Endangered to Critically Endangered.

The 2023 Council agreed with the 2022 Taskforce suggestion to consider the naming and scope of the entry, rescoping this entry to 'Always Online Games' covering all games that have to always be online, whether that is due to being MMOs, server-based games or single-player games with Always-Online DRM. Games that have online components but are not required to always be online fit into the new 'Games with Online Play Components' entry.

2024 Interim Review

The 2024 Council identified a trend towards even greater risk based on shifts in business models and increased litigation over the last year, resulting in more shutdowns which impact preservation efforts. It also raises time sensitivity for action; if there are no efforts to preserve and those existing are further shutdown, this raises the likelihood of loss.

Additional Information

Preservation for Always Online games in a playable state requires preservation or re-creation of the servers that are used to run these games. Even then, for MMOs or multiplayer games, it would be impossible to recreate these games at their various peaks. This nicely encapsulates why video recordings of (online) gameplay are important. They will never have the same configuration of subscribers, to say nothing of the innumerable changes made to the software over the years, which have significantly altered how the game works and looks. Loss is inevitable, and it has already happened. The social and cultural aspects of play are incredibly important, and on-screen recording is the most robust way to capture that.

Whilst it is expected that MMOs and always multiplayer games (such as Fortnite) would always require an internet connection due to their reliance on servers, single player games, or those where the primary gameplay is single player, being always online due to DRM provides an added risk to preservation. If the server shuts down, then even the single player components might not be playable, thus loss happens faster than a single player game that does not have a reliance on servers. For more details, see the 'Shut Down or Discontinued Video Games' entry.

Case Studies & Examples

- **Ubisoft reportedly revoking The Crew from owners' libraries following server shutdown**, Wales, M. , (2022), *Eurogamer* https://www.dpconline.org/digipres/champion-digital-preservation/bit-list/critically-endangered/bitlist-always-online#:~:text=%22We%20announced%20on%20December%2014,server%20infrastructure%20and%20licensing%20constraints.%22 [accessed at 2024-09-05].

- PCGamingWiki has an automated list of games that has Always-online DRM as well as a list of games that had Always-online DRM that have shut down. See **List of games using Always Online DRM**, PCGamingWiki, https://www.pcgamingwiki.com/wiki/List_of_games_using_Always_Online_DRM [accessed at 2023-10-24].

- GOG is a digital distribution platform for video games and films that only distributes games that are DRM-free. See **UPDATE #2: OUR COMMITMENT TO DRM-FREE GAMING'**, GOG, (2022), https://www.gog.com/news/bgog_2022_update_2b_our_commitment_to_drmfree_gaming [accessed at 2023-10-24].

- Dr Megan Winget's ethnographic research project focused on supporting the collection and preservation of massively multiplayer online (MMO) games. See **Winget (Megan) Videogame Development Research Collection**, Dr. Winget, M., (2009), *The University of Texas at Austin* https://repositories.lib.utexas.edu/handle/2152/8465 [accessed at 2023-10-24].

- **International Video Game Preservation Survey Report**, Arneil, C. , (2024), *National Film and Sound Archive of Australia* https://www.nfsa.gov.au/collection/curated/international-video-games-preservation [accessed at 2024-10-22].

- **Embracing a wider screen culture**, BFI, https://blog.bfi.org.uk/long-read/our-ambitions/embracing-a-wider-screen-culture/ [accessed at 2023-10-24].

Games with Online Play Components

Video games with online components that have online elements for interactive aspects of play, in particular those with online multiplayer components. This does not exclude games that have both online multiplayer and offline single player components (e.g., Dark Souls has online multiplayer components but also has the ability for single player play in offline mode) but rather the focus on the entry is on the online components for interactive play and there is more potential to preserve the interaction. This excludes games that are no longer available legally.

Categories

- Gaming

Examples

Dark Souls, Minecraft, Portal 2, Stardew Valley, Super Mario Bros. Wonder.

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of skills, commitment or policy from corporate owners; always online DRM; uncertainty over IPR or the presence of orphaned works; lack of offline backup; changing business model of providers; limited recognition of value of game play; over dependence on goodwill of ad-hoc community; dependency on bespoke hardware or interfaces; Complex hardware dependencies or bespoke hardware

Mitigations

Emulation pathway; source code; trusted repository; large user community; IPR supportive of preservation; strong documentation

Bit List History

Added to list: 2023

Trends

No trend data

Last Review

2023 Review

This entry, alongside the Games with 'Offline Play Components' entry, was created from rescoping the previous 'Old or Non-current Video Games' entry as part of the 2023 Bit List review. It was rescoped to highlight the differences in preserving online components as opposed to offline components in video games, specifically the dependency of servers.

2024 Interim Review

The 2024 Council agreed These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

Whilst this entry focuses on games with online play components that are still accessible (any servers that have been shut down or discontinued would fall into the Shut Down or Discontinued Video Games entry), these games could potentially lose server support thus necessitating urgent action to preserve the online component whilst the server is still active. A recent example of this is the announcement in October 2023 of Nintendo discontinuing online services for Nintendo 3DS and Wii U software from early April 2024. Whilst the offline component will still be accessible, the game will have lost features and/or game modes that use online components thus changing the nature of the game.

Case Studies & Examples

- **Announcement of Discontinuation of Online Services for Nintendo 3DS and Wii U Software**, Nintendo, (2023), *https://www.nintendo.co.uk/Support/Nintendo-3DS-2DS/Connecting-to-the-Internet/Announcement-of-Discontinuation-of-Online-Services-for-Nintendo-3DS-and-Wii-U-Software-2455285.html* [accessed at 2023-10-24].

- The Play It Again: Preserving Australian Videogame History of the 1990s (Play It Again 2) project, which builds on the original Play It Again project focused on curating a collection of Australasian games of the 1980s. This second iteration of Play It Again was assembled with the aim of building a collection of significant 1990s Australian videogames for the Australian Centre for the Moving Image (ACMI), preserving these and evaluating relevant emulation platforms, including EaaSI to determine their efficacy for accessing born digital artefacts. See **Play It Again Project**, *https://playitagainproject.com/* [accessed at 2024-04-06].

- **International Video Game Preservation Survey Report**, Arneil, C. , (2024), *National Film and Sound Archive of Australia* *https://www.nfsa.gov.au/collection/curated/international-video-games-preservation* [accessed at 2024-10-22].

- **Australian cultural institutions unite to collect videogames**, ACMI , (2022), *https://www.acmi.net.au/about/media/media-releases/australian-cultural-institutions-unite-to-collect-videogames/* [accessed at 2023-10-24].

- The Videogame Heritage Society, led by the National Videogame Museum, founded in 2022 to bring together organizations and collectors working with videogames. It provides advocacy, expertise, and support in collecting, preserving and displaying video games. See **Videogame Heritage Society**, National Video Museum, *https://thenvm.org/about/vhs/* [accessed at 2023-10-24].

- The British Film Institute's "Embracing a wider screen culture" strategy notes the cultural significance of video games and states that they intend to embark on sector research,

engagement and knowledge exchange (including on the preservation of video games and digital media). See **Embracing a wider screen culture**, BFI, _https://blog.bfi.org.uk/long-read/our-ambitions/embracing-a-wider-screen-culture/_ [accessed at 2023-10-24].

Pre-WWW ViewData and TeleText Services where no Archival Agency has Captured and Retained the Signal

Pre-WWW television information services broadcast within the TV signal that allowed a degree of search and retrieval of up-to-date information, based on Teletext or ViewData technologies and variants.

Categories

- Sound and Vision

Examples

AerTel; Electra; MetroText; Antiope-based systems; Ceefax; TeleText

Imminence	Effort
●●●●●	●●●○○
Immediate action necessary. Where detected, should be stabilized and reported as a matter of urgency.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of understanding; structure of information silos; lack or loss of documentation; uncertainty over IPR or the presence of orphaned works; lack of funding or impetus

Mitigations

Captured within on-air broadcast recordings; active research and recovery programme

Bit List History

Added to list: 2017

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2017, with a number of important developments reported since, raising hope that collections can be recovered and reused under certain circumstances. In particular, the development and on live capture of broadcast output at the BBC and British Film Institute noted by 2018 and 2019 Bit List Juries. Embedded signals can be assembled to access the Teletext content, and methods to recover such signals have been demonstrated (where the appropriate broadcast archive exists). In 2019, a proposal to add teletext graphics characters to the Universal Coded Character Set was also accepted, making it easier to transmit and archive teletext and legacy computer graphics for archiving and preservation. Version 13.0 of the Unicode Standard, released in March 2020, included the addition of graphic characters that provide compatibility with early teletext broadcasting standards. For these reasons, the 2020 Jury identified a 2020 trend towards reduced risk. The 2021 Jury identified a 2021 trend toward more reduced risk. They noted that while not yet fully clear how such a signal could be made searchable or made available at scale, research had progressed with different threads of research and enthusiast and community-led initiatives being brought together in the development of preservation tools and techniques. In light of those these developments and active research and recovery efforts, the 2021 Jury additionally supported a change from the Practically Extinct to Critically Endangered classification. The 2022 Taskforce agreed on the classification with risks remaining on the same basis as before, with no change to the trend.

The 2023 Council noted there may be a lower significance of the digital material of this entry than the previous 2021 Jury but agreed with the Critically Endangered classification with overall risks remaining on the same basis as before ('No change' to trend).

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

From a cultural studies point of view, it is a huge loss: an important source of information about news and social mores of the time. Is there progress to report on this entry, meaning that elements of the problem have been resolved? Major national agencies have collections of off-air recorded television on videotape carriers, which are likely to contain the teletext data. Few, if any, have undertaken substantial extraction and preservation of the teletext in its own right. However, many have digitized videotape carriers to digital file formats, which are now under preservation. An active Teletext enthusiast community has developed and created programmatic solutions to extract the teletext from the video files and emulation of the teletext display. A next step could be a collaboration between the official agencies and the enthusiast community to develop a systematic programme to extract teletext from off-air recordings for both preservation and access via emulation.

Case Studies & Examples

- **The Teletext Archive**, *https://www.teletextarchive.com/* [accessed at 2023-10-24].

- **Teefax**, *http://teastop.plus.com:8080/* [accessed at 2023-10-24].

- **The Teletext Art Research Lab**, *http://teletextart.co.uk/* [accessed at 2023-10-24].

- **The Viewdata Frame Database**, *https://db.viewdata.org.uk/* [accessed at 2023-10-24].

- **Announcing the Unicode Standard**, Unicode, Inc., *http://blog.unicode.org/2020/03/announcing-unicode-standard-version-130.html* [accessed at 2023-10-24].

Games with Offline Play Components

This entry is focused on games that can be played offline, often those designed for single player play while in offline mode. This does not exclude games that can be played online or have online interactive components (e.g., Sims 3 can be played online or offline), but rather the focus is on the preservation of offline single player components over the online components.

Categories

- Gaming

Examples

Sims 3, Planescape: Torment, Hades, The Elder Scrolls V: Skyrim

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of skills, commitment or policy from corporate owners; complex hardware dependencies or bespoke hardware; dependence on obsolete, low usage operating systems with no emulation pathway; uncertainty over IPR or the presence of orphaned works; use of older magnetic media; loss of underlying code or gaming engine

Mitigations

Emulation pathway; source code; trusted repository; large user community; IPR supportive of preservation; strong documentation

Bit List History

Added to list: 2023

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry, alongside the 'Games with Online Play Components' entry, was created from rescoping the previous 'Old or Non-current Video Games' entry as part of the 2023 Bit List review. It was rescoped to highlight the differences in preserving offline components as opposed to online components in video games, specifically a lack of dependence on servers.

2024 Interim Review

The 2024 Council agreed these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

Whilst this has the same risk classification as the Games with Online Play Components entry, the risk could be considered to be slightly lesser due to the lack of reliance on servers as well as examples of games having their online services shut down but players still being able to access the offline game modes/features, such as the case with Nintendo discontinuing online services for Nintendo 3DS and Wii U software in early April 2024 where they explicitly state that "Players will still be able to use features and game modes that do not require online communication".

This entry is also interlinked with the entries covering games played on different hardware (Console games, PC games and smartphone games entries) as the risks can change based on this

Case Studies & Examples

- **Announcement of Discontinuation of Online Services for Nintendo 3DS and Wii U Software**, Nintendo, (2019), *https://www.nintendo.co.uk/Support/Nintendo-3DS-2DS/Connecting-to-the-Internet/Announcement-of-Discontinuation-of-Online-Services-for-Nintendo-3DS-and-Wii-U-Software-2455285.html* [accessed at 2023-10-24].

- The Play It Again: Preserving Australian Videogame History of the 1990s (Play It Again 2) project, which builds on the original Play It Again project focused on curating a collection of Australasian games of the 1980s. This second iteration of Play It Again was assembled with the aim of building a collection of significant 1990s Australian videogames for the Australian Centre for the Moving Image (ACMI), preserving these and evaluating relevant emulation platforms, including EaaSI to determine their efficacy for accessing born digital artefacts. See **Play It Again Project**, *https://playitagainproject.com/* [accessed at 2024-09-06].

- **International Video Game Preservation Survey Report**, Arneil, C., (2024), *National Film and Sound Archive of Australia* *https://www.nfsa.gov.au/collection/curated/international-video-games-preservation* [accessed at 2024-10-22].

- **Australian cultural institutions unite to collect videogames**, ACMI, (2022), *https://www.acmi.net.au/about/media/media-releases/australian-cultural-institutions-unite-to-collect-videogames/* [accessed at 2023-10-24].

- The Videogame Heritage Society, led by the National Videogame Museum, founded in 2022 to bring together organizations and collectors working with videogames. It provides advocacy, expertise, and support in collecting, preserving and displaying video games. See

Videogame Heritage Society, National Video Museum, (2020),
https://thenvm.org/about/vhs/ [accessed at 2023-10-24].

• The Video Game History Foundation is a 501(c)3 non-profit dedicated to preserving and
teaching the history of video games. See **Video Game History Foundation** ,
https://gamehistory.org/our-mission/ [accessed at 2023-10-24].

• The British Film Institute's "Embracing a wider screen culture" strategy notes the cultural
significance of video games and states that they intend to embark on sector research,
engagement and knowledge exchange (including on the preservation of video games and
digital media). See **Embracing a wider screen culture**, BFI, *https://blog.bfi.org.uk/long-read/our-ambitions/embracing-a-wider-screen-culture/* [accessed at 2023-10-24].

Custom Online Databases

Data collected, presented and disseminated in custom online databases
that is not stored elsewhere, particularly data at risk when it is locked in
the database because no export or harvest options are available.

Categories

• Databases

• Research Outputs

• Web

Examples

Custom databases created project websites for research, citizen science

Imminence	Effort
●●●○○	●●○○○
Action is recommended within three years, detailed assessment within one year.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Lack of export options; lack of system maintenance; expired domain; lack of export functionality; lack
of technical knowledge and skills; limited or dysfunctional data management planning; web capture
challenges that means unlikely to be picked up by automatic crawlers; uncertainty over IPR or the
presence of orphaned works

Mitigations

Backup and documentation; preservation capability in designated repository; use of open formats
and open source or other licencing that enables preservation; enabled export options; robust data
management planning; documented and managed professionally

Bit List History

Added to list: 2023

Trends

- **2024:** No change.

Last Review

2023 Review

This was a new Bit List entry nominated and approved by the 2023 Council to draw attention to the particular challenges of preservation for custom online databases. This entry focuses on distinct risks relating to online databases that cannot go through traditional web archiving tools. While there are challenges to preserving databases both off- and online, it was nominated in the context of projects which set up a custom online database to record, present, and disseminate collected data, but this data is not stored elsewhere (e.g. in a long-term digital archive) and often is locked in the database because no export or harvest options are available. Identified areas of risks for these online databases can include: the maintenance of the system after the end of a project when it is not ensured, and online databases disappear because of security issues or because the domain expires; not all data is open and, after the end of a project, no one is responsible for granting access; the data is not stored elsewhere (e.g. in some trusted repository); the data is locked in and cannot be exported in (e.g. CSV) for further re-use.

Additionally, the nomination of the entry also highlighted a gap in the Bit List for databases more broadly. The 2023 Council agreed a new higher-level Databases digital species group should be created to address this gap, inviting nominations for other database-related entries to be considered for the next major revision of the Bit List.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

The preservation is highly dependent on the software used but, no matter what, once the project has reached its end, it starts to become vulnerable.

Often, the online databases are of interest to a sub-discipline-specific group of people, e.g. archaeologists specialized on cuneiform tablets. But the material itself often is then invaluable for this group because of the great effort invested in compiling it.

Databases for citizen science also provide an example where the upload of information directly into it makes it distinctive.

Emulation can be used to preserve these databases. For example, Yale University is preserving databases, especially SQL databases for websites, using EAASI. There are technical challenges, but the databases can be preserved, and have found issues are often around access to data and workforce development of technical skills to undertake preservation actions. There is a risk, however, that some of the databases cannot be exposed to the web as they have no survival time and/or cannot make them available as they were intended to be used.

Case Studies & Examples

- **Academics Retire and Servers Die: Adventures in the Hosting and Storage of Digital Humanities Projects**, Cummings, J. , (2019), *Digital Humanities Quarterly* *http://digitalhumanities.org/dhq/vol/17/1/000669/000669.html* [accessed at 2023-10-24].

Recordings of Video Gameplay Uploaded to Online Platforms

Recordings of game playing and e-sports that show how games are experienced and played, especially multi-user online games and tournaments.

Categories

- Gaming
- Sound and Vision
- Social Media
- Web

Examples

Material uploaded to Amazon Twitch, game channels on YouTube and other playback services

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Uncertainty over IPR or the presence of orphaned works; lack of offline backup; changing business model of providers; limited recognition of the cultural and historic value of game play; over-dependence on goodwill subsidy of an ad-hoc community; lack of preservation know-how at service providers; dependency on bespoke hardware or interfaces; DMCA takedowns

Mitigations

Offline backup; managed intellectual property rights; players and audiences invested in data

Bit List History

Added to list: 2015

Trends

- **2024:** No change.

- **2023:** Towards greater risk.

Last Review

2023 Review

This entry was added in 2019 as a subset of an entry made in 2017 for 'Gaming,' which the Jury split into four more discrete entries. There are overlaps with Social Media entries, except this category specifically draws attention to gaming and e-sports and therefore is a subset of both. By including as a separate entry, the 2019 Jury encouraged greater consideration of the cultural and historic value that such recordings are likely to acquire as well as the technical and economic challenges to preservation. The content is not particularly distinctive in technical terms, but there are aggravating circumstances, namely an almost complete reliance on commercial third parties (Google/YouTube and Amazon/Twitch) for the infrastructure around video capture and hosting. As the majority of this material is experienced and hosted on user-driven and ephemeral platforms such as YouTube and Twitch, it is less 'collectable' than the actual games and is unlikely to exist in private or public collections. Involves platforms that the digital preservation sector does not have much experience working with, e.g., Twitch. For these reasons, the 2021 trend was towards greater risk.

In 2023, whilst the Endangered classification and overall risk remained the same, the Council noted an increased trend towards greater risk due to an increase of copyright strikes on content, especially Nintendo gameplay videos. The 2023 Council also noted this is tied to other entries under the gaming species, especially 'Shut Down and Discontinued Games', as gameplay videos might be all that remains of the original experience of some games as well as being used to document the moment a server closed. Additionally, there can be intersecting issues with entries under the social media species, such as 'Born Digital Photographs and Video shared via Social Media Platforms' when addressing broader issues surrounding the capture and sharing of video content via a social media platform.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

Copyright claims on video content by publishers such as Nintendo also complicate things. One example of major copyright claims affecting content creators has been Nintendo issuing copyright strikes on a Legend of Zelda: Breath of the Wild video due to its demonstration of a fan-made mod and, once the issue had been raised, more copyright strikes were issued on Legend of Zelda: Breath of the Wild videos that did not contain footage of the mod. In other cases, content creators have had to deal with copyright strikes being issued due to the in-game music.

The significance of loss here is high because recordings, including commentary, and onscreen interactions with other players, seem likely to be the best way of preserving the experience of playing certain games at certain times. We are familiar with the challenges of preserving video, but we need to think about how established approaches will work in the context of the aggravating circumstances outlined above. There is a degree of urgency associated with working out how (legally and technically) preserving the materials that they hold may be preserved.

Important for social context, and from a DP point of view videos should not be too hard, but if we are capturing the experience to inform digital preservation actions and intents, then do not these videos exist in places such as YouTube and wouldn't they be brought in as part of the 'documentary' evidence of DP actions taken on the game or sports that have come into the archive?

Case Studies & Examples

- **Posting copyrighted Ubisoft material**, Ubisoft, (2024), *https://www.ubisoft.com/en-gb/help/connectivity-and-performance/article/posting-copyrighted-ubisoft-material/000062649* [accessed at 2024-09-06].

- The Game Archivist YouTube channel is dedicated to recording full gameplay videos with no commentary to "keep their experience preserved". See **The Game Archivist** , *https://www.youtube.com/@thegamearchivist* [accessed at 2023-10-24].

- **Nintendo's Copyright Strikes Push Away Its Biggest Fans**, Bedingfield, W. , (2023), *WIRED https://www.wired.co.uk/article/nintendo* [accessed at 2023-10-24].

Contractual Documents and Related Records

Documents, correspondence and other records created in the course of contractual dealings between individuals and agencies, especially where the subjects are of long duration and may be subject to legal scrutiny at undefined points in the distant future.

Categories

- Legal Data

Examples

Contracts, receipts, correspondence, licence agreements, building consent, warranties, and any other document or record that represents a legally binding transaction or permission. Such records may be useful in the avoidance or resolution of disputes, whether in court or prior to proceedings. Includes Online Terms and Conditions for e-commerce or end-user agreements for services

Imminence	Effort
●●●○○	●●○○○
Action is recommended within three years, detailed assessment within one year.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Loss of context; loss of authenticity or integrity; external dependencies; poor storage; lack of understanding; churn of staff; poorly framed or over-zealous disposal; ill-informed records management; misplaced fears with respect to data protection, encryption; Uncertainty over IPR or the presence of orphaned works

Mitigations

Managed appraisal; well managed data governance and infrastructure; preservation and appropriate retention strategies enabled at the point of creation; carefully managed authenticity; use of persistent identifiers; finding aids; well managed records management processes; application of records management standards

Bit List History

Added to list: 2014

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 as a subset of an entry introduced in 2017 for 'Digital Legal Records and Evidence,' which was split into four more discrete entries. There is overlap with Pension Mortgage and Insurance Records entry, but this entry addresses digital records with value over the long term that may not be immediately obvious at the point of creation.

The likelihood of liquidation, mergers or acquisitions means that these records are trending towards greater risk. There was a 2020 trend towards greater risk in light of the Covid Pandemic causing profound dislocation across the economy, placing many companies and agencies at financial risk. There was no identified trend for 2021 or 2022. While there are some noticeable improvements in record-keeping legislation and regulations and guidance to this area, with a growing number of agencies handling and advising on what needs to be kept and for how long, both the 2021 Jury and 2022 Taskforce agreed there remained room for improving how legal agreements are managed through records management standards and processes.

The 2023 Council agreed with the Endangered classification with the overall risks remaining on the same basis as before ('No change' to trend).

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

Closer collaboration over the digital record lifecycle with recordkeeping organizations such as IRMS/ARA and digital preservation organizations would help to ensure best practice from (before) record creation to its long-term preservation and would help to identify any risks and bridge gaps 'from the cradle to the grave'. Joining forces and resources will enable the community to raise awareness of the impact of best practices on the organizational governance and related efficiencies.

Managed appraisal of these documents and records would also aid in lowering the risk of this entry.

Case Studies & Examples

- The work and outputs of the EDRMS Preservation Taskforce, such as the EDRMS Preservation Toolkit, may be helpful for guidance as many of these records will be held in EDRMS type systems. See **EDRMS Preservation Toolkit**, Digital Preservation Coalition, (2021), *https://www.dpconline.org/digipres/implement-digipres/edrms-preservation-toolkit* [accessed at 2023-10-24].

Shut Down or Discontinued Video Games

Video games where the servers have been shut down or where the game has been delisted across digital platforms and is no longer able to be legally purchased directly from the digital marketplace (loss has already happened). It includes older and non-current video games designed and played on platforms and devices that are no longer supported. This group also includes older editions of games that have been delisted and replaced by newer or remastered editions.

Categories

- Gaming

Examples

City of Heroes, The Matrix Online, Club Penguin, P.T., Sims 1, Metro 2033, Darkspore, Maplestory 2

Imminence	**Effort**
●●●●○	●●●●○
Action is recommended within twelve months. Detailed assessment is a priority.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

N/A

Mitigations

Lack of skills, commitment or policy from corporate owners; emulation pathway; access to source code; IPR supportive of preservation; support of grassroots efforts; support for preservation from game publishers/developers; removal of Always-Online DRM

Bit List History

Added to list: 2023

Trends

- **2024:** Towards greater risk.

- **2023:** Towards greater risk.

Last Review

2023 Review

This entry was added in 2017 under the Gaming species under the title 'Old or Non-current Offline Video Games' and was rescoped in 2023. The 2023 Council noted that the original description and scope of the entry conflated several issues and was unclear in its purpose, referring both to games that have experienced loss and older games that are still available. The 2023 change of risk classification from Critically Endangered to Practically Extinct reflects the adjusted scope as games that fall under this entry have already experienced loss, in terms of servers, the actual game and users. Efforts to keep these games 'alive' or in circulation are reliant on legally dubious measures such as private servers and key reselling. It was also decided to remove reference to the age of the video game in this entry given that there has been an increase in server-reliant games shutting down within a year or two of launch that are more at risk than older games that are still being sold.

2024 Interim Review

The 2024 Council identified a trend towards even greater risk based on shifts in business models and increased litigation over the last year, resulting in more shutdowns which impact preservation efforts. It also raises time sensitivity for action; if there are no efforts to preserve and those existing are further shutdown, this raises the likelihood of loss.

Additional Information

The key element of this entry is that loss has already happened. Whilst there are numerous hobbyists working on preserving individual games and servers, the critical mass of users has now been lost for these games. Additionally, whilst the work being done by hobbyists is often vital to the survival of these games, the legality of these projects are in question and are often prone to shutting down without warning. Organizations like the Videogame Heritage Society provide a space to share advice and guidance on preserving video games but the sheer breadth of shut down and discontinued games means that a collaboration between hobbyists, organizations and game developers is what is required to begin solving this issue.

It is also worth noting that Always-Online DRM is a key issue in this area as if the servers shut down where a game has this type of DRM, then even the singleplayer part of the game can no longer be played without DRM circumvention, which is not legal. Darkspore is a key example of this happening.

A unique example to point out here around shut down MMOs is Maplestory 2, which closed its global services in May 2020, less than two years after the game received a global release. Nexon, the game's publisher, released a number of assets including designs, concept art and illustrations on their website as well as uploaded all the music onto YouTube. These assets were released for non-commercial and personal use only.

Case Studies & Examples

- **Spec Ops: The Line Disappeared. It Won't be the Last Beloved Game to Vanish**, Farokhmanesh, M. , (2024), *WIRED* https://www.wired.com/story/spec-ops-the-line-disappeared-cultural-touchstone-gone/ [accessed at 2024-09-06].

- **Ubisoft reportedly revoking The Crew from owners' libraries following server shutdown**, Wales, M. , (2024), *Eurogamer* https://www.eurogamer.net/ubisoft-reportedly-revoking-the-crew-from-owners-libraries-following-server-shutdown# [accessed at 2024-09-06].

- Atlus sued the creators of a private Shin Megami Tensei Imagine MMO server (Rekuiemu). The case was voluntarily dismissed with prejudice of defendant in January 2023 pursuant to a settlement agreement where the server was taken offline. In response to the initial lawsuit, another private server ReIMAGINE also voluntarily closed down. See **Atlus sues makers behind private server of defunct Shin Megami Tensei MMO**, Partis, D. , (2022), *Games Industry.biz* https://www.gamesindustry.biz/atlus-sues-makers-behind-private-server-of-defunct-shin-megami-tensei-mmo [accessed at 2023-10-24].

- **Computer-Aided Design and Computer Graphics in 1980s Australia: Lyn Tune, Sally Pryor and Andrew Quinn**, Swalwell, M., (2023), _ RMIT Design Archives Journal_ https://issuu.com/rmitculture/docs/rda_journal_28_13.1_issu/s/21982513 [accessed at 2023-10-24].

- **Archiving Australian Media Arts: A cross-continental adventure**, Moye, C. , (2021), *Digital Preservation Coalition* https://www.dpconline.org/blog/bit-list-blog/blog-cynde-moya-wdpd.

- **Preserving Virtual Worlds**, Mcdonough, J., Olendorf, R., Kirschenbaum, M. and Kraus, K. , (2010), https://www.researchgate.net/publication/46301847_Preserving_Virtual_Worlds_Final_Report [accessed at 2023-10-24].

- **Darkspore Servers Shut Down**, Richard, M. , (2016), *TechRaptor* https://techraptor.net/gaming/news/darkspore-servers-shut-down [accessed at 2023-10-24].

- **MapleStory 2 Archive**, Nexon, https://web.archive.org/web/20220519185154/https://maplestory.nexon.net/ms2archive [accessed at 2023-10-24].

- **Announcement of Discontinuation of Online Services for Nintendo 3DS and Wii U Software**, Nintendo, (2023), https://www.nintendo.co.uk/Support/Nintendo-3DS-2DS/Connecting-to-the-Internet/Announcement-of-Discontinuation-of-Online-Services-for-Nintendo-3DS-and-Wii-U-Software-2455285.html [accessed at 2023-10-24].

- **Play It Again Project** , https://playitagainproject.com/ [accessed at 2024-09-06].

- **International Video Game Preservation Survey Report**, Arneil, C. , (2024), *National Film and Sound Archive of Australia* https://www.nfsa.gov.au/collection/curated/international-video-games-preservation [accessed at 2024-09-06].

- **Survey of the Video Game Reissue Market in the United States**, Salvador, P. , (2023), https://doi.org/10.5281/zenodo.8161056.

- The 'MMO Timeline' in Bio Break, a blog devoted to MMORPGs, RPGs, and other games, offers a quick reference sheet for the notable MMOs and proto-MMOs from 1980 to present (with closures noted in red). See **MMO Timeline**, Syp, https://biobreak.wordpress.com/mmo-timeline/ [accessed at 2023-10-24].

- Delisted Games, which is a website dedicated to tracking delisted games and server shutdowns. See **Delisted Games**, *https://delistedgames.com/about/* [accessed at 2023-10-25].

- **Videogame Heritage Society**, National Video Museum , *https://thenvm.org/about/vhs/* [accessed at 2023-10-25].

- **Video Game History Foundation**, *https://gamehistory.org/our-mission/* [accessed at 2023-10-24].

- **Embracing a wider screen culture**, BFI, *https://blog.bfi.org.uk/long-read/our-ambitions/embracing-a-wider-screen-culture/* [accessed at 2023-10-24].

Legacy Interfaces and Services Offered Online by Major Companies

Online services with unique interfaces that change regularly and through those changes provide a different experience and different content to their users.

Categories

- Social Media

Examples

Interfaces to user-driven social platforms such as MySpace, Bebo, Facebook, Twitter/X, Reddit, LinkedIn, Sina Weibo, Flickr, and social platform API-driven services such as Botometer and Tweetdeck and many others, and superseded policies and terms governing access and re-use of platform data

Imminence	Effort
●●●●●	●●●●●
Immediate action necessary. Where detected, should be stabilized and reported as a matter of urgency.	Loss seems inevitable. Loss has already occurred or is expected to occur before tools or techniques develop.

Hazards

N/A

Mitigations

Robust and extensive web archives with strong documentation of platform documentation (such as terms of service) over time; search algorithms (where available); ranking and personalization of interfaces; awareness of IPR and the presence of orphaned works

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 to highlight the configuration of interfaces and, therefore, the ever-changing arrangement and presentation of content. Personalization means that the same query can produce quite different results to different users at the same time; the application of machine learning to behavioural surplus means the same may obtain different results at different points in time. That is over and above the rapid churn in the appearance of web interfaces. There is little appreciation of the implications for the use of online services and the potential for manipulations that arise. Moreover, the digital preservation community, which is historically concerned with data rather than interface, has only rudimentary tools to address this challenge. The 2021 Jury agreed but noted a trend towards greater risk due to security issues posed by hosting legacy technology software and services which have prompted disposal of content imminently without adequate review or selection. The 2022 Taskforce agreed these risks remain on the same basis as before (no change to the trend).

The 2023 Council agreed with the Practically Extinct classification and noted an increase in imminence, recognizing that while the need for major efforts to prevent or reduce losses continues, it is now much more likely that loss of material has already occurred, and will continue to do so, by the time tools or techniques have been developed.

2024 Interim Review

The 2024 Council agreed these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

In light of the recent changes to API services by major platforms (e.g. Twitter/X, Reddit), they recommend maintaining the entry separate from 'Consumer Social Media Free at the Point of Use' to highlight the risk of loss of access to underlying data. Since 2023, a number of major global platforms (such as Twitter/X and Reddit) have begun restricting researcher access to platform APIs, making API services cost-prohibitive to academic and collecting institutions, or shutting them down completely (Davidson et al., 2023, Summers, 2023). With restrictions to this method of collection, data will be (and likely has already been) lost to the institutions responsible for maintaining access to research data and heritage collections.

While this entry focuses on interfaces and the risk of losing the display and functionality characteristics crucial to interpreting social media content, it also encompasses the multi-modal access and terms under which social media platforms make content available. While the algorithms that serve content to a user's feed or browser-based interpretations of interfaces make a major impact on how content might be interpreted, the terms and conditions, data services, and policy frameworks around platforms also provide crucial context.

2024 has seen increased awareness of social media as a manipulating factor in social life, such as congressional hearings in the US around social media impact on youth, pop culture books such as Cathy O'Neill's publication The Shame Machine: Who Profits in the New Age of Humiliation (which includes social media amongst a host of other shame-creating industries), and a proposed warning label for teens on social media platforms from the surgeon general. New EU legislation, such as the Digital Services Act, also demonstrates wider awareness of the impact of social media and the need to hold platforms accountable through reliable records of content published by users as well as policies and service documentation.

Though Twitter/X and Reddit, for example, still exist and user data has not disappeared at a platform level, access to that data in a way that supports computation access or archiving as scale has been shut down. Without access to data via API, research and analysis methods are severely hampered. Given the fast-paced, ephemeral nature of social media, this means research potential has already been lost and tools crucial for using social media data to support public policy and services have become obsolete.

Additionally, the Council recommends further research as part of the next 2025 review, in particular, to highlight the particular risk of loss of legacy platform policies and T&Cs, which are required to interpret datasets or web crawls of social media at a certain point in time. This issue of lack of access to former policies has been raised in academic forums and this should be addressed more explicitly. They propose describing this issue and recommending 1) further research into existing archives of platform information and policy pages (such as in IA collections and LoC 'Business in America' collections) and 2) adding this issue as 'Aggravating Circumstances' for other entries or potentially creating a new entry (but preferably not).

Additional Information

The interfaces alone have less impact if they are gone (since many research users are interested in the extracted data). However, it's an important distinction for us to make that we could end up preserving social media data as datasets in the long run, meaning that the look and feel (which serves a different sort of purpose) will be lost. A number of social media researchers in groups like the Association of Internet Researchers (AoIR) may care more about the data, but perhaps it is worth exploring a bit more about their interest in interfaces.

Without the interfaces and the underlying software that enables social media platforms, it will be impossible to preserve the look and feel and even meaning of a large portion of content that depends on particular functionality or interface to be accurately or authentically interpreted, including for evidential uses, artworks, design research, and historical/qualitative research. The loss of these interfaces (or lack of any indication of robust documentation by platforms) means a significant gap in the cultural heritage of many communities and even entire nations. For example, some content creators on YouTube may lose access to their content and accounts due to copyright infringement claims or reports of inappropriate content, which may or may not be supportable. The risk of loss is higher if the content is not stored anywhere else. Though some mitigation methods are available through the platform, this issue may only affect a small number of accounts.

Some of the content/iterations of these are likely preserved to an extent within existing web archives but not as targeted collection efforts. As we've seen with MySpace and other platforms where the platform producers decide to remove content or shut down with little notice, loss may be sudden.

Replaying social media content from 2014 through modern interfaces poses a challenge to authenticity and reliability. If no recording or documentation for legacy interfaces has been preserved, it will not be possible to recreate older interfaces. You'd think the platform owners may preserve older versions, but these copies (if they exist) are not available to research or collecting institutions, and it would be worth engaging in a conversation about making them available for scholarly projects and cultural heritage.

Some of this information is almost certainly lost already (some through deliberate disposal). The imminence of action depends on the type of institution.

Case Studies & Examples

- In the 2023 NDSA Web Archiving Survey Report, one of the major takeaways related to respondents' concerns about ability to collect social media—in particular, Twitter, Instagram, Facebook, and Reddit. Content housed within social networks has always been difficult to capture for a myriad of reasons and recent changes to numerous social platforms have made this task harder. See **Web Archiving Survey Results: An NDSA Report**, National Digital Stewardship Alliance (NDSA), (2023), *https://doi.org/10.17605/OSF.IO/N5MYR* [accessed at 2024-09-11].

- **Platform-controlled social media APIs threaten open science**, Davidson, B.I., Wischerath, D., Racek, D. et al., (2021), *https://doi.org/10.1038/s41562-023-01750-2*.

- **Looking Forwards**, Summers, E. , (2023), *https://web.archive.org/web/20230720201537/https* [accessed at 2024-09-04].

- **The Shame Machine: Who Profits in the New Age of Humiliation**, O'Neil, C. (2022), (2022), <>.

- **The Digital Services Act**, European Union, (2024), *https://web.archive.org/web/20240611053451/https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en* [accessed at 2024-09-04].

- **Preview: Senate Judiciary Committee to Press Big Tech CEOs on Failures to Protect Kids Online During Landmark Hearing Today**, US Senate Committee on the Judiciary, (2024), *https://web.archive.org/web/20240607212543/https://www.judiciary.senate.gov/press/releases/preview-senate-judiciary-committee-to-press-big-tech-ceos-on-failures-to-protect-kids-online-during-landmark-hearing-today* [accessed at 2024-09-04].

- **Social Media and the Teen Mental Health Crisis'**, US Senate Committee on the Judiciary, (2023), *https://web.archive.org/web/20240406092903/https://www.judiciary.senate.gov/committee-activity/hearings/social-media-and-the-teen-mental-health-crisis* [accessed at 2024-09-04].

- **Surgeon General Calls for Warning Labels on Social Media Platforms**, Barry, E. and Kang, C. , (2024), *New York Times https://web.archive.org/web/20240620034730/https://www.nytimes.com/2024/06/17/health/surgeon-general-social-media-warning-label.html* [accessed at 2024-09-04].

Pre-WWW Videotex Data Services and Bulletin Board Services

Pre WWW telephone and television information services that allowed a degree of user interaction and data retrieval with modem-based two-way communication.

Categories

- Sound and Vision

Examples

Prestel, Minitel, VidiTel and Videotex NL, Alex, BelTel, FidoNet

Imminence	Effort
●●●●●	●●●●●
Immediate action necessary. Where detected, should be stabilized and reported as a matter of urgency.	Loss seems inevitable. Loss has already occurred or is expected to occur before tools or techniques develop.

Hazards

N/A

Mitigations

Offline backup documented and available for recovery

Bit List History

Added to list: 2012

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2017 and, since then, has stayed Practically Extinct. The 2019 Jury noted that while there may be examples residing in offline backups of services taken at the time, these are likely to have deteriorated rapidly; Therefore, they called on anyone with such collections to act quickly to stabilize and recover content. The 2020 and 2021 Juries, and 2022 Taskforce, also agreed with the classification with risks remaining on the same basis as before (no change to the trend).

The 2023 Council likewise agreed, keeping the classification of Practically Extinct with overall risks on the same basis as before ('No change' to trend). They, however, did incorporate some changes to the entry, agreeing that there should be a lower significance of loss but also increased imminence of

action required and effort to preserve. The need for major efforts to prevent or reduce losses continues, but it is now much more likely that loss of material has already occurred and will continue to do so by the time tools or techniques have developed. For these reasons, there is a greater urgency to prioritize the assessment of these materials and develop tools or techniques to prevent or reduce further losses in this group.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

The value of this type of content has been questioned in previous entry reviews. It was noted that loss of this content would impact our understanding of pre WWW digital/electronic communications and whilst the loss in terms of impact on human life would be minimal, from a cultural studies point of view, loss would have a far higher impact.

In terms of preservation, whilst there is no structured collection of this material, many individuals will have their own personal archives and a campaign of the nature of Missing Believed Wiped from the BFI might be effective in collating these disparate collections.

This entry can also be linked to the Community Archives species as early online forums were a place of community development and community creation.

Pension, Mortgage and Insurance Records

Records of transactions for long-lived financial products and services contracted between individuals and corporations. These records typically contain or depend on significant amounts of personal information and outlast the infrastructure on which they were created.

Categories

• Sensitive Data

Examples

Applications, correspondence and ancillary records relating to pensions, mortgages and insurances and other contracts of long duration. This includes corporate databases, email, web archives and EDRMS, and may require some coordination of paper, microfiche, born-digital and digitized records. These records often include the scope and duration of the contract as well as any agreed changes during the lifetime of the product. It may also include evidence of mis-selling or other sharp practice, which only becomes apparent after the fact. This entry pertains to corporate records rather than personal records

Imminence	Effort
●●●○○	●●○○○

Imminence	Effort
Action is recommended within three years, detailed assessment within one year.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Lack of corporate preservation planning; lack of preservation within the procurement of corporate systems; companies conflating backup with preservation; loss of integrity and authenticity; loss of context and connections to provide meaning; lack of preservation capability within agencies; lack of preservation voice at executive level; poor planning and roadmap for corporate infrastructure; proliferation of legacy systems; slapdash procurement or migration of new systems; mergers and acquisitions leading to confusion of corporate systems; lack of compliance, audit or accountability at operational levels; encryption; Uncertainty over IPR or the presence of orphaned works owners.

Mitigations

Backup and documentation; use of open formats and open source software; considered data management planning; licencing that enables preservation; preservation capability in designated repository; resilient to hacking; selection and appraisal in place; authenticity and integrity of records managed; resilient funding and recognition at executive level; technology watch; regular preservation audits; accreditation and participation in the professional preservation community.

Bit List History

Added to list: 2017

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2017 but was outside the competence of the judges to assess at that time. It was assessed in 2019 with additional expertise invited to the panel to support this assessment and reviewed again in 2020. The 2021 Jury agreed with the 2019 assessment and subsequent 2020 review, which classified these digital materials as Vulnerable with no trend towards greater or reduced risk.

The 2023 Council agreed with the Vulnerable classification with the overall risks remaining on the same basis as before ('No change' to trend).

2024 Interim Review

The 2024 Council recommends that a major rescoping of the Sensitive Data species is necessary, with plans to remove it as a species and incorporate key elements and examples to relevant entries for the next 2025 Bit List. This is because it is not clear how sensitive data works as a species, when

many of the other species mentioned could have sensitive data concerns, and the sensitivity of the data is more like an extra category of risk that potentially applies across any species.

Additional Information

The importance of retaining documentation in any kind of legal agreement offers this kind of material more protection than most but legal organizations may conflate backup with preservation and not always have consistent records management systems.

Case Studies & Examples

*   The work and outputs of the EDRMS Preservation Taskforce, such as the EDRMS Preservation Toolkit, may be helpful for guidance in this context. See **EDRMS Preservation Toolkit**, Digital Preservation Coalition, (2021), *https://www.dpconline.org/digipres/implement-digipres/edrms-preservation-toolkit* [accessed at 2023-10-24].

---

d)   Organisational Limits

Community-generated Content in Arts and Heritage

Digital materials produced and shared in and by ad-hoc community art and heritage projects, typically through digitization, where the creation of digital materials was a significant purpose of the initiative.

Categories

*   Community Archives

Examples

Locally organized programmes associated with public remembrance and celebration such as World War One centennial commemorations; Holocaust commemorations; City of Culture; Olympic Games; World Cup

Imminence	Effort
●●●●○	●●●●○
Action is recommended within twelve months. Detailed assessment is a priority.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Poor documentation; lack of replication; lack of continuity funding; lack of residual mechanism. dependence on a small number of volunteers, lack of preservation mandate; lack of preservation thinking at the outset; failure of digital legacy planning; conflation of backup with preservation;

conflation of access and preservation; inaccessible to web archiving; lack of knowledge or application of standards to ensure good quality preservation actions; lack of internet access; distrust of 'official' archives; uncertainty over IPR or the presence of orphaned works

Mitigations

Residual archive with residual funding able to receive and support collections; strict adherence to digitization guidelines; quality assurance; active user community; intellectual property managed to enable preservation

Bit List History

Added to list: 2012

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 as a subset of 'Community Archives and Community-Generated Content' which was split into two entries to provide greater specificity in recommendations. The 2020 Jury noted a trend towards greater risk based on how community art and heritage groups, which often rely on volunteer effort, had been unable to meet for extended periods in 2020. Moreover, the local galleries, museums and arts centres on which they depend had closed, in some cases for good. The 2021 Jury review agreed with this assessment with the overall risks remaining on the same basis as before ('No change' to 2021 trend). The 2021 Jury commented that risks have not increased as so much as the challenges have remained, specifically those of funding and sustainability. Community-generated materials are often produced and shared through funded projects and tell a similar story of loss through inaction, but the challenge is the same as before; sustainability with project-based funding. There should be greater responsibility of funding bodies to ensure that digital preservation is built into any funded community preservation project plan and outcome. Digital content in this context is often seen as a by-product of engagement and is annoyingly left to disappear by organizations that do not have digital preservation experience or infrastructure. With good governance and sustainable digital repository support, this should not be an issue.

The 2023 Council agreed with the Critically Endangered classification, with the overall risks remaining on the same basis as before ('No change' to the 2023 trend). However, they noted there was a higher significance of loss as well as a higher inevitability of loss, and from these also a greater imminence of action needed to assess, prioritize, and develop tools and techniques for reducing future loss of materials.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

Local archives address these collections on an ad hoc basis. Loss seems likely due to the precarity of the funding streams, or lack thereof, for these projects. Once digitization has been carried out, many projects do not know what to do with them or have the means to make them accessible, and a lack of understanding of copyright is a barrier to sharing as well. Funding paths that enable digitization may not include planning for long-term storage and access. This often tends to be the case in local archives, where a higher imminence of action is critical to assess and address these issues before content is lost.

Communities who live in rural and remote areas may have a lack of access to services such as broadband connectivity, which is a well-reported issue and is often referred to as the "digital divide". Inadequate internet connectivity would diminish the capacity for these communities to access digital preservation solutions, such as cloud storage for digital assets. This is especially prevalent with personal photos and videos on mobile phones as possession of a mobile phone does not necessarily mean the user has adequate internet connectivity to be able to upload videos to web-based platforms.

There may also be a distrust of 'official' archives and government agencies due to the need for culturally appropriate handling of restricted/sensitive content. If the photographs, videos or audio depict culturally sensitive elements (e.g., sacred sites, ceremonies or secret Dreaming stories), the communities may want to uphold strict practices of restrictions (see also the 'First Nations Secret/Sacred Cultural Material' entry). There are instances where 'official' archives have changed their workflows and processes to accommodate the cultural wishes of particular First Nation communities, especially for secret/sacred content but these practices are not yet common-place. A detailed look into the preservation issues of secret/sacred content can be found in the New First Nations Secret/Sacred Cultural Material entry.

Case Studies & Examples

- The Alfred Landecker Foundation's Landecker Digital Memory Lab: Connective Holocaust Commemoration project with mission of exploring how Holocaust museums, memorial sites and archives can make better use of digital technologies to ensure the Holocaust continues to be commemorated long into the future: "Memorial sites, museums, and archives are increasingly working with digital media, yet there's a notable gap in digital literacy, capabilities, and use of synergies. To fill this gap, the Alfred Landecker Foundation is supporting the University of Sussex in the establishment of the Landecker Digital Memory Lab. The new lab will be led by Dr. Victoria Grace Walden and enhance interdisciplinary and inter-sector dialogue about digital Holocaust memory. The Lab's team will also develop a 'living database,' an accumulative collection of projects and experiences in the field, offering Holocaust memory and education institutions valuable insights from historical digital practices. The aim is to inspire not only productive academic thought but also new practical initiatives." See **Alfred Landecker Foundation's Landecker Digital Memory Lab: Connective Holocaust Commemoration project**, *https://www.alfredlandecker.org/en/projects/landecker-digital-memory-lab#* [accessed at 2024-07-07].

- The PARADISEC Pacific and Regional Archive for Digital Sources in Endangered Cultures project to digitize analog records of materials from endangered cultures from all over the world. This includes audio recordings and video recordings of performance, narrative, singing, and other oral tradition, amounting to over 207 terabytes and representing 1,370

languages, mainly from the Pacific region. See **Pacific and Regional Archive for Digital Sources in Endangered Cultures**, PARADISEC, *https://www.paradisec.org.au/* [accessed at 2023-10-24].

- The State Library New South Wales 'Reconnecting collections to communities' Mukurtu implementation, in which the Indigenous Engagement team works to make available the Library's Aboriginal Historical and Indigenous Languages collections to communities on Country and online. See **Reconnecting collections to communities**, State Library New South Wales, (2022), *https://www.sl.nsw.gov.au/about-library/services/indigenous-engagement/reconnecting-collections-communities* [accessed at 2023-10-24].

- The Aboriginal and Torres Strait Islander Data Archive (ATSIDA) is a specialized trusted research data management facility for Australian Indigenous research data and is managed by the University of Technology Sydney (UTS) Library. ATSIDA is a thematic archive within the Australia Data Archive (ADA) with its datasets stored securely at the Australian National University's National Computational Infrastructure (NCI). See **Aboriginal and Torres Strait Islander Data Archive (ATSIDA)**, *https://www.atsida.edu.au/* [accessed at 2023-10-24].

- A new preservation toolkit resource, 'Digital Preservation for Community Archives', was released for World Digital Preservation Day on 7 November 2024. The preservation toolkit has been created alongside the University of Glasgow as part of the Our Heritage, Our Stories (OHOS) project, and has been developed in collaboration with a range of community archives. It will include a range of practical guidance, support, and advice to those working with digital records in the community context. See **Digital Preservation for Community Archives**, (2021), *Digital Preservation Coalition* *http://doi.org/10.7207/communitytoolkit24-01*.

- **Aboriginal and Torres Strait Islander Cultural Protocols**, Sentance, N. and University of Sydney Library, (2018), *https://doi.org/10.25910/hrdq-9n85*.

- Digital inequality is recognized as one of the targets in the National Agreement on Closing the Gap, the objective of which is to "enable Aboriginal and Torres Strait Islander people and governments to work together to overcome the inequality experienced by Aboriginal and Torres Strait Islander people, and achieve life outcomes equal to all Australians.". Target 17 states that "By 2026, Aboriginal and Torres Strait Islander people have equal levels of digital inclusion". See **Closing the Gap Targets and Outcomes**, Closing the Gap, *https://www.closingthegap.gov.au/national-agreement/targets* [accessed at 2023-10-24].

- The "digital divide" in Australia has an impact on digital preservation of content generated by communities living in remote areas. See **Australia's digital divide means 2.8 million people remain 'highly excluded' from internet access**, Parke, E. , (2022), *ABC* *https://www.abc.net.au/news/2022-10-16/australia-digital-divide-millions-cannot-access-internet/101498042* [accessed at 2023-10-24].

- The "digital divide" in Australia has an impact on digital preservation of content generated by communities living in remote areas. See **Digital divide': report finds some Australian rural mobile data speeds 90% slower than urban**, Al Khawaldeh, K. , (2022), *The Guardian.* *https://www.theguardian.com/australia-news/2022/dec/13/digital-divide-report-finds-some-australian-rural-mobile-data-speeds-90-slower-than-urban* [accessed at 2023-10-24].

Consumer Social Media Free at the Point of Use

Social media platforms free at the point of use with a business model based on reselling user data for consumer behavior and/or advertising analysis, mainly for profit-driven corporations. This entry broadly includes digital content created, shared and hosted on social media platforms as well as current interfaces of social media platforms.

Categories

- Social Media

Examples

Instagram, Facebook, Twitter/X, Pinterest, Yahoo Groups, Truth Social, Reddit, Mumsnet, Sina Weibo, Flickr, Bebo, and legacy BBS

Imminence	Effort
●●●●●	●●●●○
Immediate action necessary. Where detected, should be stabilized and reported as a matter of urgency.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Lack of preservation capacity in provider; Lack of preservation commitment or incentive from provider; proprietary products or formats, including user interface; poor data protection; inaccessibility to web archiving; political or commercial interference; Lack of offline equivalent; super-abundance; Uncertainty over IPR or the presence of orphaned works; Lossy compression in upload scripts

Mitigations

Offline backup and documentation of media assets; Migration plan; Early warning from vendors; Roadmap from vendors; Accessible to web harvest; Suitable export functionality; Licencing enables preservation; Preservation commitment from vendor; Preservation capability in vendor; Resilient to hacking; Selection criteria

Bit List History

Added to list: 2017

Trends

- **2024:** Towards greater risk.

- **2023:** Towards greater risk.

Last Review

2023 Review

This entry was added by the 2019 Jury as a subset of a broader social media entry first introduced in 2017. It was created as a standalone entry to draw attention to the different threats faced by online services that are paid for versus 'free at the point of use' (both depend on the business model of the vendor and the terms they impose). The 2021 Jury raised the risk classification from Endangered to Critically Endangered based on concerns arising with **Trends** towards harmful and malicious hate speech as well as misinformation and deliberate deletion. The 2022 Taskforce agreed on a trend towards even greater risk based on the continued, significant trend towards hate speech, misinformation and disinformation, and deliberate deletion in light of ongoing global conflicts that include (but are not limited to) social and economic inequalities and climate change. In particular, they mentioned the sale of Twitter prompting a moment of instability in consumer social media, with the scale of Twitter, evident acrimony between parties prior to the sale and the hostile news coverage afterward, elevating the risks associated with social media. They also brought to attention issues surrounding platforms enabling extreme views not permitted on mainstream platforms, which emerged and proliferated noticeably and, from a preservation standpoint, could be argued are potentially at very high risk and historically significant.

Based on the assessment of the rescoped entry, the 2023 Council agreed on the Critically Endangered classification and noted an increase in the imminence of action required as well as the effort to preserve. The need for major efforts to prevent or reduce losses continues, but it is now much more likely that loss of material has already occurred and will continue to do so by the time tools or techniques have developed. There is a greater urgency to prioritize the assessment of these materials and develop tools or techniques to prevent or reduce further losses in this group.

The 2023 Council recommended further rescoping and adjusting of this and other social media entries in light of how web-based and cloud-based business products and services have developed in recent years. This included:

Clarifying the scope. This entry broadly refers to the preservation of content and interfaces of social media platforms, with these platforms designed to facilitate the creation and sharing of media through interactive social networks. These services, particularly those provided by largely unregulated (or underregulated) platforms, pose critical risks for not only capturing and preserving the content hosted on the social media platform but also the interfaces of the platforms themselves.

Similarly, the entry specifically refers to risks for digital materials created, shared, and hosted via social media services offered 'free at the point of use,' in which the business model and sustainability can only be guessed, and contracts tend to be asymmetrical in favor of the supplier. Moreover, because these services have a low barrier to entry, they may be favored by agencies or individuals least able to respond to closure or loss.

As part of this rescoping, relevant information concerning cloud-based aspects were incorporated into the 'Cloud-based Services and Communications Platforms' entry to more clearly differentiate the risks associated with cloud hosting and computing technologies, allowing this entry on consumer social media free at the point of use to focus on challenges, notably those relating to harvesting and managing content and interfaces of web-based social networking platforms.

The 2023 Bit List Council additionally recommended that the next major review for the Bit List include:

A restructure and splitting of the entry to create separate entries for 'digital content hosted on social media platforms' and for 'interfaces of social media platforms', where each can be teased out to

provide greater clarity about specific risks, aggravating factors and recommended actions. This should include expanding on API access to data, providing examples of legacy content already lost, and pointing to examples where risk is especially high (e.g., things that are still up but alarmingly fragile!)

A consideration of merging the 'Data Posted to Defunct or Little-used social media platforms' entry with this entry, to incorporate examples of loss in the presence of aggravating conditions.

A consideration of merging the 'Born Digital Photos and Video Shared on Social Media' entry with this entry, to provide examples of particular types of digital content hosted on social media platforms that are lost or at risk. This is mostly due to the fact that so many of the 'regular' social media platforms have tended toward more ways to mimic or copy TikTok-style videos, and making the distinction will become harder in the future since they all have similar functionality and ways to create photo/video content.

A consideration of merging the 'Legacy Interfaces and Services' entry with this entry to provide examples of particular interfaces of social media platforms that are lost or at risk.

2024 Interim Review

The 2024 Council identified a trend towards even greater risk due to a number of factors, summarized below.

Creators and archivists relying on consumer social media free at the point of use inhabit a precarious position. Free services may be favored by agencies or individual creators who are least able to respond to closure or loss because of the low barrier to entry associated with 'free at the point of use' services. Proprietary interfaces and services pose risks, as companies prevent third-party attempts to preserve either hosted content and/or the end-user experience of the environment. An inability to preserve social media interfaces diminishes future potential for emulation and may inhibit researchers' ability to glean important context, as described in the Bit List *2023 Review*.

Additional barriers to preservation via web capture are also present in terms of service for user accounts that explicitly prohibit crawling. For example, the X Terms of Service state "You may not access the Services in any way other than through the currently available, published interfaces that we provide. For example, this means that you cannot scrape the Services, try to work around any technical limitations we impose, or otherwise attempt to disrupt the operation of the Services" and "crawling or scraping the Services in any form, for any purpose without our prior written consent is expressly prohibited" (X, 2023). Another example, from the Facebook Terms of Service, states "You may not access or collect data from our Products using automated means" (Facebook, 2022).

An additional recommendation for the next 2025 review is to assess if 'proprietary formats' (e.g. the platform interfaces) adequately demonstrates the scope of this entry and answers the first bullet point of the 2023 Council recommendation. The 2023 recommendations for re-scoping and combining entries will also be assessed in more detail in 2025.

2024 Council members also raised concerns regarding Artificial Intelligence and Machine Learning, noting that for this entry and, more broadly, anything related to Social Media, an emerging risk is AI training fears. This manifests in two ways:

Social media users deleting their content entirely in an attempt to prevent it from being used or sold for training a Large Language Model (Edwards, 2024). However, as noted by the Council members, it

can be difficult to have a good sense of how widespread this is or how much it's affecting content that people want to preserve – but it is nonetheless worth noting.

Website operators blocking scrapers for similar reasons, (Voorhees, 2024). This primarily targets AI crawlers but may also affect web crawlers used for preservation. It can similarly be difficult to have a good sense of how widespread this is, but it seems likely that anti-crawler provisions are only going to get worse, not better.

Additional Information

Social media free-at-the-point-of-use remains at a critical risk due in large part to the policies of unregulated (or underregulated) corporate platforms such as Facebook, X (previously Twitter), and their parent companies. The content shared on these platforms and the history of the development of platform infrastructure and policy itself provide a critical source of information for policy-makers and researchers. The complete lack of preservation provision and deliberate obstruction of archiving attempts for public interest puts this valuable content at high risk of loss and draws attention to the critical risk posed by these examples of platforms.

Content hosted on social media platforms (that users might not have stored elsewhere) is at risk and users may lose the opportunity to keep their own data for personal archiving or to donate to an organization. Collecting organizations may lose the opportunity to archive hosted content within their collecting remit using web or API harvesting tools. In both instances, data remains at high risk because it is hosted by companies that could change policies or access on a whim. Also, the inability to archive even free content unless you have a login as an archivist (like with Browsertrix). Additionally, there are social media companies requiring payment to access data for preservation.

There are interfaces of social media platforms that researchers may want to see to study the evolution of the platforms over time (through web harvesting typically) that are at risk. Preservation is affected by researcher API access being shut down, halting preservation of entire platforms. There are also differences between the themes/collecting policies of institutions and researchers who are scraping their own data and depositing it in repositories.

Preserving this stuff en masse is still incredibly difficult, but many of these platforms allow the downloading of their own personal content/archives. However, these lose all the context of social media and therefore, whilst they do preserve the data, they do not preserve the essence of the material. Platforms like X (previously Twitter) have both opened and closed their API further in recent years, but others like Yahoo have closed, and Facebook, as well as X (formerly Twitter), continues to be almost hostile towards archiving and preservation attempts.

With digital materials from premium or institutional social media services, the business model and sustainability are more obvious, and contracts may be enforceable more readily. Moreover, because these services have a slightly higher barrier to entry, they may be favored by agencies that are better able to respond to closure or loss. Traditional web archiving can be employed where the user pays for a service, but the content is ultimately publicly available (such as Flickr). But much is unclear about how to preserve internal social media / closed networks that web archiving cannot get to, or existing tools do not cover.

Social media capture via web harvesting has become increasingly difficult. Social media platforms have done nothing to address the barriers to automated capture that prevent the preservation of even so-called public content. For example, campaign websites or other election-related content that is only published on Facebook or on X (previously Twitter) because these services are 'free.' This

content is of particular concern as it appears on no other website. Web archivists are constantly shifting strategies and approaches and trying out new (but limited) tools to best capture this content. If we cannot successfully preserve these platforms, we are missing out on documenting organizations, campaigns and elections around the globe. Much of this data exists as data sets based on aggregated use rather than individual files.

Often these are external proprietary platforms bound by intellectual property law and potentially privacy law which will impede the imminence of action. What recourse do archives or digital repositories have to deal with this and capture the materials?

Case Studies & Examples

- Mentioned examples of additional barriers to preservation via web capture present in terms of service for user accounts that explicitly prohibit crawling included X (2023) 'Terms of Service', Effective: September 29, 2023. See **X: Terms of Service**, *https://web.archive.org/web/20240611040225/https://x.com/en/tos* [accessed at 2024-09-06].

- **Stack Overflow users sabotage their posts after OpenAI deal**, Edwards, B., (2024), *Ars Technica https://arstechnica.com/information-technology/2024/05/stack-overflow-users-sabotage-their-posts-after-openai-deal/* [accessed at 2024-09-06].

- **How We're Trying to Protect MacStories from AI Bots and Web Crawlers – And How You Can, Too**, Voorhees, J., (2024), *https://www.macstories.net/stories/ways-you-can-protect-your-website-from-ai-web-crawlers/* [accessed at 2024-09-06].

- **Preserving Social Media**, Thomson, S., (2016), *Digital Preservation Coalition http://doi.org/10.7207/twr16-02*.

- The National Library of Scotland 'The Archive of Tomorrow: Health Information and Misinformation in the UK Web Archive' project, to record the proliferation of misinformation about coronavirus. See **Archive of Tomorrow**, National Library of Scotland, *https://www.nls.uk/about-us/working-with-others/archive-of-tomorrow/* [accessed at 2023-10-24].

- The archiving of the 'In Her Shoes' collection, part of the Archiving Reproductive Health (ARH) project. Working with key stakeholders, including activist organisations like Abortion Rights Campaign, Together for Yes, Terminations for Medical Reasons, Coalition to Repeal the Eighth, and many others, ARH gathered and preserved a selection of digital objects and research data, including social media, that tells part of the story of this historic campaign. ARH published collections of design and publicity material from activist groups, as well as a sequence of stories from the popular Facebook page 'In Her Shoes', a page where people anonymously shared stories of their experiences of being unable to access abortion in Ireland. This initiative received a 2022 Digital Preservation Award for Safeguarding the Digital Legacy. See **Archiving Reproductive Health**, Archiving Reproductive Health Project, (2022), *Digital Preservation Awards 2022. https://www.dpconline.org/events/digital-preservation-awards/dpa2022-archiving-reproductive-health* [accessed at 2023-10-24].

- **ArchiveSocial**, CIVICPLUS, *https://archivesocial.com/* [accessed at 2023-10-24].

- In the 2023 NDSA Web Archiving Survey Report, one of the major takeaways related to respondents' concerns about ability to collect social media—in particular, Twitter, Instagram,

Facebook, and Reddit. Content housed within social networks has always been difficult to capture for a myriad of reasons and recent changes to numerous social platforms have made this task harder. See **Web Archiving Survey Results: An NDSA Report**, National Digital Stewardship Alliance (NDSA), (2023), _https://doi.org/10.17605/OSF.IO/N5MYR_ [accessed at 2024-09-11].

- **Slop is the new name for unwanted AI-generated content**, Willison, S., (2024), _Simon Willison's Weblog_ _https://simonwillison.net/2024/May/8/slop/_ [accessed at 2024-09-12].

- **Mapping social media archiving initiatives: state of the art, Trends, and future perspectives**, Cannelli, B. , (2018), _IIPC Blog_ _https://netpreserveblog.wordpress.com/2022/11/30/mapping-social-media-archiving-initiatives-state-of-the-art-**Trends**-and-future-perspectives/_ [accessed at 2023-10-24].

- A 2022 report on a nationwide questionnaire survey conducted to obtain the responses of people to hypothetical scenarios of social media archiving by the National Diet Library in Japan, noting legal and ethical concerns as well as respondent views on the preserving of private data publicly available on social media. See **People's perceptions on social media archiving by the National Library of Japan**, Shiozaki, R., (2022), _https://doi.org/10.1177/01655515221108692_.

Digital Archives of Community Groups

Digital materials including ephemera, correspondence and campaign materials created as a by-product of small scale or ad-hoc community action groups.

Categories

- Community Archives

Examples

Archives of smaller and ad-hoc political and campaigning organizations; environmental protests; sports clubs; smaller religious groups; amateur music or drama; fan groups

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Poor documentation; lack of replication; lack of continuity funding; lack of residual mechanism; dependence on a small number of volunteers, lack of preservation mandate; lack of preservation thinking at the outset; conflation of backup with preservation; conflation of access and preservation;

inaccessible to web archiving; dependence on social media providers; distrust of 'official' agencies; uncertainty over IPR or the presence of orphaned works

Mitigations

Residual archive with residual funding able to receive and support collections; active user community; intellectual property managed to enable preservation

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

The Jury created this entry in 2019 as a subset of 'Community Archives and Community-Generated Content' which was split into two entries to provide greater specificity in recommendations for approaching the preservation of created as a by-product of small scale or ad-hoc community action groups (versus digital materials generated for significant purpose of a community initiative). The 2020 Jury identified a 2020 trend towards greater risk based on community groups such as sports clubs, religious communities, arts and political groups, often relying on volunteer effort, being unable to meet for extended periods in 2020. Moreover, the local community centres, clubs or places of worship on which they depend had closed, in some cases for good. This trend continued for 2021; the 2021 Jury commented that much of the content in community archives has easily preservable content, but resources are not directed towards them, basic digital preservation practices are not well embedded amongst the general population, and selective approaches are needed to get a handle on the situation and to find the resources to do the work.

The 2023 Council agreed with the classification of Critically Endangered. with the overall risks remaining on the same basis as before ('No change' to the 2023 trend). However, they also noted an increase in the significance of loss due to the fact that community heritage tends to be part of wider conversations within the international landscape.

2024 Interim Review

The 2024 Council agreed these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

Typically, born-digital material is more at risk as community groups may not know about the risk of loss. Many are unaware of digital preservation terminology. It is the ad-hoc nature of these groups and projects which is of great concern.

There is a significant need to raise awareness and provide a 'home' but also to do so with sufficient sensitivity so as to ensure community groups remain in control of their own material.

Communities who live in rural and remote areas may have a lack of access to services such as broadband connectivity, which is a well-reported issue and is often referred to as the 'digital divide'. Inadequate internet connectivity would diminish the capacity for these communities to access digital preservation solutions, such as cloud storage for digital assets. This is especially prevalent with personal photos and videos on mobile phones as possession of a mobile phone does not necessarily mean the user has adequate internet connectivity to be able to upload videos to web-based platforms.

AI could potentially be used to assist with easy access to simple, succinct explanations and principles of digital preservation and archiving solutions which would give these communities a wider understanding of the work being done and empower them to be able to do minimum digital preservation themselves.

Case Studies & Examples

- A new preservation toolkit resource, 'Digital Preservation for Community Archives', was released for World Digital Preservation Day on 7 November 2024. The preservation toolkit has been created alongside the University of Glasgow as part of the Our Heritage, Our Stories (OHOS) project, and has been developed in collaboration with a range of community archives. It will include a range of practical guidance, support, and advice to those working with digital records in the community context. See **Community Archives Digital Preservation Toolkit**, Digital Preservation Coalition, (2024), *Digital Preservation Coalition* http://doi.org/10.7207/communitytoolkit24-01.

- In the 2023 NDSA Web Archiving Survey Report, one of the major takeaways noted that since the first Web Archiving Survey in 2011, the landscape of web archiving has dramatically changed, and the 2022 survey results show an increase in web archiving practices by grassroots organizations, volunteer groups, and non-academic communities who seek to document their lived experiences. See **Web Archiving Survey Results: An NDSA Report**, National Digital Stewardship Alliance (NDSA) , (2019), https://doi.org/10.17605/OSF.IO/N5MYR [accessed at 2024-09-11].

- Digital inequality is recognized as one of the targets in the National Agreement on Closing the Gap, the objective of which is to "enable Aboriginal and Torres Strait Islander people and governments to work together to overcome the inequality experienced by Aboriginal and Torres Strait Islander people, and achieve life outcomes equal to all Australians.". Target 17 states that "By 2026, Aboriginal and Torres Strait Islander people have equal levels of digital inclusion". See **Closing the Gap Targets and Outcomes**, Closing the Gap, https://www.closingthegap.gov.au/national-agreement/targets [accessed at 2023-10-24].

- The "digital divide" in Australia has an impact on digital preservation of content generated by communities living in remote areas. See **Australia's digital divide means 2.8 million people remain 'highly excluded' from internet access**, Parke, E., (2022), *ABC* https://www.abc.net.au/news/2022-10-16/australia-digital-divide-millions-cannot-access-internet/101498042 [accessed at 2023-10-24].

- The "digital divide" in Australia has an impact on digital preservation of content generated by communities living in remote areas. See **'Digital divide': report finds some Australian rural mobile data speeds 90% slower than urban**, Al Khawaldeh, K., (2022), *The Guardian*

https://www.theguardian.com/australia-news/2022/dec/13/digital-divide-report-finds-some-australian-rural-mobile-data-speeds-90-slower-than-urban [accessed at 2023-10-24].

- **Byte-sized Bit List: Using the Bit List to garner action in support of Community Archives**, Wilson, A., (2023), *Digital Preservation Coalition* *https://www.dpconline.org/blog/bit-list-blog/bit-list-to-garner-action* [accessed at 2023-10-24].

Family or Personal Records

Digital content and communications generated for personal consumption in a domestic setting. These records are highly valuable to family members and those interested in genealogy. They can also have wider historical/research significance to collecting institutions.

Categories

- Personal Archives

Examples

Childhood photographs and videos; School or graduation photos; wedding photos and movies; electronic correspondence (email, messenger, WhatsApp)

Imminence	Effort
●●○○○	●●●●○
Action is recommended within five years, and detailed assessment within three years.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Storage on portable media or poor storage; dependence on devices or processes; dependence on obsolete or proprietary formats; storage media out of warranty; single copies; inappropriate dependence on service provider; inappropriate encryption or password protection; lack of awareness or planning; loss or lack of documentation; over-abundance; inability to act in a timely manner; confusion over intellectual property; lack of digital literacy; uncertainty over IPR or the presence of orphaned works

Mitigations

Replication; action in a timely manner; open formats; selection and appraisal; archival agency; education of digital preservation

Bit List History

Added to list: 2017

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This was introduced to the Bit List in 2017. Although research and advice on the preservation of personal records have been available for some time, outreach and training have not reached the audience, and there has been no material improvement in the risks faced by this category since 2017. It is reasonable to assume that the number of digital objects in this category has increased; thus, the consequences of loss have expanded but the 2021 Jury determined there had been no significant trend towards greater risk; content is being lost all the time despite digital materials that can easily be preserved with tools not widely available outside of institutions. Therefore, this is a public awareness campaign issue, and more tools need to be made easily available for people to be able to better preserve their own digital content.

The 2023 Council agreed with the Critically Endangered classification but noted that the definition for this entry did not mention the potential wider historical/research significance of some personal archives to collecting institutions and recommended rescoping the entry to make this clearer.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

Personal papers can provide insight into the lived experience of a wider range of people - archives of 'everyday' people are invaluable to social historians and personal archives of people with national/international significance complement institutional/public records.

There is a strong overlap with community archives, except noting that responsibility is even more localized. There is room breaking the entry down further into a series of components to represent the complexity more effectively and present a more nuanced action plan.

This matter needs awareness-raising. Education is needed, such as digital preservation as a survival skill for teenagers. Also, simple and cheap tools or pathways to preservation are needed.

Education to the public is critical for advocacy - these are the societal records of the future! Though having said that, what has survived in hardcopy has largely been through luck, and the same thing I think will be the same for digital. The same issues exist with glass plate negatives, photographs and certain emulsions and even printed digital photographs, brittle paper, fading ink etc.

Case Studies & Examples

- Such materials are increasingly moved into the cloud and the Google Photos case provides a good example. See **Last Chance To Download Your Photos: Google's Album Archive Will Be Deleted Tomorrow**, Monckton, P., (2023), *Forbes* *https://www.forbes.com/sites/paulmonckton/2023/07/17/google-warning-album-archive-will-be-canceled-in-two-days-download-your-photos-now/* [accessed at 2023-10-24].

- **Your Memories. Their Cloud**, Hill, K., (2022), *New York Times* *https://www.nytimes.com/2022/12/31/technology/cloud-data-storage-google-apple-meta.html* [accessed at 2023-10-24].

Media Inside Paper Files

Media inside paper files occurred in records since the 1980s and will continue to do so for many years.

Categories

- Portable Media

Examples

Digital media mixed with paper files in records offices and filing cabinets of almost every kind of enterprise

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Unsustainable effort to assess; exotic or obsolete media; poor storage; lack of descriptive labelling; uncertainty over IPR or the presence of orphaned works

Mitigations

Carefully labelled; managed programme of assessment and retrieval; robust media used

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 to report the significant amounts of digital media being transferred to archives folded into traditional files. The 2019 Jury noted that it is relatively simple to preserve this material once identified using standard tools, but it can be an 'unknown unknown,' and that assessment can seem overwhelming and, therefore it may overlap with other portable media risks

but has a higher risk classification. The 2021 Jury agreed on a 2021 trend towards greater risk due to the increased time sensitivity and need for conducting collection audits as soon as possible, in order to determine what you have to then work out a plan about opening carriers, assessing files, and extracting them if significant. The 2022 Taskforce agreed, with risks on the same basis as before ('No change' to trend)

The 2023 Council agreed with the risk classification of Critically Endangered with the overall risks remaining on the same basis as before ('No change' to trend). .

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

This is highly dependent on who is looking after the portable formats. There are good examples, for example in libraries, where disks are stored at the back of books or in front of magazines and can be processed at the point of acquisition. In archives, however, dealing with bit-level preservation of external media (often on legacy formats) is largely an unquantified problem, and so resource commitments will not be in place. So, there is a method and tools but simply no time committed and no proper assessment either. In other agencies, the issue will not have even been considered, and for them, it will be much harder over time with some inevitable loss.

Records of Local Government

Records from local government (i.e., below the state level) which are required for transparency and may be in many diverse forms, but in which the local authority may lack the capacity to manage the complex digital preservation requirements that arise.

Categories

• Public Records

Examples

Born digital records of small and medium-sized agencies; fasting-changing internal manuals, advice or policies shared electronically; records of care services; Documentation supporting long-lived contractual relations like Public Finance Initiatives; Organizational Slack channels; network drives; EDRMS; Email

Imminence	Effort
●●●●○	●●●○○
Action is recommended within twelve months. Detailed assessment is a priority.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of preservation infrastructure; conflation of backup with preservation; loss of authenticity or integrity; Long-lived business processes; poor storage; churn of staff; significant volumes or diversity of data; poorly developed digitization; ill-informed records management; poorly developed migration or normalization; long standing protocols or procedures that apply unsuitable paper processes to digital materials; encryption; political instability; lack of sustained funding; Uncertainty over IPR or the presence of orphaned works

Mitigations

Well managed data infrastructure; preservation enabled at the point of creation; carefully managed authenticity; use of persistent identifiers; finding aids; well managed records management processes; recognition of preservation requirements; strategic investment in digital preservation; preservation roadmap; participation in digital preservation community

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 as a subset of a previous entry for 'Records of long duration from Local Government or Other Government Agencies.' The split was intended to allow greater concentration on the challenges that these distinct types of agency face. Local government typically operates across a broad range of digital formats and services, but it is unclear and unlikely that relatively small archival agencies are properly funded locally to support the wide range of digital preservation requirements that arise. The 2020 Jury noted a trend towards greater risk based on significant political and economic upheaval placing additional strain on local government and its agencies, making already vulnerable records at greater risk. **Trends** towards greater risk was also noted by the 2021 Jury and 2022 Taskforce, contributing examples like Grenfell to demonstrate the precarity of local government records, especially when these risks overlap with records of non-governmental agencies, resulting in significance and impact of loss, the impetus for action and call to governing frameworks where failing in enforcement (and depending on the jurisdiction).

The 2023 Council generally agreed with the Critically Endangered classification with the overall risks remaining on the same basis as before ('No change' to trend). The 2023 Council additionally recommended revisiting and rescoping this entry as part of the next major revision of the Bit List. Some Council members recommended splitting this entry into separate entries to differentiate the various risks associated with different types of digital public records, Others raised concerns regarding the breadth of records held by local government, and that it is perhaps not appropriate to have a distinct entry or split entries for records of local governments but rather provide examples of different kinds of public records in and across other entries.

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

The diversity of 'local government records' makes this category quite difficult to score. First, local governments have differing responsibilities in different jurisdictions. For example local governments in the UK have more responsibilities than in Australia. Also, given the number of local government agencies in a state or country, the quality of recordkeeping and digital preservation practice can vary greatly. Additionally, the variety of records that are created by local governments means that some formats or record types may be generally at low risk, while others may be practically extinct. Given this complexity it is important to make clear that the imminence of action, significance of loss, and effort to preserve are context-dependent and generalized.

The main factors that reduce risk for these records are that local government is regulated, and there are clear recordkeeping standards that apply to digital records. Also they have consistent funding (although it may not be enough and may not be directed at digital preservation).

We feel that due to the breadth of records held by local governments, it is perhaps not appropriate for them to have a distinct record series, but rather be a featured example of other series. This approach would still assist in advocacy for local government as they would be able to cross reference their digital holdings against these classifications.

Significant research by the UK National Archives into Local Government Archives in England underlines the digital skills shortages that exist, especially with respect to preservation. There may be a benefit from splitting into a) legally required public record and b) **Additional Information** that may enrich our digital preservation of society. My assumption was that the roles and requirements for records management are clearly defined, but if this is not the case and there are inadequate resources to match the requirement, then the risk goes up.

Case Studies & Examples

- The Grenfell Tower fire and Grenfell Tower Inquiry illustrate the precarity of local government records, especially when third-party contractors are involved. Not only does it show the potential impact of aggravating conditions for Records of Local Government, but it also applies to those of Records of Non-Governmental Agencies. See **Grenfell Tower Inquiry**, *https://www.grenfelltowerinquiry.org.uk/* [accessed at 2023-10-24].

- In Scotland, there is record keeping legislation that is relevant and governs some of this, such as the Public Records Scotland Act of 2011. See **Public Records (Scotland) Act 2011**, National Records of Scotland, *https://www.nrscotland.gov.uk/record-keeping/public-records-scotland-act-2011* [accessed at 2023-10-24].

- The work and outputs of the EDRMS Preservation Taskforce, such as the EDRMS Preservation Toolkit, may be helpful for guidance in this context. See **EDRMS Preservation Toolkit**, Digital Preservation Coalition, (2021), *http://doi.org/10.7207/edrmstoolkit21-01* [accessed at 2023-10-24].

- The Kickstart Cymru project, which builds on the work that has been undertaken in Wales to preserve and provide access to digital information now and in the future. Underpinned by

the Digital Preservation Policy for Wales, it is a multi-stranded initiative involving archivists, researchers, consultants, students and IT professionals to promote digital preservation in the local authority, education and cultural sectors. This included funding for programme partnership of six archive services to support local government collaboration to solve shared problems with one issue identified being the need to provide long term access and to preserve records on business systems with operational lifespans less than the need to preserve the records. It is responsive to specific sectoral needs, but with an overarching aim of enhancing digital preservation capacity. Elements of the initiative include building skills; addressing specific digital preservation issues, co-creation of documentation and providing kits to undertake practical preservation. See **Kickstart Cymru: Enhancing digital preservation capacity in Wales**, Archive Wales, (2022), *Digital Preservation Awards 2022* https://www.dpconline.org/events/digital-preservation-awards/dpa2022-kickstart-cymru [accessed at 2023-10-24].

- The issues and approaches raised by the Tuvalu Future Now Project, a set of three major initiatives designed to preserve its nationhood, governance and culture in the event of a worst-case scenario. The third initiative is the development of a digital nation. It includes digitising and transferring access to government and consular services and all accompanying administrative systems into the cloud to enable elections to continue to be held, and government bodies to continue in their roles. It also includes a virtual copy of Te Afualiku, the first island in Tuvalu to be digitally recreated through satellite imagery, photos and drone footage, creating a digital twin to not only help inform decisions around urban planning and development but also examine how to use augmented and virtual reality to allow displaced and future generations of Tuvaluans to continue to exist as both a culture and a nation, complete with ancestral knowledge and value systems. If this concept becomes a reality, the Tuvaluan people will be able to interact with one another in a digital dimension, in a way that imitates real life and helps to preserve shared language and customs. See **Facing extinction, Tuvalu considers the digital clone of a country**, Fainu, K., (2023), *The Guardian* https://www.theguardian.com/world/2023/jun/27/tuvalu-climate-crisis-rising-sea-levels-pacific-island-nation-country-digital-clone [accessed at 2023-10-24].

Records of Non-Governmental Agencies

Records of independent agencies and contractors that act on behalf of the state in the delivery of public services, and which may be present in many diverse forms, but for which the NGO or contractors may lack the capacity to meet the complex digital preservation requirements that arise, or may have a business motive to minimize or ignore requirements for the maintenance of the record.

Categories

- Public Records

Examples

Born digital records of small and medium-sized agencies; fasting-changing internal manuals, advice or policies shared on intranets or EDRMS; records of care services; historic guidelines and manuals

which evidence 'best practice at the time'; Documentation supporting long-lived contractual relations like Public Finance Initiatives; Organizational Slack channels; network drives; EDRMS; Email

Imminence	Effort
●●●●○	●●●○○
Action is recommended within twelve months. Detailed assessment is a priority.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of preservation infrastructure; conflation of backup with preservation; loss of authenticity or integrity; Long-lived business processes; poor storage; churn of staff; significant volumes or diversity of data; poorly developed digitization specifications; ill-informed records management; poorly developed migration or normalizations specifications; longstanding protocols or procedures that apply unsuitable paper processes to digital materials; encryption; political instability; lack of sustained funding; denial of responsibility; failure to include archives within contract from commissioning agency; Uncertainty over IPR or the presence of orphaned works

Mitigations

Well managed data infrastructure; preservation enabled at the point of creation; carefully managed authenticity; use of persistent identifiers; finding aids; well managed records management processes; application of records management standards; recognition of preservation requirements at highest levels; strategic investment in digital preservation; transfer protocols to public archive; participation in digital preservation community

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 as a subset of a previous entry for 'Records of long duration from Local Government or Other Government Agencies.' The split was intended to allow greater concentration on the challenges that these different types of agencies face. Non-governmental organizations typically operate across a broad range of digital formats and services acting on behalf of the public sector. The 2020 Jury noted the trend towards greater risk based on 2020 being a year of significant political and economic upheaval, putting additional strain on NGOs in these circumstances where already vulnerable records are likely to be at greater risk. **Trends** towards greater risk were also noted by the 2021 Jury and 2022 Taskforce, contributing examples like Grenfell to demonstrate the

precarity of non-government agencies, especially when these risks overlap with those of local government, resulting in significance and impact of loss, the impetus for action and call to governing frameworks failing in enforcement for these agencies (e.g., examining current recordkeeping regimes keeping them accountable).

The 2023 Council generally agreed with the Critically Endangered classification, with the overall risks remaining on the same basis as before ('No change' to trend). The 2023 Council additionally recommended revisiting and rescoping this entry as part of the next major revision of the Bit List. Some Council members recommended splitting this entry into separate entries to differentiate the various risks associated with different types of digital non-governmental records. Others suggested that it is perhaps not appropriate to have a distinct entry or split entries for records of non-governmental agencies but rather provide examples of different kinds of these digital materials in and across other entries.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

There is a large variation in the types of records held by NGOs. Additionally, the quality of digital preservation performed by NGOs can vary widely. Therefore, the same approach to scoring was taken for this entry as the one above.

We consider records of NGOs to be at greater risk due to there being less regulation, and the regulations that exist being less stringently enforced.

An additional risk factor for these records is a blurring of the lines of responsibility, which can lead to records 'falling through gaps', or to difficulties funding digital preservation practice. This can be further complicated by outdated legislation which does not take into account the complexity of privatisation and public/private partnerships. For example, the legislation that PROV operates under is 50 years old. This, in turn, can lead to regulation and enforcement being more complex than it is for government agencies.

Case Studies & Examples

- The Grenfell Tower fire and Grenfell Tower Inquiry illustrate the precarity of local government records, especially when third-party contractors are involved. Not only does it show the potential impact of aggravating conditions for Records of Local Government, but it also applies to those of Records of Non-Governmental Agencies. See **Grenfell Tower Inquiry**, *https://www.grenfelltowerinquiry.org.uk/* [accessed at 2023-10-24].

- There can be some grey areas depending on the legislative context. The Public Records Scotland Act 2011, for example, covers government agencies and any non-government org contracted to do work on behalf of government agencies. See **Public Records (Scotland) Act 2011**, National Records of Scotland, *https://www.nrscotland.gov.uk/record-keeping/public-records-scotland-act-2011* [accessed at 2023-10-24].

- There can be some grey areas depending on the legislative context. The Public Records Scotland Act 2011, for example, covers government agencies and any non-government org

contracted to do work on behalf of government agencies. See **Policy Commons**, *https://policycommons.net/* [accessed at 2023-10-24].

Records of Quasi Non-Governmental Agencies

Records from agencies at arms-length to government whether locally, nationally or internationally. They may be required to maintain archives for the purposes of transparency, sometimes for extended periods, and **sometimes** in diverse and complicated forms.

Categories

- Public Records

Examples

Records of non-executive state or national agencies; museum or leisure trusts; industry or public regulators; public audit services; public-good funding and investment agencies; autonomous and semi-autonomous public agencies; sovereign wealth funds; public/private partnerships; publicly owned companies.

Imminence	Effort
●●●●○	●●●○○
Action is recommended within twelve months. Detailed assessment is a priority.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of preservation infrastructure; conflation of backup with preservation; loss of authenticity or integrity; Long-lived business processes; poor storage; churn of staff; significant volumes or diversity of data; poorly developed digitization specifications; ill-informed records management; poorly developed migration or normalizations specifications; long-standing protocols or procedures that apply unsuitable paper processes to digital materials; encryption; political instability; lack of sustained funding; Uncertainty over IPR or the presence of orphaned works

Mitigations

Well-managed data infrastructure; preservation enabled at the point of creation; carefully managed authenticity; use of persistent identifiers; finding aids; well-managed records management processes; application of records management standards; recognition of preservation requirements at highest levels; strategic investment in digital preservation; preservation roadmap; participation in digital preservation community

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 as a subset of a previous entry for 'Records of long duration from Local Government or Other Government Agencies.' The split was intended to allow greater concentration on the challenges that these different types of agencies face. Records of quasi non-governmental agencies are at arm's length to government, but the 'QuaNGO' or 'ALEO' (Arms-Length Executive Organization) may lack the capacity to meet complex digital preservation requirements that arise, nor be able to deposit in the government archive. The 2021 Jury added that arm's length bodies are still public bodies, and they have a duty of care for maintaining evidence of their actions and transactions. They often receive public funding, and depending on the archives, legislation may be required to transfer to an archive. The issue is when there is a lack of clarity regarding the recordkeeping requirements or neglect of records and information once it has outlived its usefulness. These bodies still create records that affect citizen lives and have a duty to document, and therefore changed the classification from Endangered to Critically Endangered. The 2021 Jury and 2022 Taskforce noted a trend towards greater risk when looking at the precarity of records in QuaNGO agencies in periods of significant political and economic upheaval creating greater strains for funding to support preservation capacity.

The 2023 Council generally agreed with the Critically Endangered classification with the overall risks remaining on the same basis as before ('No change' to trend). The 2023 Council additionally recommended revisiting and rescoping this entry as part of the next major revision of the Bit List. Some recommended splitting this entry into separate entries to differentiate the various risks associated with different types of digital records of quasi non-governmental agencies. Others suggested that it is perhaps not appropriate to have a distinct entry or split entries but rather provide examples of different kinds of these digital materials in and across other entries.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

There is a large variation in the types of records held by QuaNGOs and/or ALEOs. Additionally the quality of digital preservation performed can vary widely. Therefore, the same approach to scoring was taken for this entry as the one above.

Similar to the risks of NGOs, we consider these records to be at greater risk due to there being less regulation, and the regulations that exist being less stringently enforced.

An additional risk factor for these records is a blurring of the lines of responsibility, which can lead to records 'falling through gaps', or to difficulties funding digital preservation practice. This can be further complicated by outdated legislation which does not take into account the complexity of privatization and public/private partnerships.

Although the split draws attention to the different pressures faced by QuaNGOs it could be further subdivided into legally required public records and **Additional Information** that may enrich our digital preservation of society. The classification assumes that the roles and requirements for records management are clearly defined, but if this is not the case or there are inadequate resources to match the requirements, then the risk goes up.

While the 2022 trend shows increases in risk there are some green shoots of hope in Ireland found when working actively with the agencies, and communicating some of the concerns they have for their data so there's better awareness and hopefully that will turn into action.

Case Studies & Examples

- **Digital preservation of digital documents of an NGO in Mexico**, Muñoz López, C., (2021), *Digital Preservation Coalition* https://www.dpconline.org/blog/wdpd/cmunozlopez-wdpd21 [accessed at 2023-10-24].

Collections Information Management Data and Systems

Descriptive information and data, covering both the systems (databases) and the data they contain. This includes information made publicly available, and information only available for internal use.

Categories

- Museum and Gallery

Examples

Covered under this entry are third-party and in-house collections information management systems and databases, current and legacy, both large and small (e.g., Microsoft Access, FileMaker Pro)

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Poor or no documentation; lack of technical and preservation infrastructure; complex interdependencies on specific hardware, software or operating systems; significant volumes or diversity of data; conflation of access with preservation; dependence on proprietary products; lack of preservation capacity in museum or gallery; poorly developed or no processes for migration or normalization; uncertainty over IPR or the presence of orphaned works

Mitigations

Strong documentation; preservation capability; strong repository and preservation technical infrastructure; good descriptive cataloguing; use of open formats and open source software; considered data management planning; licencing that enables preservation

Bit List History

Added to list: 2023

Trends

- **2024:** No change.

Last Review

2023 Review

This entry was added in 2019 under 'Digital Materials in Museums and Galleries' and previously rescoped in 2021 to 'Supporting Digital Materials for Museums and Galleries'.

The 2023 Bit List Council superseded the entry, splitting it into six discrete entries as the scope of the single entry was too broad to provide the guidance needed. The recommendation to break this entry down was also made by the 2021 Jury, as the types of digital collections content in museums can be vast and offer particular risks in museum and gallery contexts. For this entry on collections information management data and systems, context is important. For example, smaller 'in-house developed' and 'cottage industry' systems may be at higher risk than larger third-party systems with significant international buy-in and support. They additionally recommended that the next major 2025 review consider whether or not to split out the data held in Collections Information Management Systems from the systems themselves.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

While there are no changes to the risk trend, the context in the wider cultural sector of the 2023 British Library Cybersecurity incident should be taken into account that this type of data may be impacted if museum and galleries are on the receiving end of cybersecurity incidents. Museums and galleries may have older legacy systems in place (and potentially with fewer backups and redundancies built in), and so the impact may be significant for them.

Additional Information

Databases and catalogues can have a knock-on effect. The information they contain is valuable for contextualizing and understanding the resources they describe. Without them, meaning may be lost even if bits are not.

Case Studies & Examples

- The British Library cyber incident as a case example of issues arising. See **Learning Lessons from the Cyber-attack: British Library cyber incident review**, The British Library, (2024), *https://www.bl.uk/home/british-library-cyber-incident-review-8-march-2024.pdf/* [accessed at 2024-09-06].

Corporate Records of Long Duration on Network Drives, Intranets and EDRMS

Records on internal corporate network drives, intranets or document management services where access is limited to a distinct group of users, and in which the lifecycle of the record or the business processes they support is greater than the technology on which they are created or retained.

Categories

- Sensitive Data

Examples

Born-digital records of small and medium sized enterprises; fasting-changing internal manuals, advice or policies shared on intranets or EDRMS; records of long-lived products and services; Historic guidelines and manuals which evidence 'best practice'; Documentation supporting long-lived contractual relations; Online terms and conditions; Corporate Slack channels; Google Drives; EDRMS; Email

Imminence	Effort
●●●●○	●●●○○
Action is recommended within twelve months. Detailed assessment is a priority.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of systematic preservation function; lack of preservation path or plan for data; dependence on proprietary products or formats; poor management of data protection; political or commercial interference; lack of offline equivalent; over-abundance through poor disposal or naming and version control; lack of capacity; lack of commitment; loss or lack of documentation; sector-specific software or data types; encryption; Uncertainty over IPR or the presence of orphaned works owners

Mitigations

Preservation infrastructure and pathways; replication; appraisal and selection including de-duplication

Bit List History

Added to list: 2017

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2017 to draw attention to the pressing need for digital preservation in business, especially in small to medium enterprises. The 2020 Jury noted how the Covid Pandemic has caused profound dislocation across the economy and placed many companies and agencies at financial risk. The likelihood of liquidation, mergers or acquisitions means that these records are trending towards greater risk. The 2021 Jury agreed with the trend towards greater, adding that increased risk is not necessarily because there are no assigned parent archives to take on these materials; rather, it is because they too often sit in these spaces for some time before being transferred to the archives. They are often not well managed or maintained by their creating agencies, putting them at risk of accidental deletion or corruption. There remain increased risks without business continuity and trust.

The 2023 Council agreed with the continued Endangered classification with the overall risks remaining on the same basis as before ('No change' to trend), though also noting an increased imminence of action needed for this entry and highlighted the importance of processes in the preservation of these records. They recommended that the next review considers rescoping this entry, possibly adding a new entry about covering sensitive data in databases that would fall into both the Sensitive Data and new Databases digital species groups.

2024 Interim Review

The 2024 Council recommends that a major rescoping of the Sensitive Data species is necessary, with plans to remove it as a species and incorporate key elements and examples to relevant entries for the next 2025 Bit List. This is because it is not clear how sensitive data works as a species, when many of the other species mentioned could have sensitive data concerns, and the sensitivity of the data is more like an extra category of risk that potentially applies across any species. Additionally, further input from those working with corporate records in this context is invited.

Additional Information

Corporate records should form part of organizational records management schemes, and so responsibilities should be clear; however, this may be much more challenging for smaller organizations without dedicated roles or with complex data types.

Processes become as important as technology when it comes to preserving this kind of material. If an organization does not have good records organization, naming conventions etc. that may make material as vulnerable to loss as technological failure or format obsolescence could.

Closer collaboration over the digital record lifecycle with recordkeeping organizations such as IRMS/ARA and digital preservation organizations would help to ensure best practice from (before) record creation to its long-term preservation and would help to identify any risks and bridge gaps 'from the cradle to the grave'. Joining forces and resources will enable the community to raise awareness of the impact of best practices on the organizational governance and related efficiencies.

Case Studies & Examples

- The work and outputs of the EDRMS Preservation Taskforce over the last year, for example the EDRMS Preservation Toolkit, may be helpful for guidance and examples in this context.

Digital Radio Recordings

Primary and/or original recordings of radio broadcasts generated live but often poorly stored thereafter, for example offline recordings on single LTO (Linear Tape Open) Tapes.

Categories

- Sound and Vision

Examples

Broadcast archives include but are not limited to national, local, public, commercial radio

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of archival mandate; lack of capability of archive; lack of policy or capacity within broadcaster; small or unprofitable broadcaster; uncertainty over IPR or the presence of orphaned works; overzealous rights management protection; device or software dependence; dependence on proprietary or obsolete formats; lack or loss of documentation; little use or inaccessibility; storage (typically tapes) older than warranty; lack of media refreshment plan; lack of error or integrity checking process; single copies

Mitigations

Archival responsibility accepted and acted upon; replication; refreshment of media; good documentation; active in digital preservation community; trusted repository; content re-used

Bit List History

Added to list: 2017

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2017 as a separate entry due to concern over recordings on LTO tapes. These provide between 15 and 30 years' storage which may be less depending on usage and storage conditions. LTO1 and LTO2, released in 2000 and 2003 respectively, have largely reached the final phases of viability. Reader compatibility may be more problematic than media resilience, however. Drives supporting newer releases of the format are typically only compatible within two generations, and experience with the recently released LTO8 suggests that it is only backwardly compatible to one generation. For instance, one major national archive and library had decided to expedite migration away from LTO6, which is becoming obsolete more quickly than anticipated. Through time, the risks to collections that have not been refreshed or replicated from early LTO tapes expand. Thus, the overall trend is towards greater risk when collections are not migrated. Older formats, perhaps as recently as LTO6, extinction events should be anticipated within two to five years. The 2020 and 2021 Juries, as well as the 2022 Taskforce, agreed with the Endangered classification, noting no change to the trend (they agreed these risks remain on the same basis as before).

The 2023 Council similarly agreed with the Endangered classification with the overall risks remaining on the same basis as before ('No change' to trend). The 2023 Bit List Council also recommended that the next major review for the Bit List consider creating a new 'Broadcast' species group where this entry may be better suited.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

Depending on the legislative context, digital radio recordings may fall under published works if they are broadcast. Additionally, if the recordings are broadcast by a taxpayer-funded broadcaster, record-keeping guidelines may already exist to advise how long content should be kept, which would then inform a selective approach to caring and looking after them for as long as required.

While broadcasters may keep their own programmes, they are often not comprehensively collected or archived by memory institutions as contributions to cultural heritage.

Case Studies & Examples

- The Radio Preservation Task Force (RPTF), created in early 2014, grew out of the Library of Congress's National Recording Preservation Plan (2012). See **Radio Preservation Task Force** , (2014), *Library of Congress* https://www.loc.gov/programs/national-recording-preservation-plan/about-this-program/radio-preservation-task-force/ [accessed at 2023-10-24].

- The British Library Save Our Sounds National Radio Archive project (2018) to create a digital radio archive that will preserve a representative proportion of ongoing UK radio output and make this available for research. See **National Radio Archive**, (2018), *British Library* https://www.bl.uk/projects/national-radio-archive [accessed at 2023-10-24].

- **The Valorization of the Tunisian Radio Archive in the Artificial Intelligence Era**, Meddeb, S. and Cecchine, R. , (2023), *https://hdl.handle.net/2142/121105*.

Electronic Hospital and Medical Records

Personal medical records and records of hospital treatment are increasingly—if not uniformly—born digital. By implication, those records should be retained through the lifetime of the patient, or in some instances longer as required for intergenerational study; and yet there is little evidence of the medical profession participating in the digital preservation community.

Categories

- Sensitive Data

Examples

Medical scans; records of treatment and care plans; health advice and notifications

Imminence	Effort
●●●●●	●●●●●
Immediate action necessary. Where detected, should be stabilized and reported as a matter of urgency.	Loss seems inevitable. Loss has already occurred or is expected to occur before tools or techniques develop.

Hazards

Loss of context; loss of authenticity or integrity; poor storage; lack of understanding; churn of staff; significant volumes of data; significant diversity of data; ill-informed records management; poorly developed transfer and integrity checking; poorly developed migration or normalizations specifications; long standing protocols or procedures that apply unsuitable paper processes to digital materials; encryption; Uncertainty over IPR or the presence of orphaned works owners

Mitigations

Well-managed data infrastructure; preservation enabled at the point of creation; carefully managed authenticity; use of persistent identifiers; well-managed records management processes; application of records management standards; recognition of preservation requirements at highest levels; strategic investment in digital preservation; preservation roadmap; participation in the digital preservation community

Bit List History

Added to list: 2017

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was first submitted in 2017 under 'Medical and hospital records.' At that time, there was limited capacity to address the topic. It was published as 'of concern' to revisit and review by the 2019 Jury and also independently received as a submission to the open nomination process under 'Electronic hospital and medical records.' The entry covers a broad range of material, and it may be useful in future years to split the entry into more discrete entries. Still, the 2021 Jury agreed to keep the current description and classification to draw attention to the scale of the digital preservation challenges which arise in hospitals and the medical profession. The same reasoning for greater risk in 2020 was used for 2021; there has been significant strain through the Covid pandemic, with resources stretched to meet an overwhelming demand and rigid, exacting protocols. In this environment, it is hard to avoid the sense that records are also now at greater risk. The Jury further commented that hospital records may be at greater risk than we think, where there may already be poor maintenance of records during their lifecycle, poor migration planning, etc. The 2022 Taskforce recommended that the 2023 Council bring in additional subject matter expertise for feedback and comment on any changes in risks relating to growth and volume of born-digital records, increasing or peculiar budget strain conditions, changes pertaining to sensitivity and potential destruction linked to ransomware or conflicts.

The 2023 Council agreed with the previous Endangered classification with the overall risks remaining on the same basis as before ('No change' to trend) though also suggesting an increased timeline for imminence of action and greater inevitability of loss.

2024 Interim Review

The 2024 Council recommends that a major rescoping of the Sensitive Data species is necessary, with plans to remove it as a species and incorporate key elements and examples to relevant entries for the next 2025 Bit List. This is because it is not clear how sensitive data works as a species, when many of the other species mentioned could have sensitive data concerns, and the sensitivity of the data is more like an extra category of risk that potentially applies across any species.

Additional Information

Increasing sensitivity and awareness of data protection requirements could act inadvertently as a barrier to lifecycle data management. It is striking how little evidence is of the health technology companies participating in the global digital preservation community.

The processes implemented by Sao Joao hospital (see below) are encouraging, but too many medical establishments are operating in an excessively ad-hoc way when it comes to records management. As well as preservation, issues of data protection and ethical obligations are to the forefront when working with this kind of material.

Case Studies & Examples

- The São João University Hospital Center (SJUHC) Health Records Repository project offers an example of changing practices relating to the project's implementation of a long-term digital preservation repository capable of ingesting, preserving and providing access to digital clinical information. As part of the Hospital's digital transformation strategy, the Health Records Repository promotes change in the management of daily medical records through the implementation of procedures for preparation, digitization and preservation of health records. The results of the last two years of activity of the Health Records Digital Repository reveal a higher efficiency in the access and reuse of clinical information in the context of healthcare. This initiative was nominated for a 2022 Digital Preservation Award. See **Long-**

term preservation of Digital Health Records, SJUHC and the Portuguese National Archives, (2022), *Digital Preservation Awards 2022* https://www.dpconline.org/events/digital-preservation-awards/dpa2022-digital-health-records [accessed at 2023-10-24].

- The National Library of Scotland 'The Archive of Tomorrow: Health Information and Misinformation in the UK Web Archive' project as it relates to capture of health advice published on the web. See **Archive of Tomorrow**, *National Library of Scotland* https://www.nls.uk/about-us/working-with-others/archive-of-tomorrow/ [accessed at 2023-10-24].

- **Conti cyber attack on the HSE: Independent Post Incident Review**, Health Service Executive Ireland , (2018), https://www.hse.ie/eng/services/publications/conti-cyber-attack-on-the-hse-full-report.pdf [accessed at 2023-10-24].

Older Open Source Intelligence Sources

Older open source intelligence produced, collected and analysed from publicly available social media and web content with the purpose of answering a specific intelligence question and that supports crowd-sourced investigation and fact-checking to verify or refute claims of state agencies and rebel groups in the context of historic political or military conflict.

Categories

- Legal Data

Examples

Social media sources relating to the Arab spring.

Imminence	Effort
●●●●●	●●●●○
Immediate action necessary. Where detected, should be stabilized and reported as a matter of urgency.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

N/A

Mitigations

Offline backup documented and available for recovery; clarity over IPR and no orphaned works

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 from a nominated entry that was split into three subsets by the 2019 Jury relating to current, recent, and historic sources. This entry relates in particular to materials published at the time of the 'Arab' spring. Social media companies had initially taken little or no action with respect to social media content in conflict zones, taking the view either that they were mere technical platforms and therefore not responsible for editorial; or that the platforms were being used largely for social good, loosening the control of the media from oppressive regimes. However, as the Arab Spring progressed, the companies came under significant pressure to monitor content with more care, in part because terrorist groups had begun using social media platforms for propaganda purposes. The social media companies responded by implementing algorithms that removed or deleted content. This had the unintended consequence of deleting or suppressing content that was being used in open source investigation for journalistic or judicial purposes and may have resulted in refutation or prosecution. The 2019 Jury recognized the duty of care that social media companies have towards their users and is in no sense seeking to have that material re-published on the open web but noted the unintended consequence for journalists and investigatory authorities from the rush to deletion, illustrating how this entry further underlines the relative fragility of all social media content. The 2021 Jury agreed with the current classification and description with no change to trend. The 2022 Taskforce also found no change to trend.

The 2023 Council agreed with the current Practically Extinct classification with the overall risks remaining on the same basis as before ('No change' to trend). The 2023 Council also added clarification to the meaning of 'open source' for this entry, to explain its meaning in relation to intelligence openly available online, noting that open source can also refer to a specific software or content licence that permits limited uses of IP so this distinction would be helpful for readers.

This is important for social context but may be picked up inadvertently through other ways; it remains ambiguous about who has ultimate responsibility for collecting and preserving this.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

The Council acknowledges the continuing challenge of ensuring the preservation of complete and accurate resources given that: platform owners continue to be obliged to remove content that violates community standards; copyright and ownership increasingly hinders capture of the open source materials; and with the rise in fakes, preservationists must attend to standards for legal admissibility and authentication which vary from one jurisdiction to another.

Additional Information

The Council also added clarification to the meaning of 'open source' for this entry, to explain its meaning in relation to intelligence openly available online, noting that open source can also refer to

a specific software or content licence that permits limited uses of IP so this distinction would be helpful for readers.

This is important for social context but may be picked up inadvertently through other ways; it remains ambiguous about who has ultimate responsibility for collecting and preserving this.

Case Studies & Examples

- **Virtual Uprisings: On the Interaction of New Social Media, Traditional Media Coverage and Urban Space during the 'Arab Spring.''**, Al Sayyad, N. and Guvenc, M., (2010), *Urban Studies* *https://doi.org/10.1177/0042098013505881*.

- **Preserving Open Source Digital Evidence a Guide for Practitioners Working on Dealing with the Past Essential**, Baumhofer, E. and Reilly, B., (2022), *Swiss Peace* *https://www.swisspeace.ch/assets/publications/Essentials/Essentials20221031_Preserving-Open-Source-Digital-Evidence.pdf* [accessed at 2024-10-07].

- **Documenting Revolution in the Middle East**, Dougherty, R., (2021), *Focus on Global Resources, CRL* *https://www.crl.edu/focus/article/7437* [accessed at 2023-10-24].

- **Bellingcat and beyond. The future for Bellingcat and online open source investigation**, Higgins, E., (2019), <Bellingcat and beyond. The future for Bellingcat and online open source investigation> [accessed at 2023-10-24].

---

e)   **Out of Band**

Correspondence and Informal Records of Research

Correspondence and other records which describe the configuration and delivery of research, but which are ancillary to the core research outputs, including reviews, drafts and correspondence between researchers.

Categories

- Research Outputs

Examples

Email boxes of senior academics; social media posts; personal spaces on institutional networks

Imminence	Effort
●●●○○	●●●●○
Action is recommended within three years, detailed assessment within one year.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Originating researcher no longer active or changed research focus; staff on temporary contracts; dependence on single student or staff member; weak or fluid institutional commitment to subject matter; weak institutional commitment to data sharing; uncertainty over IPR or the presence of orphaned works; encryption

Mitigations

Recognition of value of correspondence; integration with CRIS; routine use of EDRMS; documented and managed professionally; separation of personal and corporate identities

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was introduced in 2017 under 'Research Data,' though without specific reference to the correspondence and records of research. In 2019, the Jury split this entry into a range of contexts for research outputs. There was a 2020 trend towards greater risk based on education and research institutions facing budget uncertainties. A number of institutions introduced early severance schemes or put staff on short-term contracts at greater risk of redundancy. While this puts other types of research output at risk, the personal nature of correspondence means that the risks are intensified, and so this entry faced greater risks than those identified in 2019. The 2020 and 2021 Juries, as well as the 2022 Taskforce, agreed there had been no significant change in the trend over the preceding years with risks remaining on the same basis as before ('No change' to trend).

The 2023 Council similarly agreed with the Critically Endangered risk classification with risks on the same basis as before ('No change' to trend).

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

In an ideal circumstance, correspondence should be stored in EDRMS systems separately from research data and subject to different retention schedules, i.e., 10-20 years. There may be challenges connecting the EDRMs holdings to the research data and vice versa.

Advocacy and research in regards to the scale of the problem may be required to encourage academics to use EDRMs, for example, correspondence and integration with CRIS. Simplified tools and workflows to move data from CRIS to Repository to Preservation systems

There are significant cultural issues to preserving these materials. Researchers may be unlikely to see the value in correspondence and other documents; these may be seen as ephemeral. There is also a risk to the preservation of correspondence through channels outside of the university email that may be harder to capture and preserve, such as Teams or WhatsApp. There should also be an encouragement to researchers to keep only what is needed and only for so long as the retention period requires. More often than not, records of correspondence will not require long-term preservation.

Research project management records may have funding agency retention periods varying from 3 years to 10 years after the completion of the project for standard projects, to 20 years for more innovative projects, stretching to permanent retention to first-of-a-kind research. Advocating researchers to identify which group their records belong to, let alone get them to transfer different types of correspondence to a platform which has retention and preservation capability as part of it, is a massive task.

Grey Literature

Semi-published research outputs such as blogs, dissertations, informal conference papers or commissioned reports which are not formally published but which can contain original and insightful contributions within scholarly communications. This entry covers a wide spectrum of very diverse types of materials which all have different preservation considerations.

Categories

• Research Outputs

Examples

Blogs, technical reports, conference papers, dissertations, commercial research

Imminence	Effort
●●●○○	●●●●○
Action is recommended within three years, detailed assessment within one year.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Originating researcher no longer active or changed research focus; staff on temporary contracts; dependence on single student or staff member; weak or fluid institutional commitment to subject matter; weak institutional commitment to data sharing; uncertainty over IPR or the presence of orphaned works; encryption; lack of recognition; non-disclosure agreements

Mitigations

Use of persistent identifiers; embedded within repository infrastructure; quality assurance

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was introduced in 2017 under 'Research Data,' though without explicit reference to grey literature. In 2019, the Jury split this entry into a range of contexts for research outputs. This entry represents activities which build towards formal publications and research outputs but which do not typically accumulate in institutional repositories. The 2020 Jury noted a trend towards greater risk because higher education and research institutions faced budget uncertainties, and a number of institutions introduced early severance schemes or put staff on short-term contracts at greater risk of redundancy. While this puts other types of research output at risk, the ad hoc nature of grey literature placed it at greater risk. The 2021 Jury agreed with the Critically Endangered classification but argued the content of grey literature is not entirely unique if it eventually makes its way into published outputs, and they also noted improvements and initiatives towards preservation of semi-published research data and outputs over the last year, which together led to a consensus on a 2021 trend towards reduced risk. The 2022 Taskforce agreed with the 2021 assessment, with risks remaining on the same basis as described (no change to the trend).

The 2023 Council agreed with the Critically Endangered classification and that overall risks remained on the same basis as before ('No change' to trend), but noted that there will always be an element of risk to materials under this entry due to its semi-official nature. The Council also noted that this entry covers a wide spectrum of material, and all had different preservation considerations.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

Loss of material like this would be common in the analogue world, but in the digital age, we have the capacity and perhaps something of a responsibility to ensure that it is captured: more of an opportunity lost to extend the available research resource. The ADS's Grey Literature Library demonstrates what could be done if information architectures are deployed to mirror and extend professional practice.

Workflows and policies regarding tagging, collecting and EDRMS may help protect such data into the future. Past materials are almost certainly partially lost.

Not all funder-maintained specialist repositories accept grey literature for long-term storage (e.g., UKRI-NERC EDS). These are redirected to generic open data repositories such as Zenodo which mint DOIs but do not offer data quality assurance for different data types.

Case Studies & Examples

- The Policy Commons has a mission to index and preserve grey literature from IGOs, NGOs, think tanks, governments and, to date, indexing and preserving around 4 million items from c.11,000 institutions from across the world. See **Policy Commons**, *https://policycommons.net/* [accessed at 2023-10-24].

Legacy Research Web Collections

Research related collections of digital content on the web which are now outdated and/or no longer actively maintained. This can include software and published or unpublished source code.

Categories

- Web

- Research Outputs

Examples

Academic and institutional websites from the first decade of the web containing details of research projects and interests as well as research data

Imminence	Effort
●●●●○	●●●●○
Action is recommended within twelve months. Detailed assessment is a priority.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Inaccessible to web archive; bespoke code; insufficient documentation; uncertainty over IPR or the presence of orphaned works

Mitigations

Secured by web archive; documentation and rights information published alongside material

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019. While there are overlaps with 'Semi-Published Research Data' and 'Unpublished Research Data' entries, it is a separate entry to distinguish between 'current' and 'legacy' collections with different risk profiles. In 2020, the fact that materials of legacy web collections were no longer actively maintained increased the risk classification to Critically Endangered. The 2021 Jury agreed with these distinctions, adding that loss has already occurred and future loss can be prevented through approaches such as web archiving and code preservation. They identified a 2021 risk toward greater risk based on noted security issues posed by hosting legacy technology software and services which prompted disposal of content imminently without adequate review or selection. The 2022 Taskforce agreed with this assessment, noting no change to the trend (it remained on the same basis as before).

The 2023 Council agreed with the Critically Endangered classification with risks remaining on the same basis as before ('No change' to trend) but also noted a greater inevitability of loss compared to previous reviews. Additionally, the Council recommended that a received nomination for an entry, on unpublished digital indices and transcriptions in the DIMEV Open-Access Digital Edition of the Index of Middle English Verse, would provide a valuable example to this entry rather than as a new, standalone entry. The 2023 Council additionally recommended that the next major review considers rescoping the entry, possibly splitting this entry into separate areas to assess different levels of risk relating to published and unpublished source code in legacy research web collections.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

These collections are valuable but lose funding and care as institutions re-configure their tasks and individuals retreat from tasks due to retirement or (as volunteers) to old age.

There are an endless number of legacy research web resources out there that people don't know about.

Not necessarily a technical challenge but a resource challenge

The Internet Archive and other national web archiving bodies have copies of a lot of websites that would fit into this category but by no means all. There's also a distinction between the software or code used to deliver the user experience and the data. Such code is secondary to the content.

This issue can be intensified by the legacy IT Infrastructure in cases where much of the content is hosted there, as security concerns may lead to disposal of content imminently. In these scenarios, their imminence of action becomes more urgent given the security issues posed by hosting legacy technology/software/etc.

Case Studies & Examples

- The example of the British Library cyber incident as a case example of issues arising when working with legacy systems. See **Learning Lessons from the Cyber-attack: British Library**

cyber incident review, The British Library, (2024), *https://www.bl.uk/home/british-library-cyber-incident-review-8-march-2024.pdf/* [accessed at 2024-09-06].

- One example of an at-risk legacy research web collection, provided by the nominator of this entry, is the Unpublished digital indices and transcriptions in the DIMEV Open-Access, Digital Edition of the Index of Middle English Verse. The index comprises transcriptions made by a research team of Middle English text which were gathered as XML sheets and built upon a print publication: the Index of Middle English Verse (1943). These transcriptions involved significant financial and time investment and many are transcriptions of material unavailable online as digital facsimiles (uncertain data storage of the data that underlies the web resource, or whether it is being stored by a university or could easily be recovered). See **The DIMEV: An Open-Access, Digital Edition of the Index of Middle English Verse**, Mooney, L., Mosser, D. Solopova, E., Thorpe, D., Hill Radcliffe, D., Hatfield, L., Cornelius, I. and Johnston, M., *https://www.dimev.net/* [accessed at 2023-10-24].

- The recovery of the VecNet archive of malaria-related publications offers another example that also has obvious public health implications. VecNet was founded in 2011 as a network of institutions assembled to address the concerns and recommendations of the Malaria Eradication Research Agenda initiative. It became a portal for malaria information and analysis tools, with the goal of extending present vector control interventions and enabling incorporation of additional interventions to achieve elimination. By 2019 an important component of the portal, the DataCite repository, ceased to be available. However, the Vector-Borne Disease Network Data Warehouse (VecNet-DW), a project of departments of University of Notre Dame and the Institute of Tropical Health and Medicine at James Cook University, retained the relevant data and is collaborating with Data Futures, which created the new Invenio repository. See **VecNet**, Invenio, *https://vecnet.nd.hasdai.org/* [accessed at 2023-10-24].

- Preserving the Carmichael Watson Research Project website at the University of Edinburgh: a case study on this project website, only online from 2013 until 2018, came to imminent risk of permanent loss and the strategy undertaken to transform it into a more sustainable format through web archiving and to revive its public accessibility. See **Using ArchiveWeb.page to capture the Carmichael Watson Project**, Day Thomas, S. and Hawes, A. (2021, (2021), *Web Archiving & Preservation Working Group https://www.youtube.com/watch?v=0CWMwJn6p-w* [accessed at 2023-10-24].

- **Secure your digital datasets — by letting a data centre look after them!'**, Fellgett, M., (2021), *British Geological Survey Blogs https://www.bgs.ac.uk/news/secure-your-digital-datasets-by-letting-a-data-centre-look-after-them/* [accessed at 2023-10-24].

Manuals, Documentation, and Associated Materials

Manuals that support physical, born-digital, and hybrid-digital museum objects, including technical manuals and that may contain diagrams, images, videos and more. Manuals may include those specifically created for the object, and/or from the supplier or manufacturer (and

may exist for the entire object, or for each component part). Covered here is other documentation critical for the management, preservation, access, and display of collections objects, including conservation records. Associated materials (such as sidecar or other technical files) can provide insight into the digital content held in the collection. Encompassed in this entry is both information for internal use, supporting outgoing loans, as well as an information source for current and future researchers.

Categories

- Museum and Gallery

Examples

Manuals created for the object and/or from the supplier or manufacturer; conservation records and other forms of documentation.

Imminence	Effort
●●●○○	●●●●○
Action is recommended within three years, detailed assessment within one year.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Poor documentation; lack of preservation planning; held in proprietary file formats; uncertainty over IPR or the presence of orphaned works

Mitigations

Strong documentation; preservation capability; open source or commonly used file formats

Bit List History

Added to list: 2023

Trends

- **2024:** No change.

Last Review

2023 Review

This entry was added in 2019 under 'Digital Materials in Museums and Galleries' and previously rescoped in 2021 to 'Supporting Digital Materials for Museums and Galleries'.

The 2023 Bit List Council superseded the entry, splitting it into six discrete entries as the scope of the single entry was too broad to provide the guidance needed. The recommendation to break this entry down was also made by the 2021 Jury, as the types of digital collections content in museums can be vast and offer particular risks in museum and gallery contexts. For museum objects (including physical, born-digital, and hybrid-digital) a manual for the object may be critical to being able to provide access to, display, and/or maintain, regardless of whether the object is being retained in a

museum collection as a 'working' or 'non-working' object, which can be dependent on the types of materials the object is constructed from. For complex objects - physical and digital - there can be a significant dependency on these manuals and similar documentation. The extreme result of not having access to a manual may be that the object is unable to be maintained or displayed in the future. Associated materials may provide technical insight for both internal use and for researchers. For certain researchers, manuals and documentation may provide critical insight, and be of more value to the research than the object itself.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

While scores should remain as is currently, the Council recommends a more thorough investigation as to where these manuals and other documentation are stored, especially where museums don't have an Electronic Document and Records Management System (EDRMS) for managing born-digital and digitised internal corporate information and records.

For more complex digital objects held in museums, especially interactive, AR/VR/MR/XR etc., the range of documentation in addition to a tech manual may be vast, and so a more thorough review should be undertaken of this category in 2025, incorporating a wider range of museums that represent a wide range of sectors (not only art museums).

Case Studies & Examples

- **How Are Museums Using Artificial Intelligence?**, Styx, L. , (2024), *Museum Next* *https://www.museumnext.com/article/artificial-intelligence-and-the-future-of-museums/* [accessed at 2024-09-06].

Open Source Intelligence Sources of Current Conflicts

Open source intelligence produced, collected and analysed from publicly, openly available social media and web content with the purpose of answering a specific intelligence question and that supports crowd-sourced investigation and fact-checking to verify or refute claims of state agencies and rebel groups in the context of current political or military conflict.

Categories

- Legal Data

Examples

Social media sources relating to current conflicts, such as in Yemen or Syria

Imminence	Effort
●●●●○	●●●○○

Imminence	Effort
Action is recommended within twelve months. Detailed assessment is a priority.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Loss of authenticity; lack of preservation agency; limited or no digital preservation capability; Uncertainty over IPR or the presence of orphaned works

Mitigations

Offline backup captured by the journalist or investigating authority

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added as a subset in 2019, as part of a broader 'Open Source Intelligence Sources' which the Jury split into three elements, relating to current, recent and historic sources. This entry relates in particular to materials relating to current and ongoing conflicts. Social media companies have a policy to take down or suppress content that they consider to be propaganda for terrorist groups. This has had the unintended consequence of deleting or suppressing content that was being used in open source investigation or fact-checking for journalistic or judicial purposes, and which may therefore be an impediment to refutation or prosecution. However, a new generation of cloud-based services, such as Hunchly, have emerged in the last few years, which allow investigators to copy and stabilize content to private accounts in the process of investigating it: so, the ethical requirements of social media companies and the integrity of the investigation are both served. The 2021 Jury noted that such content stays at risk, and the process of investigation is slower than algorithmic deletion. Nonetheless, there is a notable difference in the investigation of current conflicts than historic ones where evidence has been lost. The 2022 Taskforce identified a trend towards even greater risk based on the increased significance of crowd-sourced investigations and fact-checking in light of ongoing global conflicts that include (but are not limited to) those in Ukraine.

The 2023 Council agreed with the Critically Endangered classification with the overall risks remaining on the same basis as before ('No change' to trend). The 2023 Council also added clarification to the meaning of 'open source' for this entry, to explain its meaning in relation to intelligence openly available online, noting that open source can also refer to a specific software or content licence that permits limited uses of IP so this distinction would be helpful for readers.

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

The Council acknowledge the continuing challenge of ensuring the preservation of complete and accurate resources given that: platform owners continue to be obliged to remove content that violates community standards; copyright and ownership increasingly hinders capture/preservation of the open source materials; and with the rise in fakes, preservationists must attend to standards for legal admissibility and authentication which vary from one jurisdiction to another.

Additional Information

Preservation is important for social context and may be picked up inadvertently in other ways - but is ambiguous about who has ultimate responsibility for collecting and preserving this.

Case Studies & Examples

- The Ukraine Investigations by GLAN and Bellingcat Justice & Accountability project to investigate alleged atrocity crimes taking place in Ukraine. The aim of the project is to conduct a set of open source investigations into incidents causing civilian harm occurring in Ukraine according to robust legal standards with the aim of making them available to national and international prosecutors who are gathering evidence of alleged crimes. In this case, the open source content gathered during Bellingcat's investigations will be preserved by Mnemonic, an independent third-party organization maintaining an archive of digital content from Ukraine, as it has done for Syria, Yemen and Sudan. See **Methodology for Online Open Source Investigations**, Glan and Bellingcat, *https://www.glanlaw.org/online-open-source-methodology* [accessed at 2023-10-24].

- The website of the Forensic Architecture (FA) research agency, based at Goldsmiths, University of London, offers examples of OSINT. See **Forensic Architecture**, *https://forensic-architecture.org/methodology/osint* [accessed at 2023-10-24].

- The website of the Coalition for Content Provenance and Authenticity (C2PA). The C2PA addresses the prevalence of misleading information online through the development of technical standards for certifying the source and history (or provenance) of media content. C2PA is a Joint Development Foundation project, formed through an alliance between Adobe, Arm, Intel, Microsoft and Truepic. See **Coalition for Content Provenance and Authenticity**, *https://c2pa.org/* [accessed at 2023-10-24].

- **Preserving Open Source Digital Evidence: A Guide for Practitioners Working on Dealing with the Past**, Baumhofer, E. and Reilly, B.F., (2022), *https://www.swisspeace.ch/articles/preserving-open-source-digital-evidence* [accessed at 2023-10-24].

- **Bellingcat and beyond. The future for Bellingcat and online open source investigation**, Higgins, E., (2019), *iPres Conference 2019 https://www.youtube.com/watch?v=kZAb7CVGmXM* [accessed at 2023-10-24].

- **Outlining a Human-Rights Based Approach to Digital Open Source Investigations**, Dubberley, S., and Ivens, G., (2022), *The Human Rights, Big Data and Technlgy Project http://repository.essex.ac.uk/32642/1/Outlining%20a%20Human-*

Unpublished Research Data

Data sets produced in the course of research but never shared or made available outside of the initial research team.

Categories

- Research Outputs

Examples

Unpublished research data can include different kinds of unpublished research data outputs, such as unstructured or structured datasets, databases, or other organized collections of computerized information or data such as periodical articles, books, graphics and multimedia

Imminence	Effort
●●●●○	●●●●○
Action is recommended within twelve months. Detailed assessment is a priority.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Originating researcher no longer active or changed research focus; staff on temporary contracts; dependence on single student or staff member; weak or fluid institutional commitment to subject matter; weak institutional commitment to data sharing; uncertainty over IPR or the presence of orphaned works; encryption; limited or dysfunctional data management planning.

Mitigations

Replication and documentation; data management plan; preservation pathway agreed.

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** Material improvement.

Last Review

2023 Review

This entry was added in 2019 as a subset of the 'Unpublished Research Outputs' reported in 2018, which was split into entries to draw attention to the different preservation requirements and

concerns that arise. This entry relates specifically to research data which has not been shared or published by any means and is thus in contravention of the 'FAIR' principles which require data to be Findable, Accessible, Interoperable and Reusable. Without proper planning, research data can have a high barrier to re-use, especially where documentation is lacking. The 2019 Jury took the view that documentation and re-use go hand in hand, and researchers should be under no illusions that data not documented or shared faces material and immediate risks of extinction. The 2020 Jury agreed with the assessment. The 2021 Jury identified a trend towards reduced risk in light of more robust collaborative initiatives to jointly address the risk of data loss in and across research communities. The 2022 Taskforce identified a trend towards even more reduced risk based on material improvement over the last year ('Material improvement' trend), which had not only offered examples of good research data management and preservation practices but also suggested a significant shift toward a culture of change and collaboration across different research communities and stakeholders. Those mentioned included (but were not limited to) improvements and initiatives by the European Open Science Cloud (EOSC), Science Europe, Research Data Alliance (RDA), Digital Curation Centre (DCC) and related projects on the preservation of research data and outputs.

The 2023 Council, in light of the **Trends** in 2021 and 2022, changed the classification from Practically Extinct to Critically Endangered, noting a positive trend of increased research data management activity and engagement by libraries, which should help to ensure that more research datasets are properly deposited in data repositories. They added that there was a general trend across many, if not most, HEI libraries producing research to do more in terms of research data management and a much larger part of what libraries do, with activities in this area growing and scaling up. However, the scale of unpublished datasets is hard to assess, as they are by definition unknown.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

If we do not know it exists, does it exist? It may also be that in certain circumstances this includes data that is unfavourable and has intentionally not been published. If perceived as high-value, someone in the research team will likely take steps to ensure it is protected. We can be proactive and offer advice, but ultimately it is down to them. We cannot keep everything!

This is a wide field, so the scale and impact are hard to describe, but the risk is higher than papers due to potential file format complexity.

Success is dependent on how successful an institution's research data management communications are. Advocacy and research are needed to show the scale of the problem, as well as education regarding open science and preservation.

Simply having a data management plan prepared is not sufficient, it needs to be properly implemented and kept up to date and relevant for both the researcher and the repository which will take a copy of the data. DMP should be used to appraise what data is worth long term preservation (e.g. NERC Data Value Check List), and what data is of lower quality, non-reusable, and even a reputational risk should it be shared further.

Unpublished Research Data from Government Researchers

Data sets and research outputs produced in the course of government research but never shared or made available outside of the initial research. In particular, the risk classification applies to research data under government embargo, restrictions due to sensitivities, classification issues, and/or materials suppressed for ideological reasons.

Categories

- Research Outputs

Examples

Data sets or research outputs produced for agencies that have closed or have had funding withdrawn from research initiatives; research data from government agencies no longer active

Imminence	Effort
●●●●○	●●●●○
Action is recommended within twelve months. Detailed assessment is a priority.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Lack of access to archival services; sudden or unanticipated closure; loss of implicit knowledge from destabilized or demoralized staff; encryption; uncertainty over IPR or the presence of orphaned works

Mitigations

Archival responsibility well developed; documentation; published through research channels

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 under 'Unpublished Research Data from US Government Researchers'. It has significant overlaps with other entries in the research outputs group but was set as a standalone entry to draw attention to two realities: 1. Research outputs are not simply a matter for academic institutions, and that government is, in fact, a major producer of research data; and 2.

Political instability and threats to the continuity of government services are a significant preservation risk. The 2019 Jury noted that while it specifically related to the US government context, it did not mean that other jurisdictions are immune from political instability, and commented that politically inconvenient research outputs face particular and immediate threats of which the digital preservation community should be cognizant. No 2020 trend towards increased or decreased risk was identified. The 2021 Jury agreed with concerns raised by the 2019 assessment but recommended the broader applicability should be more explicit. The entry title and description were changed to broaden and include governments across national and international contexts. The 2021 Jury added that with the changes, the risk profile will range and depend on the political system, the political change and the measures in place to save and reuse data from disbanded research projects; in other words, there may be instances where the unpublished research data in one country may fall under the Vulnerable classification. The 2022 Taskforce agreed with this 2021 assessment with no change to trend.

The 2023 Council agreed with the Critically Endangered classification with risks on the same basis as before ('No change' to trend) and recommended inviting additional review from an expert in this area for the next review. A further recommendation was made to consider whether it should be an individual entry or instead be incorporated as an example under the 'Unpublished Research Data' entry.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

The US made the news as part of the last government, but this is probably an issue in other countries as well and is, therefore, a category that could be made more generic. One question to ask is whether the research data is considered of long-term value or considered ephemeral?

Completed Investigations Based on Open Source Intelligence Sources

Open source social media and web content that has been used to support the conclusions of crowd-sourced investigation and fact-checking in political or military conflict.

Categories

- Legal Data

Examples

Social media sources relating to recent conflicts

Imminence	Effort
●●●●○	●●●○○

Imminence	Effort
Action is recommended within twelve months. Detailed assessment is a priority.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Encryption; loss of authenticity; lack of preservation agency; limited or no digital preservation capability; Uncertainty over IPR or the presence of orphaned works

Mitigations

Offline backup captured by a journalist or investigating authority; materials presented and documented in court; court able to deliver preservation; authenticity protected

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 and subsequently split into three elements by the Jury, relating to current, recent and historic sources. This entry relates to materials used in evidence in completed investigations, as well as those presented to courts or other investigatory agencies. Social media companies have a policy to take down or suppress content that they consider propaganda for terrorist groups. This has had the unintended consequence of deleting or suppressing content used in open source investigation or fact-checking for journalistic or judicial purposes, which may impede refutation or prosecution. However, a new generation of cloud-based services now allows investigators to copy and stabilize content to private accounts in the process of investigating it, so the ethical requirements of social media companies and the integrity of the investigation are both served. The 2020 Jury noted that such content remains at risk. The presentation of data to a court or prosecuting authority, or the publication through news media, implies the introduction of a long-term preservation function. The 2021 Jury agreed with this assessment and Endangered classification but changed the 2021 trend towards greater risk in light of recent developments in crowd-sourced investigations and fact-checking. The 2022 Taskforce agreed on a trend towards even greater risk based on the increased significance of crowd-sourced investigations and fact-checking in light of ongoing global conflicts that include (but are not limited to) those in Ukraine.

The 2023 Council agreed with the Endangered classification with the overall risks remaining on the same basis as before ('No change' to trend), noting that some of the description and language used in the entry may be confusing. For example, one might think if investigations are complete that surely the parts used as evidence or published in articles are captured elsewhere and presumably

preserved there? While this may be the case, presuming long term preservation may lead to future loss in instances where this is not true. Here, risks can overlap with those found in 'Evidence in Court' and 'Proceedings in Court' entries, which themselves indicate that standard records management processes within designated agencies should be able to take care of the preservation of materials like this but, given that it is likely to involve complex types of data, such agencies may not be equipped to deliver preservation effectively.

The 2023 Council additionally recommended further scoping of the entry with input from those working with these materials directly, to help explain the unique challenges as well as examples where content has been lost due to deletion by social media companies and/or legal retention periods where certain content may not fall under the scope of records for long-term or permanent retention.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

The Council noted the work by the Syrian Archive, WITNESS, Data Cívica, Amnesty International's Citizen Evidence Lab, the Human Rights Data Analysis Group, Mnemonic and UC Berkeley Human Rights Center potentially contributes specific preservation strategies and technologies for open source intelligence.

Case Studies & Examples

- The 2020 Medical Facilities Under Fire report by the Syrian Archive., which provides information on how The Syrian Archive and its partners (Syrians for Truth and Justice, Justice for Life) analyzed and verified pattern of attacks by cross referencing a combination of open-source visual content, flight observation data, and witness statements. Through collecting, verifying and reporting investigative findings from these incidents, the authors hope to preserve critical information that may be used for advocacy purposes or as evidence in future proceedings seeking legal accountability. See **Medical Facilities Under Fire**, Syrian Archive, (2017), *https://syrianarchive.org/en/investigations/Medical-Facilities-Under-Fire* [accessed at 2023-10-24].

- **Bellingcat and beyond. The future for Bellingcat and online open source investigation**, Higgins, E., (2015), *iPres Conference 2019, https://www.youtube.com/watch?v=kZAb7CVGmXM* [accessed at 2023-10-24].

- European Human Rights Advocacy Centre (EHRAC) blog post providing a Q&A on 'Using Open Source Investigations in Human Rights Litigation,' noting their approach to gather and present evidence of Russian military presence in and around Ilovaisk in August 2014. See **Q&A: USING OPEN SOURCE INVESTIGATIONS IN HUMAN RIGHTS LITIGATION**, European Human Rights Advocacy Centre (EHRAC), (2022), *https://ehrac.org.uk/en_gb/blog/qa-using-open-source-investigations-in-human-rights-litigation/* [accessed at 2023-10-24].

- **Preserving Open Source Digital Evidence a Guide for Practitioners Working on Dealing with the Past Essential**, Baumhofer, E., and Reilly, B., (2022), *Swiss Peace https://www.swisspeace.ch/assets/publications/Essentials/Essentials20221031_Preserving-Open-Source-Digital-Evidence.pdf* [accessed at 2024-10-07].

Interpretive Materials

Text, graphics, videos, and other content that is used in gallery and exhibition spaces to guide audiences and provide learning experiences.

Categories

- Museum and Gallery

- Social Media

Examples

These may include explanatory text and diagrams. This type of content typically sits alongside objects being exhibited. This information may also be used in other forms of digital public engagement, including within websites and social media

Imminence	Effort
●●●○○	●●○○○
Action is recommended within three years, detailed assessment within one year.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Poor or no documentation; lack of repository and preservation infrastructure; external dependencies; poor storage; significant volumes or diversity of data; digital content being left unmanaged with Exhibitions teams (e.g., not going through any Records Management, Recordkeeping, being passed to the archives, or having preservation even considered); uncertainty over IPR or the presence of orphaned works

Mitigations

Strong documentation; preservation capability; good recordkeeping principles, practices, and management in place; strong repository and preservation technical infrastructure; well-developed migration pathways

Bit List History

Added to list: 2023

Trends

- **2024:** Towards greater risk.

Last Review

2023 Review

This entry was added in 2019 under 'Digital Materials in Museums and Galleries' and previously rescoped in 2021 to 'Supporting Digital Materials for Museums and Galleries'.

The 2023 Bit List Council superseded the entry, splitting it into six discrete entries as the scope of the single entry was too broad to provide the guidance needed. The recommendation to break this entry down was also made by the 2021 Jury, as the types of digital collections content in museums can be vast and offer particular risks in museum and gallery contexts. This entry was added to focus on risks tied to interpretive materials in museums and galleries. These materials may be tailored to various audience demographics and may take into account a variety of learning styles. These materials can also provide significant insight into how display and interpretation of objects has changed over time. While some interpretive materials may be text or still image based, others may be more technically complex as they may be time-based (e.g., video) or even a complex digital object in its own right.

2024 Interim Review

The 2024 Council identified a trend towards even greater risk. The species group noted this content may not be seen to be of long-term value to the museum, however, may be of value to the researcher and, in light of the 2023 British Library Cybersecurity incident, the Imminence of Action was changed from five years to less than three years.

To add further clarity, those in the Social Media added that the description for this entry seems to address two separate types of digital content: 1) interpretive materials used to support users' experiences of an exhibition, and 2) museum and gallery public engagement via social media. With that they have recommended that Council or species reviewers should discuss if a separate Museum (or even Collecting Institution) public engagement on social media entry should be created and sit in Social Media rather than individual organisation types.

Additional Information

Interpretive outputs through websites may have reduced risk as web outputs if captured as part of national web archiving initiatives.

Case Studies & Examples

- The mentioned British Library cyber incident as a case example of issues arising. See **The British Library**, Learning Lessons from the Cyber-attack: British Library cyber incident review, (2024), *https://www.bl.uk/home/british-library-cyber-incident-review-8-march-2024.pdf/* [accessed at 2024-09-06].

- **How Are Museums Using Artificial Intelligence?**, Styx, L. , (2024), *Museum Next* *https://www.museumnext.com/article/artificial-intelligence-and-the-future-of-museums/* [accessed at 2024-09-06].

Research Materials and Outputs in Museums and Galleries

Digital material used in, or resulting from, research carried out on materials, digital or otherwise, held in galleries, museums, or similar. Research outcomes may not be formally published, and supporting datasets may not be formally accessioned or archived by an organization or a related organization. Access to these research materials and outcomes may only be made available for internal use, to inform other public outcomes, or for individual researchers.

Categories

- Museum and Gallery

- Research Outputs

Examples

Unpublished or published research papers, datasets, databases and other supplementary materials

Imminence	Effort
●●●○○	●●●●○
Action is recommended within three years, detailed assessment within one year.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Lack of documentation; uncertainty over IPR or the presence of orphaned works; unstable funding for repository; external dependencies.

Mitigations

Strong data management planning; preservation capability; good documentation; deposit into trusted repository.

Bit List History

Added to list: 2023

Trends

- **2024:** No change.

Last Review

2023 Review

This entry was added in 2019 under 'Digital Materials in Museums and Galleries' and previously rescoped in 2021 to 'Supporting Digital Materials for Museums and Galleries'.

The 2023 Bit List Council superseded the entry, splitting it into six discrete entries as the scope of the single entry was too broad to provide the guidance needed. The recommendation to break this entry down was also made by the 2021 Jury, as the types of digital collections content in museums can be vast and offer particular risks in museum and gallery contexts. This entry on Research Materials and Outputs within the scope of Museums and Galleries differs from the ones found in Research Outputs, with the latter focus around institutional supporting higher education institutions but lacking for museums and gallery contexts. They agreed with the 2021 Jury Review recommendations that Museum & Gallery entries require further rescoping. In regards to this entry, the 2023 Council recommended that a future review should further rescope of Oral Histories and Research Materials and Outputs due to overlaps/cross referencing which, due to time constraints, was unable to be done for the *2023 Review* cycle.

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

While there are no changes to the risk trend, they note an influencing factor over the past year is the potential shift in funder requirements around Open Access, and sometimes preservation. With that in mind, they recommend a full international review in 2025 to see globally where funder requirements are placed for both preservation and (long-term) Open Access

Additional Information

To add further clarity, it might be worth differentiating use cases—for research outputs as digital material resulting from research carried out in Museums and Galleries, and for digital material used in research.

This research may be publicly or philanthropically funded. While research materials - used and/or developed in the course or research - and research outputs may not be made publicly available, they may be used to inform other outputs, e.g. exhibition, interpretation, conservation, etc.

Exhibition catalogues and interpretation of collections are often published online in research papers.

Semi-Published Research Data

Data sets produced in the course of research and shared between researchers, such as by posting to a website or portal but without preservation capability or commitment. Typically the data remains in the hands of the researchers who have the job of maintaining it.

Categories

•   Research Outputs

Examples

Departmental web servers; project wikis; GitHub repositories

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Originating researcher no longer active or changed research focus; staff on temporary contracts; dependence on single student or staff member; weak or fluid institutional commitment to subject matter; weak institutional commitment to data sharing; uncertainty over IPR or the presence of

orphaned works; encryption; limited or dysfunctional data management planning; web capture challenges that means unlikely to be picked up by automatic crawlers

Mitigations

Data in preparation for transfer to specialist repository; robust data management planning; documented and managed professionally using data stewards

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This 2019 entry was previously introduced in 2017 under 'Research Data,' though without explicit reference to semi-published research data. The 2019 Jury split the 'Research Data' entry into a range of contexts for research outputs, including this addition. The entry draws attention to represent 'self-help' data sharing which is to be encouraged as a means to facilitate open science but should not be confused with long-term preservation. The 2021 Jury agreed with the Endangered classification, noting problems with the volume of data being produced but not being kept in a meaningful way. They noted that research data is complex and has specific requirements for documentation that may only be known to subject matter experts; however, data creators (e.g., researchers) are not necessarily well placed to sustain the data in the long term.

There were also a few significant changes to the entry in the 2021 Bit List.

Removal of 'informally' from the previous entry description ('shared informally between researchers') due to possible misperception or misunderstanding; 'informal' may imply researchers would perceive the data as low value and not want it captured. This may be the case, so it is important to consider and provide advice to researchers who think there is value in their data.

Two previous entries (Geomagnetic Data and Software and Maritime Archaeological Archives) have been removed as separate entries and incorporated into this broader entry on semi-published research data to highlight the range of content and forms semi-research data can take and highlight the need for specialized knowledge and specialist repositories for preparing and managing the data throughout the lifecycle

The 2021 trend towards reduced risk, which was based on improvements and initiatives towards the preservation of semi-published research data since the entry's addition in 2019.

The 2022 Taskforce agreed on a trend towards reduced risk based on material improvement over the last year which had not only offered examples of good research data management and preservation practices but also suggested a significant shift towards a culture of change and collaboration across different research communities and stakeholders. Those mentioned included (but were not limited to) improvements and initiatives by the European Open Science Cloud (EOSC), Science Europe,

Research Data Alliance (RDA), Digital Curation Centre (DCC) and related projects on the preservation of research data and outputs.

The 2023 Council agreed with the Endangered classification and that risk remained on the same basis as before ('No change' to trend).

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

There is a positive trend of increased research data management activity and engagement by libraries and data centres, which should help to ensure that more research datasets are properly deposited in data repositories, rather than left in a 'semi-published' state.

Offering and minting researchers Digital Object Identifiers for datasets deposited at specialist repositories will encourage data citation and increase research impact of individual researchers, which traditionally relied more on publishing papers than datasets.

Case Studies & Examples

- **Practical Guide to Sustainable Research Data: Maturity Matrices for Research Funding Organisations, Research Performing Organisations, and Research Data Infrastructures**, Boccali, T., Sølsnes, A., Thorley, M., Winkler-Nees, S. and Timmermann, M. , (2021), *Science Europe* http://doi.org/10.5281/zenodo.4769703.

- **Development and outputs of the European Open Science Cloud (EOSC) Long-Term Data Preservation Task Force**, European Open Science Cloud (EOSC) , https://www.eosc.eu/advisory-groups/long-term-data-preservation [accessed at 2023-10-24].

Non-standard Public Records

Records created in the course of public administration and subject to public records legislation but created on unofficial channels and platforms and therefore subject to unlawful destruction whether by accident or design.

Categories

- Legal Data

Examples

Content and messages from cloud-based instant messaging services (such as WhatsApp or Snapchat) that pertain to public administration and are subject to public records legislation but concealed from or inaccessible to archival agencies

Imminence	Effort
●●●●○	●●●○○
Action is recommended within twelve months. Detailed assessment is a priority.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

N/A

Mitigations

Lack of skills, commitment or policy from corporate owners; Business models in conflict with principle of preservation; Archival pathway; public officials briefed on the nature of public records and the penalties for illegal disposal; boundary between public and private correspondence; cloud services administered transparently; export functions; ; Uncertainty over IPR or the presence of orphaned works

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 as a subset of an entry in 2018 for 'Digital Legal Records and Evidence,' which the Jury split into four different entries in order to draw attention to the different challenges and priorities that arise. The 2019 Jury gave this entry the strongest indication of risk available. This group includes those records which may contain politically damaging or uncomfortable realities and thus be at risk of deletion and may be concealed from archival agencies whether by accident or design. The 2019 Jury also noted that the destruction of certain classes of public records is unlawful, whether or not it is deliberate.

The 2020 Jury added the trend towards greater risk based on the 'pivot to digital' necessitated by the Pandemic resulting in widespread changes in workflow and in the platform for the delivery of government, with significant amounts of remote working. These changes happened rapidly, often without time to consider the preservation and record keeping implications. In those circumstances, it was reasonable to suppose the risks expanded in size as well as scope.

The 2021 Jury agreed but found no significant increase or decrease to the trend. They added that there should be a balance between trying to preserve what has already been created using these channels and trying to educate against/prevent records from being created this way in the future. The 2022 Taskforce found no significant trend towards even greater or reduced risk.

The 2023 Council agreed with the Practically Extinct classification and noted a slight decrease in the effort needed to preserve and the imminence of action required when compared to the 2021 Jury review, suggesting that overall risks remain on the same basis as before with major efforts needed but loss not entirely inevitable.

2024 Interim Review

The 2024 Council agreed These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

While scores should remain as is currently, the Council recommended the entry be revisited in the next review to consider whether it is better suited under the Public Records digital species instead of Legal Data.

Additional Information

Even records created on official channels and platforms are potentially being subjected to unlawful destruction through subversion of official formal processes.

This is a 'small effort to fix' in terms of the technology to export data. But loss seems likely unless there is stronger monitoring and enforcement of the policy around this.

Agencies responsible for the public record will not be able to completely control their public servants' use of unofficial channels (but could tighten), so they need methods to obtain from unofficial channels. Very important for public accountability and transparency of the state.

Obviously, these records should not be created using these channels in the first place. It is probably unlawful for those in public office. However, we know there has and always will be this kind of backdoor activity and pretty much always a scandal when it is revealed. The challenge is managing to collect it.

Maybe at the time of creating this entry, the Jury was thinking of legal data examples, but at this point, in time, I think the species best belongs under "Public Records" and so should be moved there.

Published Research Data Appended to Journal Articles

Closed research data sets produced and documented in accordance with good practice and appended to a journal article or transferred to a repository that does not have sufficient subject-matter expertise or funding commitment to ensure reliable or ongoing preservation for the long term.

Categories

- Research Outputs

Examples

Supplementary data sets added to formally published papers in repositories that are designed primarily for papers; electronic journals offering data sets without obvious preservation capacity;

institutional repositories servicing highly complex scientific data sets with insufficient subject-matter expertise

Imminence	Effort
●●●○○	●●○○○
Action is recommended within three years, detailed assessment within one year.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Complex mix of formats; deposit in repositories that lack relevant expertise or knowledge or funding; poorly designed migration or normalization processes; poorly formed ingest and quality assurance procedures; rapid churn of staff; incoherent patterns of subject matter; lack of domain knowledge; no or very small numbers of users; weak or absent collecting policy; deposit to ensure minimal compliance with funder mandate; limited or dysfunctional data management planning and documentation; uncertainty over IPR or the presence of orphaned works

Mitigations

Clear data management planning and documentation; deposit by publisher in a trusted repository; deposit by author/s in appropriate repositories with digital preservation expertise and mandate; clear licensing to enable digital preservation and access; strong user base; development roadmap; ability to transfer collections or share metadata with subject repositories or portals; demonstrable re-use of data; clear collecting policy; data management planning early in the data lifecycle

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** Material improvement.

Last Review

2023 Review

This 2019 entry was previously introduced in 2017 under 'Published Research Outputs,' though without explicit reference to the research data appended to journal articles. The 2019 Jury split the entry into a range of contexts for research outputs, including this addition and 'Research Data Published through Repositories'. The entry draws attention to services that take upon themselves commitments to preserve research data, but which may not deliver those promises through lack of capability. The 2021 Jury agreed with the Endangered classification but commented on the improvements and initiatives towards the preservation of research data outputs, with good practice documentation and replication in this space (e.g., collaborations with publishers and repositories, LOCKSS, CLOCKS, etc.). For these reasons, the 2021 trend was towards reduced risk.

The 2022 Taskforce agreed on a trend towards reduced risk based on material improvement over the last year that had not only offered examples of good research data management and preservation practices but also suggested a significant shift towards a culture of change and collaboration across different research communities and stakeholders. Those mentioned included (but were not limited to) improvements and initiatives by the European Open Science Cloud (EOSC), Science Europe, Research Data Alliance (RDA), Digital Curation Centre (DCC) and related projects on the preservation of research data and outputs.

In light of the identified 2021 and 2022 **Trends**, the 2023 Council changed the classification from Endangered to Vulnerable. They noted that many, if not most, HEI libraries that produce research are doing more in terms of research data management, and the activities in this area are growing and scaling up. Due to increased focus on this area, it was recommended that the classification change to Vulnerable with 2023 trend of 'Material improvement'.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

A Council member recommended that, to add further clarity, it might be worth differentiating use cases—for closed research data sets produced and documented in accordance with good practice and appended to a journal article, and for closed research data sets produced and documented in accordance with good practice and transferred to a repository that does not have sufficient subject-matter expertise or funding commitment to ensure reliable or ongoing preservation for the long term.

Additional Information

A number of aggravating conditions—those relating to poorly formed ingest and quality assurance procedures, rapid churn of staff, incoherent patterns of subject matter, lack of domain knowledge, no or very small numbers of users, weak or absent collecting policy, and deposit to ensure minimal compliance with funder mandate—are problems with some repositories, not all repositories.

Presenting different use cases can tease apart the use case for supplementary materials appended to journals (e.g., which CLOCKSS and Portico preserve) and those in repositories that are perhaps not tailored for this use case. Cases where data is transferred to a repository that does not have sufficient subject-matter expertise or funding commitment to ensure reliable or ongoing preservation for the long term are far more at risk.

Research data is complex and has specific requirements for documentation which may only be known to subject matter experts. However well intended, it is risky for institutions to attempt to replicate that level of expertise across all the domains within the institution, and it can be hard for smaller publishers to make commitments to sustain data in the long term.

The loss of tools, data or services within this group would impact on people and sectors around the world. Particularly those involved with reproducibility and those wishing to use the datasets for further research.

Although there have been improvements in current practice, policies and workflows, there is still a significant corpus of information that was deposited before these improvements came into force. It is unlikely that there will be the time, will or resource to bring this information up to current standards.

UK funders e.g. UKRI-NERC Environmental Data Service are educating researchers about data policies which mandate depositing master and raw data at the funder disciplinary repository. These repositories have a strong expertise in the research discipline ensuring data and metadata standardization and quality assurance. Any copies of datasets published in journal articles or similar are considered secondary copies and do not comply with data policy, hence risking obtaining future research funding by the institute attempting to use journal outputs as their funder-acknowledged datasets.

The significance and impact of this entry specifically depends on whether it is the only copy of the dataset in existence, or whether there is another copy hosted in a data repository.

Case Studies & Examples

- Analysis from Martin Eve of CrossRef shows scholarly content at risk. The findings, based on the assessment of around 7.5 million of the e-books and articles for which CrossRef provides a fixed identifier or Digital Object Identifier, suggest that around a quarter of academic publications are not being preserved for the future. For c. 2 million articles in the study there were no evidence of them being preserved, and 4.3 of works studied were preserved in at least one place. See **Digital Scholarly Journals Are Poorly Preserved: A Study of 7 Million Articles**, Eve, M. P., (2024), *Journal of Librarianship and Scholarly Communication* *https://doi.org/10.31274/jlsc.16288*.

- The FAIRsharing Collaboration with DataCite and Publishers. See **FAIRsharing Collaboration with DataCite and Publishers: Data Repository Selection, Criteria That Matter**, McQuilton, P., Sansone, S.A., Cousijn, H., Cannon, M., Chan, W.M., Carnevale, I., Cranston, I., Edmunds, S., Everitt, N. and Ganley, E., (2019), *https://doi.org/10.17605/OSF.IO/N9QJ7*.

- Resources and research outputs from the Enhancing Services to Preserve New Forms of Scholarship project, which examined a variety of enhanced eBooks and identified which features can be preserved at scale using tools currently available. Of particular note is the published guidelines for preserving new forms of scholarship. See **Guidelines for Preserving New Forms of Scholarship**, Greenberg, J., Hanson, K., & Verhoff, D, (2021), *NYU Libraries* *https://doi.org/10.33682/221c-b2xj*.

- The work by the Centre pour la Communication Scientifique Directe (CCSD) of France and the Confederation of Open Access Repositories (COAR) in creating a preprint repository directory which has been relevant to building a user community). See **Directory of Open Access Preprint Repositories**, Centre pour la Communication Scientifique Directe (CCSD) of France and the Confederation of Open Access Repositories (COAR), *https://doapr.coar-repositories.org/* [accessed at 2023-10-24].

Research Data Published through Repositories

Research data published through digital repositories or other services providers with specialist skills to manage the data and an ongoing commitment to ensure preservation.

Categories

- Research Outputs

Examples

Recognized data repositories in specialist disciplines; institutional data repositories in subject specialist centres and partnerships.

Imminence	Effort
●●●○○	●●○○○
Action is recommended within three years, detailed assessment within one year.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Lack of long-term commitment; lack of user community; lack of visibility to potential depositors; lack of institutional commitment; insufficient documentation; uncertainty over IPR or the presence of orphaned works

Mitigations

Certification and documented good practice; effective documentation requirements for depositors; proven financial sustainability; skilled staff including professionalising disciplinary and general data stewardship offering a clear career option; participation in the digital preservation community; research data management training by repositories and research funders offered to depositors, in particular new career researchers

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** Material improvement.

Last Review

2023 Review

This entry was added in 2019 as a separate entry, but it was previously introduced in 2017 under 'Published research outputs,' though without explicit reference to the capacity of the repository infrastructure. The 2019 Jury split the entry into a range of contexts for research outputs, including this addition classified as Vulnerable; the preservation of research data published through a well-founded repository with the capacity and commitment to ensure preservation and capability through their own professional development activities made it a lower risk outcome for research data. The 2021 Jury agreed with this classification but commented on the improvements and initiatives towards the preservation of research data and outputs, leading to a 2021 trend towards reduced risk. The 2022 Taskforce identified a 2022 trend towards reduced risk based on material improvement over the last year that had not only offered examples of good research data management and preservation practices but also suggested a significant shift towards a culture of change and collaboration across different research communities and stakeholders. Those mentioned

included (but were not limited to) improvements and initiatives by the European Open Science Cloud (EOSC), Science Europe, Research Data Alliance (RDA), Digital Curation Centre (DCC) and related projects on the preservation of research data and outputs.

The 2023 Council agreed with the Vulnerable classification and noted that there was a trend towards reduced risk due to increasing research data management and engagement activity by libraries, which should result in increasing amounts of datasets being deposited. The 2023 Council also noted it would be useful to see empirical data on depositing **Trends** to assess this.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

A key consideration with this entry is whether the data repository is integrated with a preservation system to facilitate long term access and usability of datasets.

The loss of tools, data or services within this group would impact on people and sectors around the world. Particularly those involved with reproducibility and those wishing to use the datasets for further research.

Although there have been improvements in current practice, policies and workflows, there is still a significant corpus of information that was deposited before these improvements came into force. It is unlikely that there will be the time, will or resources to bring this information up to current standards.

Creating additional preservation metadata to research data holdings may help render data more robust in the long term, where using a preservation system is not an option. With an emphasis on environmental sustainability, some repositories hesitate mandating additional copies of large datasets which may be in the region of hundreds of terabytes, as this adds to both storage cost and carbon footprint, especially when capturing and preserving the research methodology would enable recreating the dataset.

Case Studies & Examples

- Research outputs from the ARCHIVER project. This includes findings from Early Adopter Use Cases. See **Early Adopters Use Cases**, ARCHIVER, *https://www.archiver-project.eu/early-adopters-use-cases* [accessed at 2023-10-24].

- Research outputs from the FAIRsFAIR project. This includes findings from Implementation & Adoption Stories. See **Implementation & Adoption Stories**, FAIRsFAIR , *https://fairsfair.eu/implementation-adoption-stories* [accessed at 2023-10-24].

- A recent analysis from Martin Eve of CrossRef shows scholarly content at risk. The findings, based on the assessment of around 7.5 million of the e-books and articles for which CrossRef provides a fixed identifier or Digital Object Identifier, suggest that around a quarter of academic publications are not being preserved for the future. For c. 2 million articles in the study there were no evidence of them being preserved, and 4.3 of works studied were preserved in at least one place. See **Digital Scholarly Journals Are Poorly Preserved: A Study of 7 Million Articles**, Eve, M. P. , (2024), *Journal of Librarianship and Scholarly Communication* *https://doi.org/10.31274/jlsc.16288*.

- **Disappearing repositories – taking an infrastructure perspective on the long-term availability of research data**, Strecker, D., Pampel, H., Schabinger, R. & Weisweiler, N.L., (2023), *https://doi.org/10.48550/arXiv.2310.06712*.

- **FAIR + Time: Preservation for a Designated Community (01.00)**, L'Hours, H., Kleemola, M., von Stein, I., van Horik, R., Herterich, P., Davidson, J., Rouchon, O., Mokrane, M., & Huber, R., (2021), *https://doi.org/10.5281/zenodo.4783116*.

- **Practical Guide to Sustainable Research Data: Maturity Matrices for Research Funding Organisations, Research Performing Organisations, and Research Data Infrastructures**, Science Europe, (2021), *https://www.scienceeurope.org/media/b3odxx3s/sepractical-guide-sustainable-research-data.pdf* [accessed at 2023-10-24].

- **Development and outputs of the European Open Science Cloud (EOSC) Long-Term Data Preservation Task Force**, European Open Science Cloud (EOSC) , *https://www.eosc.eu/advisory-groups/long-term-data-preservation* [accessed at 2023-10-24].

---

ii.   Complexity Threats

a)   Complex Platforms

Native Cloud Formats

This entry covers all native cloud formats that exist within a cloud system but cannot be exported in their native format. The data for these formats is held within the system and they are rendered within a browser.

Categories

- Formats

- Cloud

Examples

The most widely known example of this is Google formats such as Google Docs, Sheets, Slides and Jamboard.

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of skills, commitment or policy from corporate owners; dependence on proprietary products or formats; lack of export functionality; insufficient documentation; lack of conformance or validation; lack of preservation commitment or planning; inaccessibility to automated web crawlers; uncertainty over IPR or the presence of orphaned works.

Mitigations

Reduction of dependencies; improved export functionality; clear migration pathways; application of records management standards; version control; integrity checking; comprehensive documentation; access to web harvesting; technology watch.

Bit List History

Added to list: 2023

Trends

- **2024:** No change.

Last Review

This was a new Bit List entry added in 2023 to draw attention to the particular challenges of preserving native cloud content that cannot be exported and preserved in their native cloud formats. While there are some similarities with the 'Cloud-based Services and Communications Platforms' entry risks relating to dependencies on service and provider business models and the terms and conditions imposed, this entry focuses specifically on the distinct risks relating to preservation of digital content and data in native cloud formats (with these formats held within cloud-based systems and rendered within web-based browsers). Currently, in order to view the files outside of the system, an export format has to be chosen (e.g., PDF, Microsoft Office, HTML). This process has issues with proving the integrity of the exports, as conventional methods (such as checksum) are not valid. There is also the issue that the original cloud formats hold all edits and versions, the export may only preserve the current version of the file without edit history and misleading revision identifiers. As the cloud formats are browser-based, web archiving options have also been explored, but there is no current automated way to harvest a large collection of files. For these reasons, major efforts are needed to develop new tools and techniques to capture and preserve the data to prevent or reduce loss.

The *2024 Interim Review* concluded that these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Case Studies & Examples

- Case studies demonstrating good progress in this area, for example from the TNA and University of Sheffield,shared as part of a DPC event on 14th November 2023. See **Does it have to be this hard? Preserving content from Microsoft 365 and Google Workspace**, Beking, A., Hooper, B., Leming, R., Tilbury, J., Oeyen, Q., Young, P. and Gardner, R., (2023), *Digital Preservation Coalition Event* *https://www.dpconline.org/events/previous-events/eventdetail/221/-/does-it-have-to-be-this-hard-preserving-content-from-microsoft-365-and-google-workspace* [accessed at 2024-09-06].

- One example, which is part of the Google Workspace and illustrates how quickly things can become unsupported, is the announced closure of the 'Jamboard' collaborative online whiteboard platform, which will be discontinued from the end of 2024. See **Google's whiteboarding app is joining the graveyard**, Shakir, U., (2023), *Verge* *https://www.theverge.com/2023/9/28/23894509/google-jamboard-whiteboarding-app-graveyard* [accessed at 2023-10-24].

- **Google Jamboard is winding down**, Google, *Google Jamboard Help Center* *https://support.google.com/jamboard/answer/14084927?hl=en* [accessed at 2023-10-24].

- **How Can We Preserve Google Documents?**, Mitcham, J., (2017), *Digital Archiving at the University of York* *http://digital-archiving.blogspot.com/2017/04/how-can-we-preserve-google-documents_35.html* [accessed at 2023-10-24].

- **Preserving Google Drive What about Google Sheets?**, Mitcham, J., (2017), *Digital Archiving at the University of York* *http://digital-archiving.blogspot.com/2017/11/preserving-google-drive-what-about_64.html* [accessed at 2023-10-24].

- **PDF/A and read-only in SharePoint**, Pinsent, E., (2017), *Digital Preservation at UoL* *https://web.archive.org/web/20180923180801/https:/dart.blogs.ulcc.ac.uk/*.

- **Forensic Analysis of Cloud-Native Artifacts**, Roussev, V. and Mcculley, S., (2016), *Digital Investigation Vol. 16* *https://doi.org/10.1016/j.diin.2016.01.013*.

- **How I Reverse Engineered Google Docs To Play Back Any Document's Keystrokes.**, Somers, J., (2014), *https://features.jsomers.net/how-i-reverse-engineered-google-docs/* [accessed at 2023-10-24].

- **What's up (with Google) Docs? – The Challenge of Native Cloud Formats**, Young, P., (2021), *Digital Preservation Coalition Blog* *https://www.dpconline.org/blog/whats-up-with-google-docs* [accessed at 2023-10-23].

Cloud-based Services and Communications Platforms

Digital content produced, stored and accessed within commercial cloud-based services and communications platforms. This entry broadly includes services based on a costed subscription and contract business models, premium or institutional versions, and also free online utilities offered at no cost to end-users but with a business model based on gathering and reselling consumer insights.

Categories

- Cloud

Examples

Google services such as Drive, Docs, and Sheets; Microsoft services such as Sharepoint and Teams; Slack, Prezi, Yammer, DropBox.

Imminence	Effort
●●●●○	●●●●○
Action is recommended within twelve months. Detailed assessment is a priority.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Unstable business model from service providers; abandonment of the service due to various reasons (e.g., service provider bought over or pivots to new market opportunities); lack of export functionality; unstable terms and conditions; lack of onsite copy of key media; lack of investment in infrastructure; lack of strategic plan for IT provision; uncertainty over IPR or the presence of orphaned works; conflating preservation and access.

Mitigations

Clear export and migration pathways; preservation responsibility shouldered by the service provider; offline backup for key media; fit to preservation and records management plan; strategic roadmap for adoption.

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

This entry was added in 2021 as a merging of two separate 2019 entries, 'Consumer Cloud-based Utilities' and 'Premium or Institutional Social Media' to place emphasis on the similarities and common threats faced by services that are both 'paid-for' and 'free-at-the-point-of-use', namely similar aggravating conditions relating to increasing dependencies on the vendor's business models and the terms and conditions imposed. The 2021 Jury also noted a trend towards increased risk in light of greater reliance on the cloud and localized disruptions to cloud services over the pandemic and wider (global) dependence on these services, especially Google Drive, for record-keeping and business workflows. The 2022 Taskforce agreed with the previous assessment, with no change to trend.

The 2023 Council agreed with the Endangered classification with the overall risks remaining on the same basis as before ('No change' to 2023 trend), but also noted increases in imminence and effort to preserve, recognizing that while the need for major efforts to prevent or reduce losses continues, it is now much more likely that loss of material has already occurred, and will continue to do so, by the time tools or techniques have been developed.

This entry was previously categorized under the Social Media digital species. The 2023 Council adjusted this and other social media entries in light of how web-based and cloud-based business products and services had developed in recent years. This included:

Narrowing the scope. The scope of the entry was narrowed to focus more specifically on the various risks associated with digital content created, stored and shared using cloud-based services, especially business-related tools or collaboration tools which are not being well preserved (e.g. Slack, Google Drive, Sharepoint). These challenges primarily relate to services using their own cloud-based format, export functionality and quality.

As part of this rescoping, relevant information concerning cloud-based aspects was incorporated from the previous 'Born Digital Photographs and Video Shared via Social Media Platforms' and 'Consumer Social Media Free at the Point of Use', and this entry also now falls under a new 'Cloud' species group to more clearly differentiate between social media and cloud services–the Council adopts the view that just because a service is web-based and allows users to upload content for cloud hosting does not necessarily mean that it is 'social' or 'media'.

The 2024 Council agreed these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

Most platforms allow users to export their own data from them, which is helpful.

Similar to the 'Born Digital Photos and Video Shared on Social Media' entry, significance and impact scores are high because some users exclusively create and store important content on these services but uses for these services vary greatly. Also, subscription services, such as Microsoft Teams, though far from having adequate preservation provision, will have more robust back-up and recovery governed by institutional contracts, whereas Google Drive / Google 'office' services free-at-the-point of use do not provide these mitigating measures.

Dropbox is a content hosting/storage service and does support downloading a file the same quality as the file uploaded. If any one of these platforms disappeared overnight or put new restrictions on access to user content, it would certainly make headlines, as witnessed with Flickr's change in storage limit capacity for non-paying users.

There are similarities and common threats faced by services that are both 'paid-for' and 'free-at-the-point-of-use', namely aggravating conditions relating to increasing dependencies on the vendor's business models and the terms and conditions imposed. However, with digital materials from consumer cloud-based utilities, the business model and sustainability can only be presumed, and contracts tend to be asymmetrical in favour of the supplier. Moreover, because these services have a low barrier to entry, they may be favoured by agencies or individuals least able to respond to closure or loss. If referring to the entire platforms and the risk of the entirety of data on these, the concern is that the corporation providing the service suddenly decides it is no longer of value to them. In these circumstances, materials could be removed quickly. That has happened previously and will certainly be seen again. Preservation is not a commitment that most providers make.

Existing tools could be modified to tackle some of the closed networks. Still, it is likely to require investments, perhaps related to corporate records in some cases (thinking about internal Slacks, for instance), and more education about the importance of preserving this material and not trusting the publishing platforms to host the content forever.

Case Studies & Examples

- The example of the Microsoft outage in July 2024, in which a software update led to the cancellation of flights, healthcare disruptions and payroll issues. See **Global IT outage: More than 5,000 flights cancelled; how security 'arms race' led to crash | As it happened**, Sky News, (2024), _https://news.sky.com/story/outages-latest-airports-business-and-broadcasters-experiencing-issues-worldwide-13180821_ [accessed at 2024-09-06].

- The website 'Killed by Google' provides a list of projects and apps that Google has shut down over the years, dating back to 2006. See **Killed by Google**, _https://killedbygoogle.com/_ [accessed at 2023-10-24].

Content on Cloud Video Services Produced by the Service Provider

Video materials, primarily films and television programs which are produced by companies that maintain their own distribution platforms and are exclusively available through these platforms.

Categories

- Sound and Vision

- Cloud

Examples

Netflix, Amazon Prime, Disney+.

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of skills, commitment or policy from corporate owners; conflating backup with preservation; loss of original recordings; lack of preservation voice at executive level; poor planning and roadmap for infrastructure; slapdash procurement or migration to new systems; mergers and acquisitions; profusion of corporate systems; uncertainty over IPR or the presence of orphaned work; single point of failure; technical protection measures that inhibit reasonable preservation actions.

Mitigations

Backup and documentation; use of open formats and open source software; data management planning; licensing that enables preservation; corporate preservation capability; resilient to hacking;

authenticity and integrity managed; recognition of preservation functions at executive level; technology watch; preservation audits; participation in the preservation community.

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

This entry was added in 2019 to represent collections that are highly significant in cultural and social terms. It was adopted as the Jury was unclear whether the content could be played outside of the producers' publication platform, with technical dependencies between content and software amplified by rights management. The 2020 and 2021 Juries agreed with the Endangered classification, with discussions around how the growth of content produced with no or limited preservation mandate leads to greater risk. The continued scale of that growth and opacities regarding preservation by companies also led to the 2022 Taskforce noting a trend towards even greater risk. However, just as the 2022 Taskforce was completing its work, they welcomed the news of BFI taking on responsibility for the preservation of key titles from Netflix, commenting on how it represents a commitment to act on previous recommendations but not yet a 'material improvement' at that stage so there was no change to the 2022 trend at that time.

The 2023 Council agreed with the Endangered classification, with the overall risks remaining on the same basis as before ('No change' to trend).

The *2024 Interim Review* concluded that these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

This entry has five aspects to consider: 1. It falls outside the scope of traditional regulatory frameworks and archiving has not yet been included in any legislative framework, unlike broadcast TV, where there is a designated archive in most developed nations. 2. As a result, the collection and preservation of content from online platforms is underdeveloped, and the content remains unavailable in public archives. 3. These risks are mitigated by the fact that the commercial archives are technologically advanced, with mature digital ecosystems and skills, and much of the content has a 'long tail' business model, and as commercial products have value, preservation incentives are clear. 4. However, these are often stored at scale on LTO tapes, and so specific issues arise with the obsolescence of LTO tape technologies for the broadcast sector. 5. Nonetheless, issues remain around archiving relevant assets which may not be valued by the production company.

It may also be worth considering broadening legal deposit legislation so there is a mandate to deposit this content with an appropriate repository - though the volume may be unwelcome as many institutions are under-resourced.

Case Studies & Examples

- Work by the BFI National Archive in 2023 in the UK. A formal agreement with Netflix in 2022 was followed by a similar agreement with Amazon Prime Video in Summer of 2023, and by October 2023, the digital preservation workflow for curator-selected UK Netflix content was established, with two complete seasons (20 episodes) under preservation, and throughput building. See **Bridgerton, Top Boy and Heartstopper join the BFI National Archive and the nation's screen heritage**, BFI, (2022), *BFI News* https://www.bfi.org.uk/news/bridgerton-top-boy-heartstopper-bfi-national-archive-netflix [accessed at 2023-10-24].

Local Network Storage

Materials routinely copied or backed up to locally managed data storage facilities and able to be restored under institutional service arrangements.

Categories

- Integrated Storage

Examples

Institutional or departmental network storage and institutional data centers based on technologies such as (NAS) Network Attached Storage, (SAN) Storage Area Networks, Gluster FS and related.

Imminence	Effort
●●○○○	●○○○○
Action is recommended within five years, and detailed assessment within three years.	Loss of material in this group could be entirely avoidable if provided the means to deploy proven tools and techniques.

Hazards

Encryption; lack of routine maintenance; lack of storage replication; over-dependence on a single supplier, technology or technician; insufficient documentation; single point of failure; political or commercial interference; failure of dependencies (e.g., power supply, controller software); overly aggressive compression; poor information security; lack of integrity-checking; lack of strategic investment; lack of warranty; unenforceable warranty, encryption; Uncertainty over IPR or the presence of orphaned works.

Mitigations

Backup to different technology; backup to diverse locations; documentation of assets; integrity checking; preservation planning; refreshment planning; export functionality; resilient to hacking; selection and appraisal criteria; version control; resilient funding; technology watch; enforceable warranty; disaster planning and documentation.

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

This entry was added in 2019 to ensure that the range of media storage is properly assessed and presented.

The 2023 Council agreed with the current Vulnerable classification with overall risks remaining on the same basis as before ('No change' to trend), while also noting a slight decrease in the effort needed to preserve and the imminence of action required when compared to the 2021 Jury review.

The 2024 Council agreed These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

There has been a renewed interest in tape as offline storage is the only sure protection against advanced ransomware.

Case Studies & Examples

- **Storage**, Digital Preservation Coalition, (2016), *Digital Preservation Handbook, 2nd Edition* *https://dpconline.org/handbook/organizational-activities/storage* [accessed at 2023-10-24].

Cloud Storage

Materials routinely copied or backed up to an independently managed, off-site data storage facility and able to be restored under contractual terms

Categories

- Cloud

- Integrated Storage

Examples

Remote network storage provided by a third-party service under contracts, such as DropBox, Amazon, Microsoft Azure, Dell EMC, Google Cloud Platform, Google Drive, IBM, Rackspace, Iron Mountain, SAP, and others.

Imminence	Effort
●●○○○	●●●●○

Imminence	Effort
Action is recommended within five years, and detailed assessment within three years.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Lack of skills, commitment or policy from corporate owners; Encryption; lack of routine maintenance; lack of storage replication; over-dependence on a single supplier; insufficient documentation; lack of local alternative; political or commercial instability; overly aggressive compression; poor information security; lack of transparent integrity-checking; lack of strategic investment; lack of migration plan; lack of exit strategy; unenforceable penalties; unstable pricing; unpredictable removal costs; uncertainty over IPR or the presence of orphaned works.

Mitigations

Backup to different technology; backup to diverse locations; documentation of assets; integrity checking; preservation licensing and planning; export functionality; resilient to hacking; version control; resilient funding; technology watch; enforceable contract; disaster planning and documentation; stable pricing; budgeted removal costs.

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

This entry was added in 2019 to ensure that the range of media storage is properly assessed and presented. The 2021 Jury noted increased risk in light of greater reliance on the cloud and localized disruptions to cloud services over the pandemic. A 2021 trend towards greater risk was based on the wider (global) dependence on these services, especially Google Drive, for record-keeping and business workflows. The impact of loss increased with more reliance on cloud services leading to greater risk; however, this should not deter people from using cloud storage. The 2022 review agreed with this assessment but noted no significant increase in trend for 2022.

The 2023 Council moved this entry to a new higher-level Cloud species as the previous Integrated Storage species worked less well (for hardware technologies). The Council agreed with the previous Vulnerable classification, with the overall risks remaining on the same basis as before so long as there are safeguards in place ('No change' to the 2023 trend). However, the Council noted that these safeguards may not, in all cases, be sufficient to address existing risks. Council members noted how some governments may cut off the internet in times of unrest, having a disastrous effect on access to cloud-based resources, and raised questions about the feasibility of recovering material after a major cloud vendor fails or due to malicious acts. For these materials, the significance of loss and effort to preserve is much greater, with the potential for a trend towards greater risk with the loss of existing safeguards.

The 2024 Council agreed these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

While overall risk remains on the same basis as before, some Council members pointed out how a lack of transparency in knowledge about how a cloud service is actually built and functions is worrying from a preservation perspective. Additionally, the overall political 'threat situation' worldwide seems to be increasing, which means that significant changes in national political regimes can affect the predictability of how the material is handled in a cloud service and, with that, the potential for increased risk.

Additional Information

To add further clarity, Council members in the Integrated Storage species group noted that there is a distinction between 'in-house' physical storage and cloud storage, especially if one relies on cloud storage as the only storage provider for digital content. As they understand it, this 'Cloud Storage' entry focuses on material copied or backed up to a third-party cloud service. This is less threatening compared to using the cloud as the sole storage provider for content preservation.

The history of digital preservation suggests that the risk of vendors going out of business or shutting down services is the key issue here, over and above any specific technical solutions or risks.

Case Studies & Examples

- The example of the Microsoft outage in July 2024, in which a software update led to the cancellation of flights, healthcare disruptions and payroll issues. See **Global IT outage: More than 5,000 flights cancelled; how security 'arms race' led to crash. As it happened**, Sky News, (2024), _https://news.sky.com/story/outages-latest-airports-business-and-broadcasters-experiencing-issues-worldwide-13180821_ [accessed at 2024-09-06].

- Case of cloud storage provider who accidentally deleted a client account – including all replicas and backups. This emphasises that a single third-party provider should only really be considered a single copy regardless of the resilience the provider puts in place. Cloud introduces new single points of failure. See **Google Cloud explains how it accidentally deleted a customer account**, Amadeo, R., (2024), _https://arstechnica.com/gadgets/2024/05/google-cloud-explains-how-it-accidentally-deleted-a-customer-account/_ [accessed at 2024-06-17].

- Case of a cloud storage provider who suffered major data loss (or its clients suffered data loss) due to a fire in its data centre. Those clients suffered most who did not include geographically redundant storage in the contract with the storage provider as this was more expensive. See **Millions of websites offline after fire at French cloud services firm**, Rosemain, M. and Satter, R., (2021), _Reuters https://www.reuters.com/article/us-france-ovh-fire-idUSKBN2B20NU_ [accessed at 2023-10-24].

- Case of fired credit union employee accessing the financial institution's computer systems without authorization and destroying over 21 gigabytes of data via remote network storage. See **Fired NY credit union employee nukes 21 GB of data in revenge**, Gatlan, S., (2021), _BleepingComputer https://www.bleepingcomputer.com/news/security/fired-ny-credit-union-employee-nukes-21gb-of-data-in-revenge_ [accessed at 2023-10-24].

- The National Archives UK (2023) 'Digital Services and carbon emissions in the heritage sector: some preliminary findings', which noted areas relating to the cloud and cloud storage.

They write "If we are looking for areas where significant carbon reductions could be made quickly, they are not to be found here. The evidence is that hosting digital services on site results in more carbon emissions than a sensibly located (i.e., in a territory with a high proportion of electricity generated from renewables) cloud host and that, where it might be felt that migrating services simply migrates emissions from scope 2 to scope 3, in practice cloud providers can offer the same storage and compute with lower emissions. Amazon in particular reports its view of the carbon 'saved' by using its services rather than your own, but these are estimates and should not be regarded as robust." See **Digital Services and carbon emissions in the heritage sector: some preliminary findings**, The National Archives (UK), (2023), _https://www.nationalarchives.gov.uk/archives-sector/digital-services-and-carbon-emissions-in-the-heritage-sector-some-preliminary-findings/_ [accessed at 2023-10-24].

b)   Ephemeral Services

Email

Documents, correspondence and other records created in the course of contractual dealings between individuals and agencies, especially where the subjects are of long duration and may be subject to legal scrutiny at undefined points in the distant future.

Categories

- Formats

Examples

Gmail, Hotmail, Yahoo Mail, Outlook, and email in all its forms including individual messages, threads of conversation, mailboxes, email servers and file attachments.

Imminence	Effort
●●○○○	●●○○○
Action is recommended within five years, and detailed assessment within three years.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Lack of skills, commitment or policy from corporate owners; uncertainty or conflicting claims over IPR or orphaned works; use of personal accounts for professional work and vice versa; proliferation and duplication of attachments; email not recognized as a record; absent, unworkable or inconsistent records management; dependence on free cloud-based services; lack of migration path; lack of preservation planning; perverse incentives to delete; encryption.

Mitigations

Application of appraisal and selection tools; timely transfer to preservation facility or archive; commitment to transparency; preservation policy; working preservation plan; clear migration path; widespread recognition of email as a record.

Bit List History

Added to list: 2017

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

This entry was added in 2017, but the Jury did not have the capacity to assess it in detail. It was reviewed and assessed in 2019, including highlights to significant developments, including the recommendations of the Email Preservation Taskforce and the development of the ePADD software. Email presents many preservation challenges, from scale through core technologies, attachments, privacy and intellectual property rights. Because this entry intersects with many others, the aggravating conditions associated with email should be considered in conjunction where relevant. The 2021 Jury agreed with the Endangered classification with an identified trend toward reduced risk. The 2021 trend was based on discussions on continued developments in email preservation tools and techniques, as well as the growing number of archives preserving email content while recognizing issues with providing access to preserved email content. They added that ongoing records management policies towards corporate or business email need to be better embedded to stop the loss of important email content, and more awareness is needed around the potential of personal email. While record-keeping legislation and mandates direct retention periods, email document decisions taken by government officials at local, regional and national levels are not always well maintained, if at all; a loss could impact people's lives along with their ability to assert rights. The 2022 Taskforce agreed on a trend towards reduced risk based on material improvement with applied examples of good practice, including (but not limited to) approaches to creating a PDF format for the preservation of email, and improvements to existing software, tools and workflows supporting complex email preservation.

The 2023 Council agreed with the Endangered classification, with risks remaining on the same basis as before ('No change' to the 2023 trend). However, they also noted a decrease in the imminence of action and effort to preserve. The Council also recommended noting areas of overlap with the 'Cloud-based Services and Communications Platforms' entry as it pertains to the saving and preservation of email in cloud-based services such as Microsoft Sharepoint.

The *2024 Interim Review* concluded that these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend). They agree with the Endangered classification, with the potential for new initiatives around a profile of PDF specifically for email archiving coming later in 2024, and how this could impact the community.

Additional Information

Email is hugely important as it has been so pervasive as a communication mechanism for society. Some methods used and responsibility adopted for collecting at the business and public body level

(again will differ globally), but this will be a fraction of the communities that use it, and few will be set up for the long-term care of this data.

Case Studies & Examples

- Resources and outputs from the EA-PDF project to identify the essential characteristics and optimal functional requirements of email messages and necessary related information in a PDF technology-based archive. See **EA-PDF**, PDF Association, (2021), https://www.pdfa.org/resource/ea-pdf/ [accessed at 2023-10-24].

- Resources and outputs from the Integrating Preservation Functionality into ePADD (ePADD+) project to integrate long-term email preservation functionality into the program and provide a tool supporting the email archiving lifecycle more robustly. See **History**, ePADD, https://www.epaddproject.org/about/history [accessed at 2023-10-24].

- Resources and outputs from the RATOM project to develop software to assist archives and other collecting organizations with email analysis, selection, and appraisal tasks. See **About**, *RATOM* https://ratom.web.unc.edu/about/ [accessed at 2023-10-24].

- **Preserving Email (2nd Edition)**, Prom, C, (2019), *DPC Technology Watch Report 19-01* http://doi.org/10.7207/twr19-01.

- **Preserving Email**, Artefactual Systems and DPC, (2021), *DPC Technology Watch Guidance Notes* http://doi.org/10.7207/twgn21-08.

- **The Future of Email Archives: A Report from the Task Force on Technical Approaches for Email Archives**, Murray K and Prom, C., (2018), *CLIR* https://clir.wordpress.clir.org/wp-content/uploads/sites/6/2018/08/CLIR-pub175.pdf [accessed at 2023-10-24].

- "Novice to Know-How: Email Preservation", was conceived of and funded by The National Archives (UK) and delivered by the DPC. This course is aimed at learners who already have a solid foundational knowledge of digital preservation (e.g. they have completed the original N2KH learning pathway) and wish to gain practical skills in relation to the preservation of email. See **Novice to Know-How: Email Preservation**, DPC and The National Archives (UK), (2023), https://www.dpconline.org/digipres/prof-development/n2kh-online-training [accessed at 2023-10-24].

- **2013 – 2023: A Review of Ten Years of Email Archiving in France**, Ville, M., (2018), *iPRES 2023 Conference, Urbana-Champaign, Illinois, USA, 19–22 September* <>.

---

c)   **Proprietary Data**

Smart TV Apps

Apps created for smartphones. Many are deprecated quickly but others survive through multiple update cycles. It is hard to maintain version control and often dependent upon the company that publishes them. There is no clear agency or mandate to record or collect.

Categories

- Apps

Examples

UlangoTV; ITV Hub; Plex; apps created for older or no longer manufactured TVs

Imminence	Effort
●●●●○	●●●●○
Action is recommended within twelve months. Detailed assessment is a priority.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Device dependence; poor documentation; lack of preservation interest or mandate by company; uncertainty over IPR or the presence of orphaned works; deletion of stores or apps

Mitigations

Strong documentation; development of tools to archive stores and preserve menus online via the WARC format, and for HTML5 and Flash-only based smart TV media; designated repository taking preservation responsibility and capacity to deliver

Bit List History

Added to list: 2023

Trends

- **2024:** No change.

Last Review

2023 Review

This was a new Bit List entry added in 2023 to draw attention to the particular challenges of content and software preservation for apps designed and used for smart TVs. While there are some similarities with the 'Smartphone Apps' entry in regard to device dependency, frequent updates, and lack of preservation interest or mandate by companies that publish them, this entry focuses on the distinct risks relating to the availability and access to apps made specifically for smart TVs, which can be more short-lived, variable, and challenging to access and find a rootable version for preservation actions. These apps, designed and used specifically for smart TVs, are deprecated quickly or undergo frequent updates. It is challenging to maintain version control and preserve the content. Preservation efforts often depend upon the availability and access to rootable versions of the apps by the companies that publish them, particularly for older TVs that are no longer manufactured. There also remains a lack of clear agency or mandate to record or collect from corporate owners and barriers to access for preservation efforts outside this context.

The 2023 Bit List Council recommended that the next major review consider creating a new Broadcast species group where this entry may be better suited. Council members additionally noted

that it may be better to have a broader entry for smart device apps rather than specifically 'Smart TV Apps' and 'Smartphone Apps' as separate standalone entries. The risks posed by these apps can apply to a variety of devices; for example, before Smartphones, there were PDAs, and there are now Smart fridges, set-top-boxes, fire sticks, e-book readers, and other devices which have apps. For this reason, they recommended that the next major review for the Bit List includes a rescoping of Smart TV Apps and Smartphone Apps to consider: What differentiates these apps from others? What cultural heritage purpose do these apps serve? Are organizations collecting them? Are there distinct aggravating factors or risk profiles?

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

In regard to preservation actions, rooting methods for old smart TVs are often non-existent or hard to do (and sometimes manageable when they do exist); it is challenging to buy or find owners of older smart TV units with removed-from-the-store apps that have rootable versions and/or is willing to root.

The content these apps contain can hold social and cultural significance; they can provide a history of television, entertainment, and exclusive or unique content only made available through the apps specifically designed for smart TVs. Although there is a small community of people interested in old smart TVs and a bigger one focused on new ones too, in the future, when the technologies become even more obsolete, there will be researchers and others interested in looking back at Smart TVs as a piece of the history of entertainment and even culture.

Case Studies & Examples

- The nominator of this entry provided UlangoTV as an example of an archive of old smart TV apps. The domain of UlangoTV, an app that aggregated live Internet Protocol Television (IPTV) streams, was taken down by the Alliance for Creativity and Entertainment (ACE). Domain records show that Ulango.tv is now owned by the Motion Picture Association, indicating an agreement and handover took place; at the time of writing, servers hosting the smart TV app archive are currently in an evidence locker with no indications for releasing without paying to get the full archive. See **ACE Shuts Down UlangoTV 'Pirate' IPTV App, Seizes Domain**, Maxwell, A., (2020), *TorrentFreak News* *https://torrentfreak.com/ace-shuts-down-ulango-tv-pirate-iptv-app-seizes-domain-200113/* [accessed at 2023-10-24].

- The postmarketOS wiki offers a central guide on smart TV rooting methods for different models. See **Category:TV**, postmarketOS, *https://wiki.postmarketos.org/wiki/TVs* [accessed at 2023-10-24].

Smartphone Apps

Applications created for smartphones. Many are deprecated quickly but others survive through multiple update cycles. It is hard to maintain version control and is often dependent upon the company that publishes them. There is no clear agency or mandate to record or collect.

Categories

- Apps

Examples

London 2012 app; BBC Olympic app; apps published for Apple iOS 10 or earlier; apps for less-common operating systems such as Palm OS, Symbian, Blackberry OS

Imminence	Effort
●●●●●	●●●●●
Immediate action necessary. Where detected, should be stabilized and reported as a matter of urgency.	Loss seems inevitable. Loss has already occurred or is expected to occur before tools or techniques develop.

Hazards

Device dependence; poor documentation; uncertainty over IPR or the presence of orphaned works; short-term contracts; lack of skills, commitment or policy from corporate owners; rapid churn of OS; shifting business requirements of app resellers; dependence on exotic or obsolete formats or OS processes

Mitigations

Strong documentation; version control for code and compiled app; emulation enabled; designated repository taking preservation responsibility and capacity to deliver

Bit List History

Added to list: 2017

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2017 to draw attention to the challenges of software preservation and the extraordinary velocity of the market for apps. Given the speed of change, it is hard to see how digital preservation efforts can keep pace. The 2019 Jury noted that splitting this entry into different groups based on the platform would clarify differences, although the risks would be largely the same. The

2021 Jury discussed this further in light of the new 2021 'Smartphone Gaming' entry, which can be considered a subset of this category as well as Gaming in which it is currently assigned. The 2021 Jury found no identifiable trend towards greater or reduced risk but discussed the impact of loss; some loss seems inevitable due to the changes that have taken place in mobile ecosystems over the past 15 years. The 2022 Taskforce noted no change to trend (they agreed these risks remain on the same basis as before with no significant increase or decrease over the preceding year).

The 2023 Council agreed with the Critically Endangered classification with no change to trend but also recommended a change to significance of loss; the significance of loss was increased from the previous year's reviews, suggesting that the loss of tools or services within this group could have a larger global impact. Obsolescence plays a large part; the rate of change of smart operating systems is high, as are the number of apps, which means the obsolescence rate is high. Furthermore, a large number of vendors are also leaving the market, which means their specific apps will also vanish. While the Council agreed the entry description should be updated to reflect these areas of discussion, overall risks remained on the same basis as before ('No change' to the 2023 trend).

Members of the 2023 Council also noted that it may be better to have a broader smart device apps entry rather than specifically 'Smartphone Apps' and 'Smart TV Apps' as separate standalone entries. The risks posed by these apps can apply to a variety of devices; for example, before Smartphones, there were PDAs, and there are now Smart fridges, set-top-boxes, fire sticks, e-book readers, and other devices which have apps. For this reason, they recommended that the next major review for the Bit List includes a rescoping of Smartphone Apps and Smart TV Apps to consider: What differentiates these apps from others? What cultural heritage purpose do these apps serve? Are organizations collecting them? Are there distinct aggravating factors or risk profiles?

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

Old versions of apps are completely lost to most users: once you upgrade an app, you typically cannot go back. Perhaps iOS is more critical - at least with Android, you can often get .apk from the internet separate from the marketplace. The NSRL contains hundreds of thousands of mobile applications which are not being actively preserved but could be if a mandate existed. An extension to Legal Deposit might be possible.

The faster we act, the less we will lose. It is unlikely that there will ever be one agent with a mandate to collect different apps available in different countries, so a network of national organizations would be needed. The companies that create these apps are the key to the licensing challenges, and conversation with them is necessary, though it would need to happen immediately in order to negotiate the right to preserve/escrow both apps, operating systems, documentation, and phone development emulators.

Messaging apps such as WhatsApp or Telegram have had an increased media presence in the last couple of years due to their role in a number of politics-related issues that have arisen, such as concerns about UK Cabinet Ministers using the auto-delete function which could compromise accountability and transparency of the UK government. Telegram has also gained importance due to its use in the Russia-Ukraine war for sharing news.

Case Studies & Examples

- **Emerging Formats**, British Library, (2022), *British Library*
  https://www.bl.uk/projects/emerging-formats [accessed at 2023-10-24].

- **Collecting complex digital publications: testing an enhanced curation method**, Smith
  Nicholls, F., (2023), *British Library* https://doi.org/10.23636/kff3-jv09 [accessed at 2023-10-24].

- **EaaSI Software Development: Considerations, Priorities, and Commitment**, Rechert, K.,
  Gates, E., Cochrane and E., Anderson, S., (2021), *Software Preservation Network*
  https://www.softwarepreservationnetwork.org/eaasi-software-development-considerations-priorities-and-commitments/ [accessed at 2023-10-24].

- **Towards a preservation workflow for mobile apps**, Van der Knijff, J. , (2021), *BitsGalore*
  https://www.bitsgalore.org/2021/02/24/towards-a-preservation-workflow-for-mobile-apps
  [accessed at 2023-10-24].

- **Cabinet policy obliges ministers to delete instant messages**, Siddique, H. , (2021), *The
  Guardian* https://www.bitsgalore.org/2021/02/24/towards-a-preservation-workflow-for-mobile-apps [accessed at 2023-10-24].

- **'Considerations in Archiving Misinformation from Encrypted Messaging Apps**, Prabhakar, T.,
  & George, D, (2019), *Tattle* https://cyber.harvard.edu/sites/default/files/2019-12/Tattle_Disinformation_Workshop_revised.pdf [accessed at 2023-10-24].

- **Experts recover messages from Boris Johnson's old mobile phone**, Allegretti, A. , (2023), *The
  Guardian* https://www.theguardian.com/politics/2023/jul/21/boris-johnson-experts-recover-messages-old-phone-covid-inquiry [accessed at 2023-10-24].

- **Telegram is the app of choice in the war in Ukraine despite experts' privacy concerns**, Allyn,
  B. , (2022), *NPR* https://www.npr.org/2022/03/14/1086483703/telegram-ukraine-war-russia
  [accessed at 2023-10-24].

Web Domains with no Legal Deposit

This entry regards the preservation of websites and domains that fall outside a remit of legal deposit (or no legal deposit mandate exists). Web archiving is able to capture large quantities of materials with routine and standards-based tools, but there are significant issues arising with intellectual property rights associated with website capture and republication. In many jurisdictions, but by no means all, those obstacles are overcome by regulations that enable a national library or other 'legal deposit' agency to copy and preserve content. Where no such permission exists, there is a significant risk of loss.

Categories

- Web

Examples

Domains registered without a country code; domains with a country code but weak or unenforceable legal deposit permission to harvest

Imminence	Effort
⬤⬤⬤⬤⬤	⬤⬤⬤⬤⬤
Immediate action necessary. Where detected, should be stabilized and reported as a matter of urgency.	Loss seems inevitable. Loss has already occurred or is expected to occur before tools or techniques develop.

Hazards

Uncertainty over IPR or the presence of orphaned works; lack of legal deposit mandate or remit; rapid churn of websites; lack of access to Internet Archive harvest; contentious content; encryption; digital rights management; non-standard content management

Mitigations

Permissive approach to Legal deposit; legislation to support and/or manage associated risks

Bit List History

Added to list: 2015

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019. It is characterized by regulatory barriers rather than technical ones, though the pace of change in web technologies, as well as the growth of web content, means that significant technical challenges still exist. The 2019 Jury noted that local conditions were also a significant factor; for example, websites often also fall under public records legislation or are important elements of corporate records, and so important parts of the web are harvested even when there is no explicit legal deposit legislation. The 2019 Jury particularly recognized the work of the Internet Archive to capture and preserve content. They noted significant gaps in web archiving and, in too many cases, regulation as the barrier. The 2021 Jury agreed with this description and classification but added that in some limited instances, pywb tools (as opposed to automated web crawlers like Heritrix) could effectively capture the look and feel of a platform interface, preserving legacy versions for users to interact with in the future. However, pywb tools are manual and, therefore, cannot address the scale of the issue. They also do not capture interfaces in a way that makes it possible to recreate them in the future, only interact with a defined set of web pages. For this growing issue of scale, the 2021 Trend was towards greater risk. The 2022 Taskforce agreed with noted no change to the trend.

The 2023 Council agreed with the Critically Endangered classification. They also noted an increase in the imminence and inevitability of loss, recognizing that while the need for major efforts to prevent

or reduce losses continues, it is much more likely that loss of material has already occurred and will continue to do so by the time tools or techniques have been developed. While the Council agreed the entry description should be updated to reflect these areas of discussion, overall risks remain on the same basis as before ('No change' to trend).

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Council members also added that the presence of a clear IPR framework for preservation is enabling, whether it is through legislation (like legal deposit) or licensing.

Additional Information

There is not only a significant risk of loss to the content but also risk of loss to access. Unless the Internet Archive is picking these up, the early web or permission regimes are in place, and these early instances are gone forever and will continue to be lost.

Case Studies & Examples

- In the 2023 NDSA Web Archiving Survey Report, one of the major takeaways was that respondents indicated that 'staff capacity' was the biggest barrier to web archiving. Few organizations dedicate more than one, full-time employee to web archiving and very rarely is someone's entire position dedicated to web archiving. See **Web Archiving Survey Results: An NDSA Report**, National Digital Stewardship Alliance (NDSA), (2023), *National Digital Stewardship Alliance (NDSA)* *https://doi.org/10.17605/OSF.IO/N5MYR* [accessed at 2024-09-11].

Data Posted to Defunct or Little-used Social Media Platforms

Older or less widely used social media platforms to which content has been uploaded but for which no guarantees have been made about the long term.

Categories

- Social Media

Examples

BeBo, MySpace, Google Buzz, Parler and others.

Imminence	Effort
●●●●●	●●●●●
Immediate action necessary. Where detected, should be stabilized and reported as a matter of urgency.	Loss seems inevitable. Loss has already occurred or is expected to occur before tools or techniques develop.

Hazards

Closure of platform; lack of offline equivalent; lack of export functionality; no preservation undertaking from service provider; unstable business plan from service provider; Uncertainty over IPR or the presence of orphaned works.

Mitigations

Offline Replication; clear notice periods and alerts; committed ongoing maintenance of service.

Bit List History

Added to list: 2018

Trends

- **2023:** No change.

- **2023:** No change.

Last Review

This entry was nominated in 2017 and added to the Bit List following the 2019 assessment to highlight the different threats faced when attempting to preserve materials on older or defunct social media, emphasizing the different threats faced by social media users who uploaded content to defunct or little-used social media platforms. Because these services are older, the need to act is more urgent than for others. Often, the significance is only brought to attention once they are lost. The 2021 Jury noted a trend towards greater risk due to the existing risks of defunct or little-used platforms with recognition of the need to develop tools or techniques for applying to others that may follow the same path. The 2022 Taskforce agreed these risks remain on the same basis as before (no change to the trend).

The 2023 Council agreed with the Critically Endangered classification. They noted an increase in imminence and effort to preserve, recognizing that while the need for major efforts to prevent or reduce losses continues on the same basis as before, it is now much more likely that loss of material has already occurred and will continue to do so by the time tools or techniques have been developed. Therefore, immediate action is necessary. They also recommended that the next major review for the Bit List consider merging this entry with the 'Consumer Social Media Free at the Point of Use' entry to provide examples of loss prompted by aggravating conditions.

The 2024 Council agreed these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

They recommend combining this entry with 'Consumer Social Media Free at the Point of Use' by illustrating how social media is Practically Extinct in the presence of aggravating conditions, such as sudden loss or loss with little notice, citing examples such as GeoCities and MySpace. Additionally, they recommend that the 2025 review address recommendations about scoping or combining with 'Consumer Social Media Free at Use' and moving 'Defunct' social media into 'Consumer SM' as examples of extinct platforms. Additionally, 'Lesser Known' platforms might be re-scoped as 'Local or Small Community' platforms, such as BBSs/Blogs maintained by a community, or national platforms like Skyblog (which has been successfully archived by Bibliothèque nationale de France). These

smaller platforms may be at a lower risk of extinction due to fewer political/corporate entanglements and an invested user community who maintains them and can support archiving.

Additional Information

The risk to this content depends on the specific service or platform, but older platforms (BeBo, MySpace) pose a higher risk of loss than current platforms (and is likely already lost) but social media wasn't used to the same extent (and not as widely used by government, corporations, research institutions, etc.) in the early 2000s/2010s when these platforms were popular, which reduces the impact slightly.

When looking at the digital preservation landscape and where we need to apply effort as well as resources, defunct early social media spaces are not high on the list; but, when considering how contemporary social media channels could become defunct, it becomes a different conversation because of how intrinsically tied they are to political discourse and influencing political opinion

It is to be hoped that some of these have been archived via traditional web archiving, and so the remnants of these sites can be found in bits and pieces in various web archives, but it may be too late to save some of the content that is likely already lost. If some of this is still available, there may be hope in trying to preserve, but it may be difficult if the platforms are not willing to share data or work with preservationists. ArchiveTeam has stepped in here too. There is undoubtedly a story here that could be used as a call for arms to raise awareness about the preservation of current social media platforms too.

Case Studies & Examples

- The mentioned example of Skyblog as national platforms which has been successfully archived by Bibliothèque nationale de France. See **La Bibliothèque nationale archive les Skyblogs**, Bibliothèque nationale de France, (2024), *BnF* *https://web.archive.org/web/20240506232352/https://www.bnf.fr/fr/la-bibliotheque-nationale-archive-les-skyblogs* [accessed at 2024-09-06].

- The shutting down of Yahoo Answers is an example of the loss of content from community board style web platforms. See **Deleting Yahoo Answers is a disastrous idea. For history's sake, we need to preserve our digital record**, Tsavkko Garcia, R., (2021), *Business Insider* *https://www.businessinsider.com/deleting-yahoo-answers-disastrous-idea-preserve-our-digital-record-2021-4?r=US&IR=T* [accessed at 2023-10-24].

Born Digital Photographs and Video shared via Social Media or Uploaded to Cloud Services

Digital images or video with no analogue equivalent and where the only copy is online with a social media platform. This entry includes images or videos created and shared as part of personal digital archiving, but also for businesses and others publishing data only via these services. Users of these services will likely lose their data if social media companies fold or make extracting or downloading data difficult.

Categories

- Social Media

Examples

Flickr; Vimeo; YouTube; Instagram; Periscope; Snapchat; TikTok; Vine; Facebook; X (previously Twitter).

Imminence	Effort
●●●○○	●●●●○
Action is recommended within three years, detailed assessment within one year.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Lack of skills, commitment or policy from corporate owners; lack of storage replication by provider; dependence on proprietary products or formats; poor management of data protection; inaccessibility to automated web crawlers; political or commercial interference; lack of offline equivalent; over-abundance; Uncertainty over IPR or the presence of orphaned works; lossy compression applied in upload scripts.

Mitigations

Offline backup; lossless compression; good documentation; access to web harvesting; clarity of intellectual property rights that enable preservation; credible preservation commitment from service provider; export pathway.

Bit List History

Added to list: 2018

Trends

- **2023:** No change.

- **2023:** No change.

Last Review

This entry was added by the 2018 Jury as a subset of a broader social media entry first introduced in 2017 to draw attention to the different challenges faced by the growing volumes of photographs and video recordings on social media continuing to rise and, therefore, the scale of the challenge to ensure a meaningful legacy aggravated by the issue of overabundance in which appraisal decisions for preservation or deletion are overwhelmed. The entry shares many challenges with others in the social media group, with a dependency on a global service provider whose business model can only be presumed and tied to users via asymmetrical contracts that favour the supplier. Moreover, because these services have a low barrier to entry, they are used by agencies or individuals least able to respond to closure or loss. Both the 2020 and 2021 reviews of the entry noted a trend towards greater risk. The 2020 trend referenced the closure of the EverAlbum photo storage and changes to

the Flickr free service, which provided examples of the short turnaround of closures within the photo-sharing community and pointed to the volatility in the market. The 2021 trend was added in light of surrounding global crises (predominantly the coronavirus pandemic, compounded by vaccine hesitancy, but also the deterioration of the world's democracies) as a result of widespread misinformation, increasing the significance and impact of loss of digital materials. The 2022 Taskforce agreed these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk (no change to the trend).

The 2023 Council agreed with the Endangered classification and noted an increase in effort to preserve, recognizing that while the need for major efforts to prevent or reduce losses continues, it is now much more likely that loss of material has already occurred, and will continue to do so, by the time tools or techniques develop.

The 2023 Council recommended adjusting this and other social media entries in light of how web-based and cloud-based business products and services had developed in recent years. This restructuring and revision included:

Narrowing the scope. The scope of the entry has been narrowed to draw attention to the challenges of preserving images and videos disseminated through social media platforms specifically (e.g., sharing-driven and social networking platforms such as YouTube, Instagram and TikTok). These challenges primarily relate to harvesting and managing the images, video recordings and data generated by users' interactions on web-based networking platforms.

As part of the above rescoping, the entry's name was changed from 'Born Digital Photographs and Video Shared via Social Media or Uploaded to Cloud Services' to 'Born Digital Photographs and Video Shared via Social Media Platforms', and information concerning cloud-based aspects were incorporated into the 'Cloud-based Services and Communications Platforms' entry to more clearly differentiate the risks associated with cloud hosting and computing technologies.

The 2023 Council additionally recommended that the next major review revisit and consider merging the Born Digital Photographs and Video shared via Social Media Platforms and Consumer Social Media Free at the Point of Use entries. This is mostly due to the fact that so many of the 'regular' social media platforms have tended toward more ways to mimic or copy TikTok style videos, and making the distinction will become harder in the future since they all have similar functionality and ways to create photo/video content.

The 2024 Council agreed these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

They added that the platforms listed are publication platforms for media, and there's a risk of vulnerability, loss and lack of tools to preserve media longer than what the platform can provide, especially for creative industries, small business, and political activists. Politicians and government agencies will post public records to these platforms. Without archived YouTube, Vimeo, etc. publication pages, those public-facing records lose contextual value. They also recommend that for the 2025 review, rescoping or absorbing this entry into another should be discussed with the wider Bit List council.

2024 Council members also raised concerns regarding Artificial Intelligence and Machine Learning, noting that for this entry and, more broadly, anything related to Social Media, an emerging risk is AI training fears. This manifests in two ways:

Social media users deleting their content entirely in an attempt to prevent it from being used or sold for training a Large Language Model (Edwards, 2024). However, as noted by the Council members, it can be difficult to have a good sense of how widespread this is or how much it's affecting content that people want to preserve – but it is nonetheless worth noting.

Website operators blocking scrapers for similar reasons, (Voorhees, 2024). This primarily targets AI crawlers but may also affect web crawlers used for preservation. It can similarly be difficult to have a good sense of how widespread this is, but it seems likely that anti-crawler provisions are only going to get worse, not better.

Additional Information

We can point to some examples of content only on YouTube, for instance, that may be desired for acquisition for a library. Typically, YouTube would be acquired through web archiving, but with recent and ongoing challenges capturing this content, it may require contacting the creator to try to acquire the original video files to preserve through other workflows. This creates the challenge of determining who to reach out to, how to transfer those files and, if the files only exist on the social media platform, how to extract them to transfer to an organization for preservation. With crawling capabilities limited these days, Libraries will have to rely more on individuals archiving their own content and donating it to organizations. It's not clear what that workflow looks like and if there are adequate methods to support it.

The types of users for these services vary greatly - from a private individual uploading a few videos to share with friends to major agencies who use the platform to disseminate important information or research. The extent to which private individuals and even large institutions are aware of digital preservation risks is unclear, though anecdotal evidence suggests that awareness is extremely low. Therefore, it can be assumed that most users (regardless of the significance of their content) do not keep local copies or take other measures to mitigate the risk of loss from these types of platforms. Additionally, risk varies from platform to platform. YouTube, for example, only allows low-quality downloads even for content owners. Therefore, if a content owner lost or deleted an original video file, it would be impossible to recover a high-quality copy from YouTube.

The vast majority of content may be accessible for as long as the platform where it is hosted is popular (and has a viable business model); however, more insidious content (such as malicious misinformation or hate speech) may be deleted by content creators (potentially backed by hostile governments) to avoid prosecution or tracing. It is unclear to what extent these platform providers are compelled to provide access to servers / deleted content or private content for evidential purposes in the course of legal or criminal investigations. The lack of transparency and standardized international regulation of these platforms make their content vulnerable to exploitation and malicious use by individuals, corporations, and hostile governments.

With digital materials from premium or institutional social media services, the business model and sustainability are more obvious, and contracts may be enforceable more readily. Moreover, because these services have a slightly higher barrier to entry, they may be favoured by agencies better able to respond to closure or loss. Traditional web archiving can be employed where the user pays for a service, but the content is ultimately publicly available (such as Flickr). But much is unclear about how to preserve internal social media / closed networks that web archiving cannot get to, or existing tools do not cover.

Case Studies & Examples

- Mentioned examples relating to AI and ML concerns See **Stack Overflow users sabotage their posts after OpenAI deal**, Edwards, B., (2024), *Ars Technica* https://arstechnica.com/information-technology/2024/05/stack-overflow-users-sabotage-their-posts-after-openai-deal/ [accessed at 2024-09-06].

- Mentioned examples relating to AI and ML concerns See **How We're Trying to Protect MacStories from AI Bots and Web Crawlers – And How You Can, Too**, Voorhees, J., (2023), https://www.macstories.net/stories/ways-you-can-protect-your-website-from-ai-web-crawlers/ [accessed at 2024-09-06].

- Example of Periscope shutting down See **Periscope shuts down today**, Kastrenakes, J., (2021), *The Verge* https://www.theverge.com/2021/3/31/22360717/periscope-shut-down-twitter-live-streaming [accessed at 2023-10-24].

- Example of loss of users' content with social media platforms such as MySpace. See **Myspace loses all content uploaded before 2016**, Hern, A., (2019), *The Guardian* https://www.theguardian.com/technology/2019/mar/18/myspace-loses-all-content-uploaded-before-2016 [accessed at 2023-10-24].

- Examples of the short turnaround of closures within the photo-sharing community and point to the volatility in this market, such as the closure of the Ever photo storage service. See **Ever, once accused of building facial recognition tech using customer data, shuts down consumer app**, Perez, S., (2019), *TechCrunch* https://techcrunch.com/2020/08/24/ever-once-accused-of-building-facial-recognition-tech-using-customer-data-shuts-down-consumer-app/ [accessed at 2023-10-24].

- The changes to the Flickr free service also provides an example. However, they have since given thought to the issue, referred to as the 'data lifeboat'. See **Flickr will end 1TB of free storage and limit free users to 1,000 photos**, Gartenberg, C., *The Verge* https://www.theverge.com/2018/11/1/18051950/flickr-1000-photo-limit-free-accounts-changes-pro-subscription-smugmug [accessed at 2023-10-24].

- **Designing a Data Lifeboat**, (2023), *Flickr Foundation* https://www.flickr.org/programs/content-mobility/data-lifeboat/ [accessed at 2023-10-24].

- The Terms of Service; Didn't Read (ToS;DR) project, started in June 2012 provides a useful resource and reference for issues relating to content sharing platforms and social media terms of service. See **Terms of Service Didn't Read**, (2012), *The Terms of Service; Didn't Read (ToS;DR) project* https://tosdr.org/ [accessed at 2023-10-04].

- Museums, Libraries, and Archives have begun to pay attention to this content through projects like Collecting Social Photo (CoSoPho), but no breakthroughs have been made. See **Collecting Social Photo**, *The Finnish Museum of Photography* https://www.valokuvataiteenmuseo.fi/en/projects/collecting-social-photo [accessed at 2023-10-24].

- **Slop is the new name for unwanted AI-generated content**, Willison, S., (2024), *Simon Willison's Weblog* https://simonwillison.net/2024/May/8/slop/ [accessed at 2024-09-12].

Digital Music and Ephemera Shared on Social Media

Digital materials created by musicians and fans as a by-product of performance or recording, shared on websites and other social media platforms.

Categories

- Sound and Vision

Examples

Fan sites; private or illicit recordings of concerts; informal music sharing between networks such as TikTok, MySpace and Facebook

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Dependence on social media provider; lack of offline equivalent; uncertainty over IPR or the presence of orphaned works; unstable or small community of interest; encryption.

Mitigations

Offline equivalent; intellectual property rights conducive to preservation; partnership with collecting institutions; availability to web archiving.

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

This entry was created in 2019 as a subset of a previous 2017 entry, 'Digital Music Production and Sharing,' which was split to draw attention to the different challenges faced by the different forms. While there are some overlaps with other entries relating to social media as well as those relating to community-generated content, it is a separate entry to emphasize the context in which music is shared and enjoyed; this context could be lost if attention were on products controlled by studios or artists. The 2020 Jury agreed that risks remained on the same basis as before, with no change to

trend. The 2021 Jury discussed content increasingly being shared across multiple platforms, which is both good and bad for risk. A multi-platform nature provides an element of protection against total loss, but the role and type of interaction with the content on each platform are also important and expanding with limited attempts at preservation. For these reasons, the 2021 trend moved towards greater risk with the need for selective approaches based on the increasing volume of material. The 2022 Taskforce noted no change to the that trend (they agreed these risks remain on the same basis as before).

The 2023 Council agreed with the Endangered classification with the overall risks remaining on the same basis as before ('No change' to trend).

The 2024 Council agreed that these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

The ephemera is increasingly stored on websites that themselves are fragile and are removed, and nothing held on these services can be relied on in archival timeframes.

Web archiving and social media archiving have matured, so a representative sample is probably readily available for particular countries which are more mature in their digital preservation activities as opposed to other countries which are not.

This entry also connects to other entries 'Consumer Social Media Free at the Point of Use,' 'Data Posted to Defunct or Little-used Social Media Platforms'. There are similarities in regard to increased uncertainty around major social media sites, such as X (previously Twitter), and the preservation risks associated with underlying social media preservation which has an impact on the digital objects that fall under this entry, however this entry draws attention to additional risks associated with preservation of the digital forms and contexts in which these materials are shared and enjoyed.

Case Studies & Examples

- The case of MySpace's removal of MP3s demonstrates a big loss of shared digital recordings through the platform and subsequent recovery efforts by an academic group and Internet Archive. See **MySpace admits losing 12 years' worth of music uploads**, Kleinman, Z., (2019), *BBC News* https://www.bbc.co.uk/news/technology-47610936 [accessed at 2023-10-24].

- The case of MySpace's removal of MP3s demonstrates a big loss of shared digital recordings through the platform and subsequent recovery efforts by an academic group and Internet Archive. See **The Myspace Dragon Hoard (2008-2010)**, Sketch the Cow, (2019), *Internet Archive* https://archive.org/details/myspace_dragon_hoard_2010 [accessed at 2023-10-24].

- The case of 'Yahoo Groups' closure serves to underline the fragility of community content hosted by third parties. See **Farewell Yahoo Groups! Shutting down on December 15, 2020**, Brinkmann, M., (2019), *gHacks Technology News* https://web.archive.org/web/20230902153115/https.

Digital Recordings Published via Cloud-based Music Sharing Platforms

Music licensed and playable through corporate platforms protected by rights management and subscription revenues and presented as compressed single-track recordings.

ENDANGERED

Categories

- Sound and Vision

- Cloud

Examples

Spotify, iTunes, Bandcamp, SoundCloud.

Imminence	Effort
●●●●○	●●●○○
Action is recommended within twelve months. Detailed assessment is a priority.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of skills, commitment or policy from corporate owners; conflating backup with preservation; loss of original multitrack recordings; poor planning and roadmap for corporate infrastructure; slapdash procurement or migration to new systems; mergers and acquisitions; profusion of corporate systems; uncertainty over IPR or the presence of orphaned works; single point of failure; technical protection measures that inhibit preservation actions; encryption.

Mitigations

Strong backup and documentation; use of open formats and open source software; data management planning for preservation; licensing that enables preservation; corporate preservation capability; resilient to hacking; authenticity and integrity managed; recognition of preservation functions at executive level; technology watch; regular preservation audits; accreditation and participation in the professional preservation community.

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

This entry was previously under the 2017 'Digital Music Production and Sharing' entry until it was split by the 2019 Jury into four subsets, recognizing the different challenges faced. It is particularly concerned with the music industry at scale and the services that connect the vast majority of artists to their audiences. These are typically large and well-funded, and typically recognize the value of the content they publish. But this is not without risks. It is perhaps surprising that the music industry does not yet have any equivalent to the non-print legal deposit regime that applies to other types of publication, including sheet music in some jurisdictions. The 2020 Jury agreed with the risk classification with no change to trend. The 2021 Jury noted a large amount of vulnerable material on user-driven platforms where material can be very ephemeral (removals resulting from, e.g., account deletion, space limitations, copyright claims) and the issue of licensing with the instability of the business model. For this reason, the scope was widened to include ad hoc sharing so that the entry broadly included all platforms such as SoundCloud and Bandcamp, which are more community-driven, as well as Spotify; this resulted in a raised classification from Vulnerable to Endangered with 2021 trend towards greater risk. The 2022 Taskforce agreed with the classification, and that risks remain on the same basis as before, with no change to trend.

The 2023 Council agreed with the Endangered classification and overall risks remaining on the same basis as before ('No change' to trend) but also noted an increase in the imminence of action needed for capture given instabilities and ephemeral aspects.

The *2024 Interim Review* concluded that these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

The preservation of recorded music is one of our generation's most important jobs, but it is unclear where responsibility lies. There are commercial incentives to do so, but also incentives to reduce costs. Whilst public archives are permitted to keep this material in some jurisdictions, they typically do not have the resources to do so. Consequently, there is an expectation that rights holders will maintain their own archival copies but may not do so. National collecting organizations may need to develop a role to address this.

If managed well, there is hope. It may not be an issue in the cases where the production company would hold original recordings and, if a streaming service lost a track (e.g., Spotify), they would go to the production company and ask for a copy. However, it is an issue for those outside of production companies and platforms such as SoundCloud and Bandcamp, which are more community-driven.

Adobe Flash Animations and Interactive Applets

Animations, games, and other interactive applets created with Macromedia Adobe Flash Player and Shockwave Flash, along with their accompanying websites. These are primarily .swf files, but they can also include networked collections of .swf files and external assets, as well as the web pages where they are displayed.

Categories

- Web

Examples

Flash and Shockwave based games; cartoons; interactives.

Imminence	Effort
●●●●●	●●●●●
Immediate action necessary. Where detected, should be stabilized and reported as a matter of urgency.	Loss seems inevitable. Loss has already occurred or is expected to occur before tools or techniques develop.

Hazards

No longer supported by web browsers.

Mitigations

Migration plan initiated; supported in multiple browsers; security vetted; emulation pathway; clear licensing that enables preservation.

Bit List History

Added to list: 2019

Trends

- **2024:** No change.

- **2023:** No change.

Last Review

This entry was added in 2019. The 2019 Jury noted that Flash animations and applets were a mainstay of interactive web design from the late 1990s. Flash animations and interactives were created using tools supplied by the Adobe of the same name. Although Flash enabled the development of sophisticated interaction at low cost over the web, it had a chequered history in terms of browser support and has been plagued by security concerns. The 2020 Jury identified a 2020 trend toward greater risk based on the indication by Adobe that there would be a withdrawal of support to Flash Animation. The 2021 Jury noted that the discontinuation and withdrawal of support did indeed occur. Flash is no longer supported, and loss has occurred with Adobe's deprecation of Flash and lack of support in modern web browsers. For this reason, the classification moved from Critically Endangered to Practically Extinct with a 2021 trend towards greater risk given the loss of dependence on Flash, which resulted in new aggravating conditions for migration pathway, emulation pathway, source code, lack of capacity or motivation to support, no commercial interest.

The 2023 Council agreed with the Practically Extinct classification and noted an increase in imminence of action required with greater inevitability of loss.

The *2024 Interim Review* concluded that these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

Flash represents a significant amount of the creativity of websites in the early 2000s including net-based art and cartoons. With extinction, archives will need to consider if it is possible to preserve interaction through the development of new web archiving and emulation tools and techniques. These are culturally significant artefacts or so of the web and were a gateway for many early game developers given the ease of which Flash games could be made.

It's important to note there are quite a number of community projects working on this, whilst the focus tends to be on Flash games there is still work around Flash animations. Projects like Flashpoint Archive do have their own workflows for getting the games/animations but have their own preservation issues. The survival of Flashpoint is reliant on a small group of people with the storage space entirely relying on one person maintaining it. Other Flash game archives exist as well, such as Flash Game Archives. The Internet Archive has a number of Flash game collections as well. There is also work being done around emulating Flash within web browsers through Ruffle which increases access to Flash games and animations.

Case Studies & Examples

- **Flashpoint Archive**, BlueMaxima, *https://flashpointarchive.org/* [accessed at 2023-10-24].

- **Flash Game Archive**, *http://www.flashgamearchive.com/* [accessed at 2023-10-24].

- **Adobe Flash Games Archive**, (2021), *Internet Archive https://archive.org/details/adobe-flash-games-archive* [accessed at 2023-10-24].

- Ruffle is a Flash Player emulator which is used by a number of websites that were known for Flash games and animations including Newground, Armor Games, Cool Math Games and Kongregate See **Ruffle**, *https://ruffle.rs/* [accessed at 2023-10-24].

- **Around for Decades, Gone in a Flash: How we dealt with Flash objects and the National Archives of the Netherlands**, van Veenendaal, R., Wijsman, L., Takema, J. and Rappard, M, (2023), *iPRES 2023 Conference, Urbana-Champaign, Illinois, USA, 19–22 September https://www.ideals.illinois.edu/items/128299* [accessed at 2025-07-30].

- **How Flash Games shaped the video game industry**, Richner, J., *https://www.flashgamehistory.com/* [accessed at 2023-10-24].

- **Software Library: Flash**, Scott, J., (2020), *Internet Archive https://archive.org/details/softwarelibrary_flash?tab=collection* [accessed at 2023-10-24].

iii. Content

a) Data Types/File-Formats

3D Digital Engineering Drawings

3D digital engineering models produced as part of building or engineering design processes. The production of such drawings has progressed from a digital analogue of paper to complex digital environments such as BIM (Building Information Modelling) which combine original drawings, libraries of compound objects, and links to external data sets such as about the performance of materials and maintenance of parts.

Categories

- Engineering Data

Examples

Building Information Management (BIM), Computer Aided Design (CAD), Product Data Management in engineering and architecture.

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of skills, commitment or policy from corporate owners; lack of preservation mandate or collecting institution; lack of preservation capability in data owner; irregularities in supply chains; complex or long data supply chains; dependencies on proprietary software or formats; lack of persistent identifiers; uncertainty over IPR or the presence of orphaned works; temporary joint-venture companies; poor records management; poor regulatory compliance; encryption.

Mitigations

Well managed data infrastructure; preservation from the point creation; carefully managed IPR; persistent identifiers; well managed records management processes; recognition of preservation requirements at highest levels; strategic investment in digital preservation; host clearly identified; participation in the digital preservation community.

Bit List History

Added to list: 2017

Trends

- **2024:** No change.
- **2023:** No change.

Last Review

2023 Review

This entry was first submitted in 2017 when the Jury lacked the capacity to consider it in detail. In 2019 it was assessed with additional expertise co-opted, with the decision to remain a very broad category, including major one-off construction and engineering projects, a long tail of more minor building programmes, and large volume but homogeneous production processes in engineering. The 2021 Jury agreed with its Endangered status with an identified 2021 trend towards greater risk. They noted the key consideration that the lifecycle of the products and the data that describes them vastly exceeds the short life cycles of the infrastructures on which they are designed. This challenge is compounded by supply chains that may involve many different stages of production, as well as the delivery of large projects through transitory joint venture companies that have no residual mechanism or capacity to preserve the data thereafter. Although there had been advancements in the development of new preservation tools and techniques for these materials, there were recent examples of the loss of 3D architectural drawings; those have had a huge impact, especially at the local level, as well as significant impacts on infrastructure, travel, and how people interact with built environments throughout the world. The 2022 Taskforce agreed these risks remained on the same basis as before ('No change' to the 2022 trend).

The 2023 Council agreed with the Endangered classification and seconded the trend reported last time; that risks continue on the same basis as before ('No change' to 2023 trend). Most of the complexities of the format and issues remained the same, such as reliance on proprietary software and complex or unknown copyrights with the datasets. Moving forward, it was highlighted by the Council that there needs to be a greater focus and understanding on the long-term preservation of these outputs within the sector.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additionally, some Council members pointed to where there may be overlaps between the 3D Design Engineering Drawings and Formats species groups for this entry, in particular in ways to further address Step, PDF/E, STL, 3DM, etc.

Additional Information

Data in this category enables the safety and security of critical infrastructure, but the responsibility to maintain data is unclear, nor are retention periods clear. Although examples of good practice exist, the extent to which there are working solutions at large seems doubtful, and it is surprising that there are not more diverse success stories to report.

Case Studies & Examples

- The Grenfell Tower Inquiry offers a case to consider how the loss of 3D Digital Engineering Drawings can have a huge impact, especially at the local level. For example, if Grenfell had been done using 3D technologies, do we have confidence that those materials would have been adequately preserved? What would have been the local impact? What would have been the impact on the inquiry? See **Grenfell Tower Inquiry**, https://www.grenfelltowerinquiry.org.uk/ [accessed at 2023-10-24].

- In 2006, it was reported that the Airbus A380 was 2 years behind schedule due to different offices using different versions of the CATIA CAD/CAM software. See **What can go wrong when you give IT the large**, Shelly, T. , (2006), *Manufacturing Management*

https://www.manufacturingmanagement.co.uk/content/features/what-can-go-wrong-when-you-give-it-the-large/ [accessed at 2023-10-24].

- The DPC Design and Construction Records technology watch report, which aims to support archival professionals as well as active designers and facilities managers, considering acquisition, preservation, and access approaches that account for both the technical and cultural components of the broad range of born-digital design and construction records created throughout the course of designing, building, and maintaining a built space. As well as bringing together a helpful summary of relevant work in this area and discussing a range of case studies it also covers the concept of visual digital literacy which is the first step towards understanding and managing this content. See **Preserving Born-Digital Design and Construction Records**, Leventhal, A, and Thompson, J., (2021), *Digital Preservation Coalition* http://doi.org/10.7207/twr21-01.

- The Library of Congress had a symposium on 3D design and assets in 2017. See **Designing the Future Landscape: Digital Architecture, Design & Engineering Assets**, Leventhal, A. , (2018), *Library of Congress* https://www.loc.gov/preservation/digital/meetings/DesigningTheFutureLandscapeReport.pdf [accessed at 2023-10-24].

Digitally Published Sheet Music

Sheet music licensed and published in various digital formats, subject to copyright restrictions and often protected by digital rights management technologies.

Categories

- Sound and Vision

Examples

This entry includes all manner of published sheet music, including choral works, orchestral works, scores published in different forms such as PDF, PDF/A and XML.

Imminence	Effort
●●●○○	●●○○○
Action is recommended within three years, detailed assessment within one year.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Encryption; uncertainty over IPR or the presence of orphaned works; uncertain business model of publisher; lack of legal deposit mandate.

Mitigations

Rights management conducive to preservation; held in a trusted repository; legislation in place such as legal deposit enabling copying.

Bit List History

Added to list: 2019

Trends

- **2024:** No change.
- **2023:** No change.

Last Review

2023 Review

In 2019, this entry was added as a subset of a previous 2017 entry, 'Digital Music Production and Sharing,' which was split to draw attention to the different challenges faced by the different forms. This entry focuses on digitally published sheet music, which often takes the form of PDF and PDF/A, and E-Book formats. The 2021 Jury made no changes to this entry classification or trend but did comment on the importance of legislation (such as legal deposit) to manage associated risks. The 2022 Taskforce noted no change to the trend (they agreed these risks remain on the same basis as before).

The 2023 Council agreed with the Endangered with the overall risks remaining on the same basis as before ('No change' to trend) though noted an increase in imminence of action.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

The commercial value of these materials should be a protection against their loss, and the history of the music industry indicates that sheet music continues to have value, so even if an individual organization fails or its DRM servers go offline, and some music becomes inaccessible, it is not lost. However, the difficulties of archiving DRM-locked files remain real.

Legacy Video Files

Video files in any format containing moving pictures and sound recordings, particularly those that are proprietary, contain or utilize encrypted Digital Rights Management (DRM) or carrier bound.

Categories

- Sound and Vision
- Formats

Examples

STARDIVA; AVI; MOV; MKV; MP3; MP4; on DVD or other carriers.

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of replication; encryption; digital rights management; Uncertainty over IPR or the presence of orphaned works; proliferation of file formats; weak or non-existent technical documentation; lack of preservation capability or commitment; poorly managed or digitization processes or QA; reliance on encoding/decoding software; Uncertainty over IPR or the presence of orphaned works.

Mitigations

Effective replication; normalization of file formats; strong technical documentation; preservation pathway; good descriptive cataloguing; clear licensing to enable preservation; trusted repository.

Bit List History

Added to list: 2019

Trends

- **2024:** No change.
- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 under 'Video files' to emphasize the issues of video preservation that pertain to offline recording, whether from broadcast, film industry, institutional and private collections. The 2019 Jury noted the connections between this entry and others relating to social media but argued for a standalone to emphasize the range of issues tied to numerous formats and standards. The 2021 Jury discussed the need for further rescoping, arguing that the entry was too broad to be useful without specifying at-risk types or formats. For this reason, its scope was narrowed to legacy videos that are proprietary, encrypted or carrier-bound. The classification remained Endangered with a 2021 trend towards greater risk given the growing content of at-risk legacy video files but a limited mandate. The 2022 Taskforce noted no change to the trend (they agreed these risks remain on the same basis as before).

The 2023 Council agreed with the Endangered with the overall risks remaining on the same basis as before ('No change' to trend). Additionally, they agreed that a submitted Bit List nomination for the NSV-based STARDIVA storage format would provide a good example of a video file format especially at risk due to aggravating conditions rather than a separate stand-alone entry.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Council members in the Formats species groups recommended further rescoping of overlaps, in particular to differences between formats and physical carries mentioned, which have different risks

and issues. The listing of DVDs or other carriers, for example, may be better suited under the Portable Media species.

Additional Information

There are simply too many formats and too many standards, but the FFMPEG project and its related tools have significantly mitigated the technical risk to most video files. This enables a practitioner to transform the vast majority of file formats to safer preservation formats while retaining significant properties. However, technical risk is only one of the factors. There needs to be institutional engagement with audio-visual data as a priority. The issue then becomes one of identifying the organizations responsible and, constrained by the cost to store video data, making effective selection decisions.

Case Studies & Examples

- The NSV-based STARDIVA storage format is a video format with multiple audio streams used in simultaneous translation session recording by agencies such as the UN. It is a proprietary format that is no longer supported, cannot be natively preserved and cannot be viewed correctly using standard video playback tools. As noted by the nominator, given its use by agencies such as the UN, the loss of this format would be a loss of an international record. The nominator added that by using MediaArea LeaveSD it can be partially normalized for preservation purposes. See **LeaveSD**, MediaArea, (2022), https://mediaarea.net/LeaveSD [accessed at 2023-10-24].

Virtual Reality Materials and Experiences

Virtual reality (VR) refers to a set of technologies which build on existing 3D rendering technologies, with the aim of creating experiences which completely immerse a user in a virtual environment. The related term of Immersive Media (also known by the acronym XR) refers to a set of technologies used to create experiences, which either completely immerse a user in a virtual environment (Virtual Reality), augment the real world with virtual elements (Augmented Reality) or combine elements of the two (Mixed Reality). Key technologies include headsets, tracking systems, real-time 3D software and 360 video.

Categories

- Media Art

Examples

Oculus Rift, VR tours, art installations.

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Lack of skills, commitment or policy from corporate owners; lack of established frameworks and tools; technology is relatively poorly understood in the digital preservation domain; many of the technologies are proprietary; technology is seen as inherently fragile and therefore risky to collect and preserve; emulators do not currently support XR applications; expected to be difficult and costly to migrate, a process itself dependent on access to vulnerable source materials; uncertainty over IPR or the presence of orphaned works.

Mitigations

Effective replication; emulation; strong technical documentation; preservation pathway; good descriptive cataloguing; trusted repository.

Bit List History

Added to list: 2021

Trends

- **2024:** No change.
- **2023:** No change.

Last Review

2023 Review

This entry was added in 2021. VR technologies are finding use in many sectors, including archaeology, architecture, contemporary art, documentary film, gaming, forensics, science and engineering. While these technologies are not new per se, having experienced a first wave in the 1990s, they have experienced renewed interest recently as a result of a new generation of hardware. There are overlaps with other entries relating to both Media Art and Gaming, but it was added as an entry to emphasize the issues of preservation that pertain to the interconnected set of specific hardware and software components that access to XR experiences is contingent on. VR is challenging to document due to the individual nature of the experience, and components tend to become rapidly obsolete due to a fast rate of technological change as the industry pushes newer, higher fidelity hardware and software. This results in the potential to lose access to XR software applications, as old VR applications can no longer communicate with new XR hardware. The reliance on proprietary software and hardware components, as well as the lack of industry standards, poses a further risk. The 2022 Taskforce agreed that the risks remained on the same basis, with no change to trend.

The 2023 Council agreed with the classification of Endangered with overall risks remaining on the same basis as before ('No change' to trend).

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

The current wave of materials made using XR technologies represent a unique point in time for the continued development of the technology and therefore represent a significant piece of computing history. Individual materials/experiences created using XR technologies present their own

significance beyond this, which, noted elsewhere in this entry, can be represented in a wide range of sectors.

The impacts of the loss of access to virtual reality materials could be widely felt, given their wide-ranging uses across many sectors — most notably collections and archives containing materials accessed using these technologies. Simultaneously there is a risk of a loss of understanding of this technologies' development during the 2010-present period, which is likely to be of historical significance in and of itself.

Media artworks are often made with a network of knowledge that can be precarious. Documentation around production processes can be minimal, and hence acting quickly with known processes can gather information before the knowledge and people networks start to disperse. This can mean preservation of production environments and associated workflows can be preserved alongside the media.

Case Studies & Examples
- The Preserving Immersive Media Knowledge Base resource, created to help share information between members of the digital preservation community who are caring for virtual reality (VR), augmented reality (AR), mixed reality (MR), 360 video, real-time 3D software and other similar materials. This site was born out of Tate's Preserving Immersive Media Project with funding from the Netherlands Institute for Sound & Vision. See **Preserving Immersive Media Knowledge Base**, https://pimkb.gitbook.io/pimkb/ [accessed at 2023-10-24].

- Resources and outputs from the Preserving and Sharing Born Digital and Hybrid Objects From and Across The National Collection project. See **Preserving and Sharing Born Digital and Hybrid Objects**, V&A Research Projects, https://www.vam.ac.uk/research/projects/preserving-and-sharing-born-digital-and-hybrid-objects [accessed at 2023-10-24].

- **Preserving and sharing born-digital and hybrid objects from and across the National Collection**, Arrigoni, G., Kane, N., McConnachie, S., and McKim, S., https://doi.org/10.5281/zenodo.7097278.

- The Tate Preserving Immersive Media project, which is developing strategies for the preservation of artworks which utilize immersive media such as 360 video, real-time 3D, virtual, augmented and mixed reality. See **Preserving Immersive Media**, Tate, https://www.tate.org.uk/about-us/projects/preserving-immersive-media [accessed at 2023-10-24].

- **Virtual Reality is a big trend in museums, but what are the best examples of museums using VR?**, Richardson, J., (2023), *Museum Next* https://www.museumnext.com/article/how-museums-are-using-virtual-reality/ [accessed at 2023-10-24].

- The issues and approaches raised by the Tuvalu Future Now Project, a set of three major initiatives designed to preserve its nationhood, governance and culture in the event of a worst-case scenario. The third initiative is the development of a digital nation. It includes digitising and transferring access to government and consular services and all accompanying administrative systems into the cloud to enable elections to continue to be held, and government bodies to continue in their roles. It also includes a virtual copy of Te Afualiku, the first island in Tuvalu to be digitally recreated through satellite imagery, photos and drone footage, creating a digital twin to not only help inform decisions around urban planning and development but also examine how to use augmented and virtual reality to allow displaced

and future generations of Tuvaluans to continue to exist as both a culture and a nation, complete with ancestral knowledge and value systems. If this concept becomes a reality, the Tuvaluan people will be able to interact with one another in a digital dimension, in a way that imitates real life and helps to preserve shared language and customs. See **Facing extinction, Tuvalu considers the digital clone of a country**, Fainu, K., (2023), *The Guardian* https://www.theguardian.com/world/2023/jun/27/tuvalu-climate-crisis-rising-sea-levels-pacific-island-nation-country-digital-clone.

- The DPC 'Preserving Digital Art' Technology Watch Guidance Note is aimed at institutions starting to collect digital art as part of a wider collecting remit. It offers basic guidance on the specificities of digital art and how it may differ from other digital content in an institution's care. See **Preserving Digital Art**, Falcão, P., (2024), *Digital Preservation Coalition* http://doi.org/10.7207/twgn24-02.

- **Preserving Virtual Reality Artworks**, Ensom, T., and McConchie, J. , (2021), https://web.archive.org/web/20240830121113/https://zenodo.org/records/5274102.

- NEW MEDIA MUSEUMS: Creating Framework for Preserving and Collecting Media Arts in V4, initiated by the Olomouc Museum of Art as a joint international platform for sharing experience with building and maintaining collections of new media artworks across different types of institutions. The aim of the project is to find workable methods for heritage institutions to build and maintain collections of media arts, which are necessary for safeguarding this area for the benefit of society. See **NEW MEDIA MUSEUMS: Creating Framework for Preserving and Collecting Media Arts in V4**, Central European Art Database, (2021), http://cead.space/Detail/projects/3797 [accessed at 2023-10-24].

- The Collaborative Infrastructure for sustainable access to digital art LIMA project, to prevent the loss of digital artworks and to commonly develop the knowledge to preserve these works in a sustainable way. The project 'Infrastructure sustainable accessibility digital art' invests in research, training, knowledge sharing and conservation to prevent the loss of both digital artworks and the knowledge to preserve them. See **Collaborative infrastructure for sustainable access to digital art**, LIMA, https://www.li-ma.nl/lima/article/collaborative-infrastructure-sustainable-access-digital-art [accessed at 2023-10-24].

PDF

Documents presented in PDF (Portable Document Format) format (ISO 32000:1 and ISO 32000:2) and other data wrapped inside them, including all variants and versions, including PDF/A.

Categories

- Formats

Examples

Documents stored offline, or online in repositories or EDRMS, including reports, agenda, minutes, correspondence, contracts, essays, articles, or research papers, PDF 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7 and 2.0. PDF/A, PDF/X and PDF/E.

	Imminence		**Effort**

●○○○○	●●○○○
Action is recommended as required, with periodic review every five years.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Lack of skills, commitment or policy from corporate owners; loss of context; loss of authenticity or integrity; external dependencies; poor storage; lack of understanding; significant diversity of data; poorly developed digitization specifications; lack of integrity checking; poorly developed migration or normalizations specifications; lack of virus control; poor storage or replication; lack of validation at the point of creation; encryption; uncertainty over IPR or the presence of orphaned works.

Mitigations

Well-managed data infrastructure; preservation planning; authenticity managed; use of persistent identifiers; reduction of dependencies; application of records management standards; recognition of preservation requirements beyond formats; strategic investment in digital preservation; preservation roadmap; clear licensing to enable digital preservation and deposit in a trusted archive; participation in digital preservation community; format validation; version control.

Bit List History

Added to list: 2017

Trends

- **2024:** No change.
- **2023:** No change.

Last Review

2023 Review

A PDF entry was added in 2017 and was split into two entries, 'PDF/A' and 'PDF other than PDF/A', in 2019 to emphasize the different threats faced by different types of PDF. The 2021 Jury agreed with this decision and noted that trends for the PDF other than PDF/A entry and the PDF/A entry were both towards a reduced risk. The 2022 Taskforce agreed these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

The 2023 Council recommended merging the two previously split entries of 'PDF/A' and 'PDF other than PDF/A'. After reviewing the two entries separately, they found more similarities than differences between the two and, indeed, across all types of PDF (not just PDF/A). Due to the level of commercial, open-source tools that are available to assist preservation, the risk of loss was less persistent than previously suggested. Therefore, a Vulnerable classification was assigned as the most appropriate for all PDF formats as a whole.

2024 Interim Review

The 2024 Council agreed these risks remain on the same basis as before ('No change' to trend).

Additional Information

There is a lot of material produced and kept in PDF. Some of it is authoritative, in other words, the only available copy, while some of it is not. However, if it is the only copy and it is lost, it can have an impact on a lot of people

The challenge in evaluating the significance and impact of the loss of PDFs is that they're quite often a surrogate of something else, whether a digitized record or a Word document, etc. Whether or not that record is retained may be a factor. We should also be considering PDF Portfolios, which are an extension of PDF 1.7. Portfolios contain embedded files and can include text documents, spreadsheets, PowerPoints, emails, Computer Aided Design (CAD) drawings.

Vulnerability also depends on if the PDF file conforms to the specific PDF/A standard or not. This is caused by a combination of 1) not conforming to the standard and 2) collection managers assuming that the file is resilient simply because it purports to be a PDF/A. This risk is less with the format and more with the understanding and experience in data management. Moreover, materials embedded in or attached to PDF/A-2 and PDF/A-3 may be at risk.

Case Studies & Examples

- **Preserving with PDF/A (Second Edition)**, Fanning, B. , (2017), *Digital Preservation Coalition* http://doi.org/10.7207/twr17-01.

Published Research Papers

Completed research papers published in serials, monographs or theses which fall under specific collecting policies of research libraries or archives and are managed through dedicated repository infrastructures.

The Bit List
VULNERABLE
Digital Species

Categories

- Research Outputs

Examples

Published research papers in scholarly E-Books and Electronic Journals; Electronic theses (E-theses)

Imminence	Effort
●●●○○	●●○○○
Action is recommended within three years, detailed assessment within one year.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Lack of skills, commitment or policy from publishers; uncertainty over IPR or the presence of orphaned work; embedded complex objects; unstable funding for repository; lack of strategic investment; complex external dependencies; lack of persistent identifiers; bespoke formats; lack of legal deposit mandate

Mitigations

Strong documentation including intellectual property rights; clarity of preservation path and ensuing responsibilities; credible preservation plan; proven capacity of repository; legal deposit preservation

copying; post-cancellation access service; persistent identifiers used consistently; non-proprietary formats used and validated; minimal or well managed external dependencies.

Bit List History

Added to list: 2017

Trends
- **2024:** No change.
- **2023:** No change.

Last Review

2023 Review

This entry was added in 2017 under 'Published research outputs,' though without reference to the capacity of the repository infrastructure. The 2019 Jury amended it to presume the existence of repository infrastructure and noted that the aggravating conditions (which introduce risks) and good practice enhancements (which reduce it) are most relevant to repository operations.

While the 2020 Jury found no change in trend, the 2021 Jury agreed it should remain Vulnerable and discussed improvements and initiatives towards the preservation of research data and outputs, pointing to a 2021 trend towards reduced risk. The 2022 Taskforce agreed risks were on the same basis as before (no change to the trend).

The 2023 Council agreed with the Vulnerable classification and risks remained on the same basis as before ('No change' to trend), also noting a slight decrease in imminence of action with no significant trends towards greater or reduced risk. Additionally, the 2023 Council recommended that a nomination received for a new 'E-theses' entry would provide a valuable example to this entry rather than as a new, standalone entry. The 2023 Council recognized that further scoping and input are needed for this entry and recommended that the next major review revisit and restructure the entry.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

The 2023 nomination for E-theses highlights distinct risks tied to these digital published materials. E-theses tend to be sole documents which when published by universities may get harvested into other aggregators or resources but in many cases the only copy (with no physical/analogue copy) sits on an Institution's repository. In addition, many are deposited in PDF format (of many varieties and many don't even attempt to use PDF/A etc.) risking long term accessibility and re-use. However, the breadth of risks goes beyond just the PDF variety, as e-theses often include databases, audiovisual materials, websites, and more.

The loss of tools, data or services within this group would impact on people and sectors around the world. Particularly those involved with reproducibility and those wishing to use the datasets for further research.

Although there have been improvements in current practice, policies and workflows, there is still a significant corpus of information that was deposited before these improvements came into force. It

is unlikely that there will be the time, will or resources to bring this information up to current standards.

Case Studies & Examples

- A recent analysis from Martin Eve of CrossRef shows scholarly content at risk. The findings, based on the assessment of around 7.5 million of the e-books and articles for which CrossRef provides a fixed identifier or Digital Object Identifier, suggest that around a quarter of academic publications are not being preserved for the future. For c. 2 million articles in the study there were no evidence of them being preserved, and 4.3 of works studied were preserved in at least one place. See **Digital Scholarly Journals Are Poorly Preserved: A Study of 7 Million Articles**, Eve, M. P. , (2024), *Journal of Librarianship and Scholarly Communication* https://doi.org/10.31274/jlsc.16288.

- **Breaking down barriers in e-only thesis submission: how digital preservation contributes to the conversation at the University of Glasgow**, Konstantelos, L., (2021), *Digital Preservation Coalition Blog* https://www.dpconline.org/blog/wdpd/wdpd2021-konstantelos [accessed at 2023-10-24].

- **From "research output'' to "research data'' - a willingness to move forward?**, Klungthanaboon, W., (2021), *Digital Preservation Coalition Blog* https://www.dpconline.org/blog/wdpd/research-output-to-research-data [accessed at 2023-10-24].

- **Preservation, Trust and Continuing Access for E-Journals**, Beagrie, N., (2013), *Digital Preservation Coalition* http://doi.org/10.7207/twr13-04.

- **Preserving E-Books**, Morrissey, S, and Kirchhoff, A., (2014), *Digital Preservation Coalition* http://doi.org/10.7207/twr14-01.

- Resources and recent outputs from Public Knowledge Project (PKP) Preservation Network, which developed to digitally preserve Open Journal Systems (OJS) journals. See **PKP Preservation Network**, Public Knowledge Project, https://pkp.sfu.ca/pkp-pn/ [accessed at 2023-10-24].

---

iv.   **Technological**

a)   **Hardware/Media**

Non-current Hard Disk Technologies

Materials saved to storage devices with a variety of underlying magnetic or solid-state technologies that are hardwired into a computer that is no longer under warranty or supported: typically, hard disks more than five years old.

Categories

- Integrated Storage

Examples

Disks installed into computers or servers that are more than five years old, or out of warranty

Imminence	**Effort**
●●●○○	●●●●●
Action is recommended within three years, detailed assessment within one year.	Loss seems inevitable. Loss has already occurred or is expected to occur before tools or techniques develop.

Hazards

Lack of replication; poor storage; non-standard connections or controllers; aggressive compression; encryption; Uncertainty over IPR or the presence of orphaned works

Mitigations

Maintenance schedule; renewable extendable warranty; best practice storage and operation; replication

Bit List History

Added to list: 2019

Trends

- **2024:** No change.
- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 to ensure that the range of media storage is properly assessed and presented. The lifecycles of most consumer hard disk technology are relatively stable in comparison to portable devices because they are integrated into systems and therefore inherit the lifecycle and replacement of the entire system. This is less true at scale; however, where disks are used in storage arrays, and refreshment is more loosely tied to the server architecture. Storage at scale also means the percentage likelihood of finding a disk failure increases, and this likelihood of failure led to the 2021 Jury's noted trend towards greater risk. It was reviewed in 2022 with no noted change towards even greater or reduced risk.

The 2023 Council agreed with the current Critically Endangered classification with overall risks remaining on the same basis as before ('No change' to trend), while also noting a greater inevitability of loss from the discontinuation of support and development for these storage technologies when compared to the 2021 Jury review.

2024 Interim Review

The 2024 Council agreed These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

A lot of early PCI-E flash devices (e.g. Fusion-IO) used proprietary drivers before the NVME standard was developed, but are now dropping off support. Intel has stopped development of Optane non-volatile RAM, some of which required specific CPU support to access although that form was usually used for data caching rather than storage.

Accessing drives with pre-SATA interfaces is increasingly difficult since interface cards and OS support can be hard to come by.

The greater density of newer disks, as well as encryption and compression, mean they can be more fragile than older disks with less density, and less sophisticated read/write technologies. The age of a disk is not the best or only indicator of its reliability.

Case Studies & Examples

- **Music industry's 1990s hard drives, like all HDDs, are dying**, Purdy, K. , (2024), *Ars Technica* https://arstechnica.com/gadgets/2024/09/music-industrys-1990s-hard-drives-like-all-hdds-are-dying/ [accessed at 2024-09-13].

Non-current Portable Optical Media

Materials saved to DVDs, CDs or other optical storage devices where the media is out of warranty and reader devices may no longer be supported or integrated easily into hardware infrastructure: typically, more than five years old.

Categories

- Portable Media

Examples

CDs, laserdisc technologies, DVD, HDVD, MiniDisc, Superdisk.

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Poor storage; inability to access readers; no replication; encryption; uncertainty over IPR or the presence of orphaned works

Mitigations

Active management; dependable access to readers; strong documentation; documentation independent from the media

Bit List History

Added to list: 2019

Trends

- **2024:** No change.
- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 to ensure that the range of media storage is properly assessed and presented. The 2019 Jury noted optical media is, in some senses, a preferred option as it is typically more stable than magnetic or solid state media, but these characteristics have been oversold. The substrates of the disks can prove unstable, and more importantly, the reader technology tends to be proprietary and can become obsolete long before the disks degrade. Storage at scale also means the percentage likelihood of failure increases. The 2021 Jury and 2022 Taskforce agreed with the entry's assigned risk classification with no noted changes towards increased or reduced risk.

The 2023 Council agreed with the risk classification of Critically Endangered with the overall risks remaining on the same basis as before ('No change' to trend).

2024 Interim Review

The 2024 Council agreed These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

We know what to do; it is the scale of the problem. It is a big scale problem - and in many library catalogues the information about these carriers and their playing requirements do not exist in metadata. Big scale surveys of collections to identify these carriers might be a barrier to starting to tackle this one.

Optical media is less and less recoverable as time goes on. The data cannot be actively managed, the readers are starting to go out of style, and the carriers will continue to rapidly degrade. Early generations of this media are not as robust as the current generation and can deteriorate significantly if not stored appropriately. Many optical media that were created by individuals using consumer-grade hardware and software are more vulnerable than media created using commercial-grade production techniques, and often user error makes the data unreadable before bitrot and other degradation is able to begin. The presence of adhesive labels can be a big warning sign of this type of vulnerability.

Case Studies & Examples

- The British Library's Flashback project, a proof-of-concept that explored the practical challenges of preserving digital content stored on physical media (magnetic and optical disks) using a sample of content from hybrid collection items dating from between 1980 and 2010. See **The Flashback Project: Rescuing disk-based content from the 1980s to the present day**, Pennock, M., May, P., Day, M., Davies, K. and Whibley, S. , (2016), *11th International Digital Curation Conference* https://doi.org/10.5281/zenodo.1321630.

- The Library of Congress's CD-R and DVD-R RW Longevity Research project page provides summarized outcomes and findings from several different complementary studies undertaken or are continuing. See **CD-R and DVD-R RW Longevity Research**, Library of Congress, https://www.loc.gov/preservation/scientists/projects/cd-r_dvd-r_rw_longevity.html [accessed at 2023-10-24].

- The Preserving Write-Once DVDs: Producing Disc Images, Extracting Content, and Addressing Flaws and Errors analytic report by George Blood Audio Video Film (GBAVF) in which, The report was one product of a contract with the company, in which they converted a set of write-once DVDs for the Library of Congress. The report describes the issues encountered

and provides some detail about methods for carrying out the work,, offering an overview of the range and extent of the issues, as well as describing the corrective tools and processes that were used. See **Preserving Write-Once DVDs: Producing Disc Images, Extracting Content, and Addressing Flaws and Errors, Final Draft Report**, George Blood Audio Video Film, (2014), https://www.digitizationguidelines.gov/audio-visual/documents/Preserve_DVDs_BloodReport_20140901.pdf [accessed at 2023-10-24].

- **Planning ahead for DVD-Video migration research**, O'Leary, K. , (2021), *Digital Preservation Coalition Blog* https://www.dpconline.org/blog/wdpd/wdpd2019-dvd-video-migration-research [accessed at 2023-10-24].

- **Longevity of Recordable CDs, DVDs and Blu-rays, Canadian Conservation Institute (CCI) Notes, 19/1**, Iraci, J., (2019), https://www.canada.ca/en/conservation-institute/services/conservation-preservation-publications/canadian-conservation-institute-notes/longevity-recordable-cds-dvds.html [accessed at 2023-10-24].

- **Disks back from the dead**, Baker, M., (2017), *Nature* https://doi.org/10.1038/545117a.

- **The first generation of CDs is already rotting and dying**, Kirn, P., (2017), *CDM Stories.* https://cdm.link/2017/02/a-generation-of-cds-is-already-rotting-and-dying/ [accessed at 2023-10-24].

Non-current Portable Solid State Media

Materials saved to flash or other solid-state storage devices where the media is out of warranty and reader devices may no longer be supported or integrated easily into hardware infrastructure: typically, more than five years old.

Categories

- Portable Media

Examples

USB sticks and pen drives; Flash storage in cameras and phones; certain types of portable hard disk.

Imminence	Effort
●●●○○	●●●○○
Action is recommended within three years, detailed assessment within one year.	It would require a major effort to prevent or reduce losses in this group, possibly requiring the development of new preservation tools or techniques.

Hazards

Poor physical storage; inability to access readers; no replication; encryption; uncertainty over IPR or the presence of orphaned works.

Mitigations

Active management; dependable access to readers; strong documentation; documentation independent from the media.

Bit List History

Added to list: 2019

Trends

- **2024:** No change.
- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 to ensure that the range of media storage is properly assessed and presented. The 2019 Jury noted solid state media – typically flash – provides very fast access to data but can fail without warning. This is because it is typically subject to a limited number of program/erase cycles, as well as 'read/disturb' effects. Storage at scale also means the percentage likelihood of failure increases. The 2021 Jury and 2022 Taskforce agreed with the entry's assigned risk classification with no noted changes towards increased or reduced risk.

The 2023 Council agreed with the continued risk classification of Critically Endangered with the overall risks remaining on the same basis as before ('No change' to trend). They noted that many libraries, archives and other organizations are using these media as the access copy as well as the preservation copy. The data can be overwritten or lost by mishandling as well as bitrot and degradation. Data should be transferred into a digital preservation system that allows for active management.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

An additional preservation risk here is that many libraries/archives use this media as the access copy as well as the preservation copy, thus increasing the risk of the data being overwritten, lost by mishandling as well as general bitrot and degradation. To decrease the risk, data should be transferred into a digital preservation system that allows for active management.

Early generations of media and cheap giveaways are not robust and can deteriorate significantly over time.

Current Portable Magnetic Media

Materials saved to magnetic tape, portable hard disks or other magnetic media in the last five years where the reader devices are still supported and can be integrated easily into hardware infrastructure.

Categories

- Portable Media

Examples

LTO tapes; portable hard disks.

Imminence	Effort
●●●○○	●●○○○
Action is recommended within three years, detailed assessment within one year.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Poor storage conditions; encryption; digital rights management; uncertainty over IPR or the presence of orphaned works; lack of replication; lack of documentation; lack of periodic testing; lack of refreshment pathway; lack of access to readers; out of manufacturer warranty or no warranty; storage within paper files.

Mitigations

Regular review and testing; replication; refreshment plan; comprehensive documentation; high quality storage; regular maintenance of readers; multiple readers available.

Bit List History

Added to list: 2019

Trends

- **2024:** Towards greater risk.
- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 to ensure that the range of media storage is properly assessed and presented. The 2019 Jury noted magnetic media is typically more fragile than optical media because it is susceptible to 'bitrot' and magnetic damage in ways that optical media are not. The 2021 Jury commented on how the types of magnetic media used had improved the last five years, notably the use of LTO, increasing good practice and therefore identified a 2021 trend towards reduced risk in this respect. They added it is important to note that LTO tapes come in different generations. Some pose greater preservation risks now (e.g., an organization with no equipment or way of reading content), so the use of LTO is good practice so long as it includes the active management of associated risks. The 2022 Taskforce agreed the risks remained on the same basis as before (no change to trend).

The 2023 Council agreed with the risk classification of Endangered with the overall risks remaining on the same basis as before ('No change' to trend).

2024 Interim Review

The 2024 Council identified a trend towards even greater risk in consideration of how magnetic tape, and data tapes, are also very much dependent on the readers and software that is used to access and write to them.

Additionally, those in the Integrated Storage species group raised areas of potential overlap with the Portable Media species group (see 'Current Hard Disk Technologies'). As people increasingly select other storage methods, such as cloud, they are less likely to maintain existing content on portable hard disks, which means the portable hard disks are more likely to be overlooked or ignored (e.g., left in drawers) rather than checked and refreshed. Questions arise concerning hard drives and SSDs

packaged as portable devices, and for this reason further cross-species review is recommended for the next 2025 review.

Additional Information

This entry is highly dependent on who is looking after the portable media but made more difficult over time. The lack of granularity in the definition means that only general advice can be offered, such as to refresh media. In time, it may yet be more useful to split all storage media (maybe 100 items long) with an indication of how long these can be expected to last. In many cases, specialists can recover obsolete media, but the cost of employing them can become an aggravating condition.

It is important to emphasize that the short lifetime of many storage devices is not a problem to be solved with new long-lasting storage technologies (and indeed, many inventions have come and gone). Cheap commodity storage has been purposely designed to deliver value at a low price for a short time. Therefore, management and preservation processes for monitoring and refreshment need to take these characteristics into account.

Current Portable Optical Media

Materials saved to DVDs, CDs or other optical media in the last five years where the reader devices are still supported and can be integrated easily into hardware infrastructure.

Categories

- Portable Media

Examples

CDs, DVDs produced in the last five years.

Imminence	Effort
●●●○○	●●○○○
Action is recommended within three years, detailed assessment within one year.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Poor storage conditions; encryption; digital rights management; uncertainty over IPR or the presence of orphaned works; lack of replication; lack of documentation; lack of periodic testing; lack of refreshment pathway; lack of access to readers; out of manufacturers' warranty or no warranty; storage in paper files; environmental disasters; technical preservation measures.

Mitigations

Regular review and testing; replication; refreshment plan; comprehensive documentation; high quality storage; regular maintenance of readers; multiple readers available.

Bit List History

Added to list: 2019

Trends

- **2024:** No change.
- **2023:** No change.

Last Review

2023 Review

The 2019 Jury introduced this entry to ensure that the range of media storage is properly assessed and presented. The 2020 and 2020 Juries, and 2022 Taskforce, agreed with the entry's assigned risk classification with no noted changes towards increased or reduced risk (no change to trend).

The 2023 Council agreed with the risk classification of Endangered with the overall risks remaining on the same basis as before ('No change' to trend). Additionally, the Council recommended that a nomination received for 'DVDs produced under copyright' would provide a valuable example to this entry rather than as a new, standalone entry.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

This entry is highly dependent on who is looking after the portable media but made more difficult over time. The lack of granularity in the definition means that only general advice can be offered, such as to refresh media. In time, it may yet be more useful to split all storage media (maybe 100 items long) with an indication of how long these can be expected to last. In many cases, specialists can recover obsolete media, but the cost of employing them can become an aggravating condition.

Case Studies & Examples

- One example, presented to the Council through a 2023 nomination, concerns the loss of digital materials stored in DVDs. A 3,500 DVD 'film collection' was destroyed during the fire at the University Library Archives and Special Collections, University of Cape Town, South Africa. There were no backup copies in large part because of the (mis)understanding that they were prevented by copyright from making backup copies. This example extends beyond the loss of material in the optical media, particularly in respect to backups and issues surrounding copyright, but in the context of the entry shows how aggravating factors can place content at greater risk due to aggravating factors not only to copyright but also global warming with more such fires, floods and natural disasters to be expected. See **Devastating Cape Town fire destroys University Library Archives and Special Collections**, News in Conservation, (2021), https://www.iiconservation.org/content/news-conservation-issue-84-june-july-2021 [accessed at 2023-10-24].

Current Portable Solid-State Media

Materials saved to flash drives or other solid-state media in the last five years where the reader devices are still supported and can be integrated easily into hardware infrastructure.

Categories

- Portable Media

Examples

USB memory sticks; flash cards in cameras; solid state portable hard disks.

Imminence	Effort
●●●○○	●●○○○
Action is recommended within three years, detailed assessment within one year.	It would require a small effort to preserve materials in this group, requiring the application of proven tools and techniques.

Hazards

Poor storage conditions; encryption; digital rights management; uncertainty over IPR or the presence of orphaned works; lack of replication; lack of documentation; lack of periodic testing; lack of refreshment pathway; lack of access to readers; out of manufacturer warranty or no warranty; storage in paper files.

Mitigations

Regular review and testing; replication; refreshment plan; comprehensive documentation; high quality storage; regular maintenance of readers; multiple readers available.

Bit List History

Added to list: 2019

Trends

- **2024:** No change.
- **2023:** No change.

Last Review

2023 Review

The 2019 Jury introduced this entry to ensure that the range of media storage is properly assessed and presented. The 2020 and 2021 Juries, and 2022 Taskforce, agreed with the entry's assigned risk classification with no noted changes towards increased or reduced risk (no change to trend).

The 2023 Council agreed with the risk classification of Endangered with the overall risks remaining on the same basis as before ('No change' to trend).

2024 Interim Review

The 2024 Council agreed these risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

This entry is highly dependent on who is looking after the portable media but made more difficult over time. The lack of granularity in the definition means that only general advice can be offered, such as to refresh media. In time, it may yet be more useful to split all storage media (maybe 100 items long) with an indication of how long these can be expected to last. In many cases, specialists can recover obsolete media, but the cost of employing them can become an aggravating condition.

Non-current, Rare Portable Magnetic Media

Materials saved to uncommon storage devices where the media is out of warranty, reader devices may no longer be supported or integrated into hardware infrastructure, and reader devices are extremely hard to acquire due to rarity: typically, more than five years old.

Categories

- Portable Media

Examples

Bernoulli, Canon Diskfile, Superdisk, Jaz, MiniDisc, and similar

Imminence	Effort
●●●○○	●●●●○
Action is recommended within three years, detailed assessment within one year.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

N/A

Mitigations

Data can be preserved only with the ability to acquire drives and make them functional; media items must be in good working condition; original documentation can be difficult to locate, and drivers or other dependencies may be impossible to acquire; much specialized work is necessary to make drives work and transfer data; uncertainty over IPR or the presence of orphaned works.

Bit List History

Added to list: 2023

Trends

- **2024:** No change.

Last Review

2023 Review

The 2023 Council added this as a new, split entry related to the 'Non-current, Portable Magnetic Media' entry to highlight the increased risk for more unusual and less common formats associated with the media. They also noted that it is likely that more current formats will fall into this category over time, there will remain a need for use and development of forensic tools and techniques. The Council also recommended that an effort to create a comprehensive list of formats that may qualify for this category be undertaken; they suggested an open call and encouragement to the community to contribute examples to add to the entry for the next major review of the Bit List, as by their nature they are harder to identify and, by addressing those not as common, there can be further development and cross-referencing across resources (e.g. registries, technology watch, etc.).

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

A further rescoping is also recommended for the 2025 review to revisit the current splitting of the non-current portable magnetic media into two (for rare and for common). The reader may not be rare (3.5-inch floppy drive or 5.25-inch floppy drive), and it is possible to get a flux stream from this type of media but there is no way of converting that media to binary files, so essentially no way of accessing the actual content. In this way, it may not be worth splitting out, because by creating a flux stream you are essentially saving the data from further degradation on the portable media format.

Additional Information

Additionally, if we don't do further research in the special types (Such a WANG disks, Lexitron, ICL computers), that data will also be eventually lost, even if we have a flux stream. This is not really a problem for the more common types, such as IBM, which even tools like FTK Imager can convert and make accessible.

It is important to distinguish these materials from the floppy, hard drive, and other common formats for which there are still a large number of readers available and tools have been developed (FC5025, KryoFlux). These less typical, unusual or 'weird' formats were momentary and ephemeral and weren't very popular, but archival data exist on them and there are very few readers available and very few tools, if any, exist to support them. There is an overall lower impact because there are few collections on these media, relatively.

Case Studies & Examples
- **Can anyone help set up a bernoulli drive?**, Spiceworks Community , (2010), https://community.spiceworks.com/topic/504222-can-anyone-help-set-up-a-bernoulli-drive [accessed at 2023-10-24].

- **'Help needed to run a SCSI Jaz drive.**, LinuxQuestions.org, (2005), https://www.linuxquestions.org/questions/linux-hardware-18/help-needed-to-run-a-scsi-jaz-drive-318549/ [accessed at 2023-10-24].

Current Hard Disk Technologies

Materials saved to storage devices with a variety of underlying magnetic or solid-state (flash) technologies that are hardwired into a computer still under warranty or supported: typically hard disks that are less than five years old.

Categories
- Integrated Storage

Examples

Direct Attached Storage (DAS) such as magnetic or solid-state drives integrated into individual laptops or workstations and into smaller scale storage facilities

Imminence	Effort
●●○○○	●○○○○
Action is recommended within five years, and detailed assessment within three years.	Loss of material in this group could be entirely avoidable if provided the means to deploy proven tools and techniques.

Hazards

Encryption; poor handling; poor storage; lack of consistent replication; failure of external (dependencies, e.g., suppliers, security); political or commercial interference; failure of internal dependencies (e.g., power supply, disk controller); overly aggressive compression; poor information security; lack of integrity-checking; lack of strategic investment; lack of warranty; unenforceable warranty; Uncertainty over IPR or the presence of orphaned works

Mitigations

Backup to different technology; backup to diverse locations; documentation of assets; integrity checking; preservation planning; refreshment planning; export functionality; resilient to hacking; selection and appraisal criteria; version control; resilient funding; technology watch; enforceable warranty; disaster planning

Bit List History

Added to list: 2019

Trends
- **2024:** No change.
- **2023:** No change.

Last Review

2023 Review

This entry was added in 2019 to ensure that the range of media storage is properly assessed and presented. It was reviewed in 2021 with a noted trend towards greater risk in light of the continued shift towards reliance on cloud storage with computers increasingly reducing hard disk for solid-state storage and commercial motivations for less support, and reviewed in 2022 with no noted increase in trend towards even greater or reduced risk.

The 2023 Council agreed with the current Vulnerable classification, with overall risks remaining on the same basis as before ('No change' to trend), while also noting a slight decrease in the effort needed to preserve and the imminence of action required when compared to the 2021 Jury review.

2024 Interim Review

The 2024 Council agreed These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

There were also noted areas of overlap with the Portable Media species group (See: 'Current Portable Magnetic Media'). As people increasingly select other storage methods, such as cloud, they are less likely to maintain existing content on portable hard disks, which means the portable hard disks are more likely to be overlooked or ignored (e.g., left in drawers) rather than checked and refreshed. Questions arise concerning hard drives and SSDs packaged as portable devices, and for this reason, further cross-species review is recommended for the next 2025 review.

Additional Information

There are also indications of increasing prevalence of soldered-in flash storage which cannot easily be accessed in the case of device failure.

Case Studies & Examples

- Some new technologies like shingling, HAMR/MAMR and multiple actuators have given HDD technology–and, more importantly for preservation, interfaces such as SATA and SAS–a new lease on life. Nevertheless, the writing is on the wall as flash and related technologies move to NVME and CXL interfaces. See **Pure: No more hard drives will be sold after 2028**, Mellor, C., (2023), *Blocks & Files* https://blocksandfiles.com/2023/05/09/pure-no-more-hard-drives-2028/ [accessed at 2023-10-24].

- For example, SSDs can be remarkably sensitive to storage conditions when unpowered. See **JEDEC SSD Specifications Explained**, Cox, A., (2013), *JC-64.8*

  https://www.jedec.org/sites/default/files/Alvin_Cox%20%5bCompatibility%20Mode%5d_0.pdf [accessed at 2023-10-24].

- **Storage**, Digital Preservation Coalition , (2016), *Digital Preservation Handbook* https://dpconline.org/handbook/organizational-activities/storage [accessed at 2023-10-24].

---

b) **Software**

Commercial Software

Computer software that is produced for sale or that serves commercial purposes, including previous editions and versions of software that are not available or no longer in use. This entry broadly includes proprietary software, access through licences or subscription business models.

Categories

- Software

Examples

Wordstar, Novell Netware 386, SAP, Oracle, Adobe Photoshop; Microsoft products, such as the Windows Operating System and MS Office, are some of the most well-known examples of commercial software.

Imminence	Effort
●●●○○	●●●●○
Action is recommended within three years, detailed assessment within one year.	Loss seems likely. By the time tools or techniques have been developed, the material will likely have been lost.

Hazards

Lack of skills, commitment or policy from corporate owners; lack of established frameworks and tools; technology is poorly understood; no emulation pathway; reliance of proprietary software

and/or subscription-based business models; limited or no commercial interest; complexities of sector-specific software or data types; lack of technical documentation; uncertainty over IPR or the presence of orphaned works; technical protection measures that inhibit preservation actions; encryption.

Mitigations

Effective replication; access to source code; emulation pathway; strong technical documentation; preservation pathway; licensing that enables preservation; use of open formats and open source software; corporate preservation capability; awareness and advocacy work with commercial software providers.

Bit List History

Added to list: 2023

Trends

- **2024:** No change.

Last Review

2023 Review

This was a new Bit List entry added in 2023 to draw attention to the particular challenges of content and software preservation for commercial software products. The entry focuses on the distinct risks relating to the availability and access to software and code, and lack of preservation interest or mandate, by companies that publish them, creating challenges to preserve digital content and software in source code form. The entry also highlighted that software, as a whole, was not currently on the list (compared to higher-level species like Apps). The Council agreed on the creation of a new higher-level Software digital species group to address this gap and challenges specific to software preservation while also recognizing overlaps with other entries including Apps, Gaming, Media Art and Research Outputs. While there are overlaps, the Council agreed it would be valuable to separate software to reflect differences in the volume of access to software, significance and motivations for preserving commercial software in and across different sectors.

2024 Interim Review

These risks remain on the same basis as before, with no significant trend towards even greater or reduced risk ('No change' to trend).

Additional Information

A large part of this requires advocacy work, and call to action to raise awareness with commercial software providers to preserve their software.

Software preservation raises (often debated) questions about key characteristics for preservation. Wordstar, for example, may be of interest for not only access to the content it facilitates but also for the preservation of the software as mass produced, commercial product. You can also argue in more practical ways that for most files there is usually something about them that you need for the original software for or else the content will be different, and therefore preservation of software is critical regardless of the significance of the software as a product.

Software dependencies on the environment (hardware and software) enable it to run and its associated context and support and in this way some of the approaches can overlap with gaming, but software is not seen as a creative product in the same way. This nomination is more about commercial uses and industrial design objects where games are artworks for entertainment and social/cultural uses.

One might argue the contents of the Museum of Brands and Packaging is not unique and yet it seems unsafe to expect it to necessarily exist elsewhere. The low regard in which commercial software (and this is before we consider bespoke research software) is held - except as a means of accessing more interesting material - seems to suggest that we may in the future look in vain for someone with a copy of even quite widely available software, let alone, say, early UK antivirus products, world leading as they were.

Case Studies & Examples

- The example of the Microsoft outage in July 2024, in which a software update led to the cancellation of flights, healthcare disruptions and payroll issues. See **Global IT outage: More than 5,000 flights cancelled; how security 'arms race' led to crash | As it happened**, Sky News, (2024), *https://news.sky.com/story/outages-latest-airports-business-and-broadcasters-experiencing-issues-worldwide-13180821* [accessed at 2024-09-06].

- The Digital Archiving at the University of York blog series about preserving Wordstar files, which notes and demonstrates how a random copy of WordStar in the archive was able to be used to get a sense of the files in their original context. See **Some observations on digital appraisal**, Mitcham, J. , (2018), *Digital Archiving at the University of York* *https://digital-archiving.blogspot.com/2018/07/some-observations-on-digital-appraisal_20.html* [accessed at 2023-10-24].