

From Waste to Resource: How Standardized Health Metrics Can Accelerate the Circular Economy in Storage Media

Rich Kenny^{1,*}, Jonmichael Hands², and Nick Hayhurst³

ABSTRACT

Current research into hard disk drives (HDD) and solid-state drives (SSD) focuses on predicting failure rates using SMART attributes to decide when to retire a drive and move to the end of life rather than identifying ways to increase product life. This paper looks instead at what attributes need to be present to reuse the drive with full confidence and which attributes have a material impact on reliability during second use.

To quantify which attributes had the highest impact we conducted a large-scale erasure project of over 117,000 unique drives to identify reasons for the failure of erasure and what drive attributes led to the storage device's failed sanitization.

These findings aim to inform the development of industry standards and practices prioritizing data security, device health and facilitating increased reuse of storage media and furthering the circular economy.

Keywords: Circular economy, HDD, SSD, Storage media.

Submitted: June 18, 2024

Published: July 15, 2024

 10.24018/ejece.2024.8.4.636

¹Interact, Managing Director, UK.

²Circular Drive Initiative, USA.

³Cedar, Software Engineer, UK.

*Corresponding Author:

e-mail: rich@interactdc.com

1. INTRODUCTION

Significant increases in the demand for storage have left a gap between the amount of storage currently available, and being produced by manufacturers, and the amount of storage required by consumers and businesses in the long term. To facilitate this supply at the lowest possible financial and environmental cost there needs to be an increase in second-use storage devices and a movement to improve circular economy practises for the storage market.

This paper will quantify the key metrics and attributes that impact the reusability of storage media and then follow with the technical and functional aspects required to de-risk circular economy practices. This includes specific system design to derisk the use of refurbished drives through improved fault tolerance in deployments.

2. TECHNICAL REQUIREMENTS FOR REUSE

2.1. Understanding Reliability vs. Lifespan

Quality and reliability are distinct metrics. Quality aims to reduce initial failures (often due to less rigorous testing in lower-end drives), while reliability concerns failures arising from extended use. Enterprise HDDs, for instance,

boast Mean Time Between Failure (MTBF) ratings of 2 M or 2.5 M hours, equivalent to 0.44% and 0.35% Annualized Failure Rate (AFR). These metrics are statistical measures derived from large-scale population testing and do not directly predict the lifespan of an individual driver. For instance, an HDD with a 0.44% AFR does not guarantee 227 years of service (1/0.0044). Instead, it indicates a 0.44% chance of failure within a year, but the drive could last much longer or fail sooner.

2.2. How SSDs Fail

SSD failures differ from HDDs, with firmware issues being the most common, followed by media and hardware problems. Firmware updates are crucial, as they often address bugs responsible for most failures.

SSD endurance is how many writes take place to an SSD before it wears out and is tied to the program-erase (PE) cycles NAND cells can undergo. Repeated cycles degrade the oxide layers, impacting charge storage. More bits per cell often mean more writes for voltage accuracy (e.g., QLC vs. SLC), varying by manufacturer and NAND process. SSDs are rated in terabytes written (TBW) or drive writes per day (DWPD), per JEDEC standards JESD218 and JESD219 [1].



Most SSDs do not report their specification sheet rated TBW (though the new OCP NVMe specification includes it), so the percentage used in SMART is key for gauging remaining endurance. Resellers should know the original specs and warranty, including any rated TBW, as endurance impacts trust in used drives. SSD endurance is well-understood and measurable, relating to data retention, not drive failure likelihood. Drives at their TBW should be decommissioned or repurposed for non-critical uses.

Enterprise SSDs retain substantial health even after deployment, highlighting their potential for extended use with proper management. A case study of millions of SSDs in production enterprise storage workloads shows that many applications consume endurance very slowly [2].

2.3. HDD Failures and Workload Rating

HDDs, composed of electrical and mechanical parts, are susceptible to mechanical failures due to their moving read/write heads and spinning platters. Factors like workload and temperature significantly impact their reliability and performance. High temperatures and extensive use can lead to increased failure rates.

As HDDs contain moving parts, they are more prone to mechanical failure than SSDs. Workload and temperature are thus vital for HDD reliability [3]. Endurance is rated in TB/year of data read/written (e.g., 200-550TB/year for Chia farming drives). While some reallocated sectors are normal, a sudden increase signals potential failure. Monitoring SMART data helps track this and other health attributes. Hyperscaler studies on HDD predictive failure have shown that use, in reads and writes, is the number one predictor of increased probability of failure (increased AFR) [4]. Even a double or triple increase in AFR might still be suitable for another deployment with higher data protection, RAID, and erasure coding strength designed specifically for used storage devices.

2.4. Designing a System for Reliability with Used Drives

Hyperscalers such as Google, Meta, and Microsoft often see HDDs with a higher AFR after a 3–5-year deployment that were in heavy use. In designing an erasure coding (EC) and data protection schema specifically tailored for used drives with an Annualized Failure Rate (AFR) of 1%–2%, it is essential to implement a robust and resilient strategy to mitigate the higher risk of drive failure. Given the increased likelihood of drive failures, the chosen EC scheme must balance redundancy, storage efficiency, and computational demands. For instance, using stronger EC configurations like N+2 or N+3 (where N represents the number of data fragments and the number following the “+” indicates the number of parity fragments) can provide enhanced fault tolerance, capable of withstanding multiple simultaneous drive failures. This approach ensures data integrity and availability even when used drives exhibit higher AFR.

Additionally, incorporating fast rebuild mechanisms and monitoring systems to identify and replace failing drives swiftly can further bolster system reliability. By optimizing these parameters, a storage system utilizing refurbished drives can achieve the necessary reliability and performance to support circular economy practices while

addressing the technical challenges associated with used storage media [5].

Used SSDs may have a lower AFR due to maintenance release firmware updates if the NAND flash endurance has not been extensively worn out. It is critical for used storage systems that device vendors easily allow used storage devices to be easily field upgradable to the latest version. Firmware updates enhance reliability, patch critical security vulnerabilities, and add feature and performance enhancements [6].

3. FACTORS IMPACTING DRIVE CIRCULARITY

3.1. Data Sanitization

The importance of effective data sanitization cannot be overstated in the context of reused storage media. Modern sanitization techniques such as cryptographic erasure, secure erase, and firmware-based methods must be standardized and rigorously implemented to ensure data privacy and security.

The 2023 paper *New IEEE Media Sanitization Specification Enables Circular Economy for Storage* [7] explores data sanitization in detail. The key is purging media sanitization, a method designed to render data on a storage device irrecoverable even with state-of-the-art equipment. Not only purge target user data but also any media that used to contain user data like spare areas and reallocated sectors or retired blocks.

To find the number of HDDs and SSDs which could be reused we carried out data erasure on 117,982 individual drives, with manufacturing dates ranging from 2015–2022, over the course of one year (2023). The project focused on erasing the drives and cataloguing any reasons why a drive failed erasure. Any failure would preclude the drive from being reused so was the first step to identify the potential reuse values in storage devices.

4. METHODOLOGY

To ensure comprehensive data sanitization, we followed industry-standard practices in erasing 117,982 individual drives, including specific methods for both HDDs and SSDs. The process was designed to ensure thorough erasure, verification, and analysis of storage devices, addressing both data protection and device reuse.

Each drive was placed in a dedicated enclosure or system designed to facilitate secure erasure and the software schedule algorithm was executed.

4.1. Pre-Testing

Objective: To assess the health and readiness of each drive prior to the sanitization process.

Procedure: Each drive was subjected to an initial test to capture a variety of S.M.A.R.T (Self-Monitoring, Analysis, and Reporting Technology) attributes. Key health indicators, such as the Reallocated Sector Count and Uncorrectable Error Count, were analysed to determine the condition of each drive.

Outcome: The pre-testing phase identified drives in good health and those at risk of failure, ensuring informed

decision-making for the subsequent erasure process. Those drives that failed initial health checks were marked as not suitable for reuse. With failure reasons specified in the table below.

4.2. Data-Erasure

Objective: To securely erase and sanitize each drive to the highest level supported by its firmware.

Procedure: The data erasure process adhered to the IEEE 2882-2022 [8] and NIST 800-88 R1 Clear and Purge standards [9]. Each drive was analysed to determine its firmware-based capabilities, applying the highest level of sanitization supported. Sanitize techniques were utilized where supported by the device firmware to ensure irreversible data erasure.

Outcome: The data on each drive was securely erased, meeting rigorous security standards, and ensuring compliance with NIST and IEEE guidelines. Where any failures occurred, they were captured and added to the table below.

4.3. Verification and Logging

Objective: To confirm the success of the data sanitization process and maintain detailed records of each drive.

Procedure: Post-erasure, each drive underwent a verification process to ensure the efficacy of the sanitization. Detailed logs were maintained to record the status of each drive and document any anomalies encountered during the process.

Outcome: Verification confirmed the successful erasure of data, and comprehensive logs provided a reliable record of the sanitization process.

4.4. Post-Testing

Objective: To assess the health of each drive after the erasure and verification steps.

Procedure: Each drive was evaluated to ensure that health-based statistics, such as S.M.A.R.T attributes, had not deteriorated further during the sanitization process.

Outcome: The post-testing phase confirmed that the sanitization process did not adversely affect the health of the drives, supporting their potential for reuse.

4.5. Failure Analysis

Objective: To analyse and understand the reasons for erasure failures.

Procedure: Drives that failed the erasure process (15,776 cases) were subjected to detailed analysis to determine the root causes. Investigations focused on hardware malfunctions, firmware incompatibilities, and other factors contributing to the failures.

Outcome: The failure analysis provided insights into the causes of erasure failures, contributing to the enhancement of future sanitization processes, and improving the handling of storage devices.

5. RESULTS

Within this dataset, we encountered 15,776 instances where erasure proved unattainable, necessitating a deeper investigation into the root causes. 94,643 were HDD, with

12,048 failures (12.7%), and 23,339 SSD with 3,728 failures (16%), for a combined failure rate of 13.37%.

The breakdown of the top fifteen erasure failures revealed several distinct categories, each shedding light on the multifaceted challenges inherent in data sanitization and reuse (Table I):

5.1. Verification Failure (I/O Error during Verification): 4510 instances

Verification failure emerges as a predominant challenge. This indicates discrepancies between expected and actual outcomes during the erasure process, or difficulties whilst reading data from the device, highlighting potential issues with data integrity.

5.2. Erasure Failure (I/O Error during Erasure): 4053 instances.

Erasure failure identifies difficulties in successfully wiping data from storage devices. These errors stem from numerous factors, including physical defects, firmware anomalies, or environmental influences.

The high volume of I/O (Input/Output) errors during the erasure process can be attributed to various underlying factors, each contributing to the overall complexity of data sanitization efforts. I/O errors can occur when the system encounters difficulties in reading from or writing to storage devices, hindering the smooth execution of data operations. Several reasons may account for the occurrence of I/O errors during erasure:

- *Physical Damage or Wear:* One of the primary causes of I/O errors is physical damage or wear to the storage medium. Over time, storage devices can develop physical defects such as bad sectors, scratches, or mechanical failures, impairing the ability to read from or write to certain areas of the disk. These defects can disrupt the erasure process, leading to I/O errors.
- *Firmware Anomalies:* Firmware issues within the storage device can also contribute to I/O errors during erasure. Firmware is responsible for controlling the device's operations, and any anomalies

TABLE I: TOP 15 REASONS FOR STORAGE DEVICE FAILURE

Failure reasons	Quantity
Verification failed (I/O Error during verification)	4510
Erasure failed (I/O Error during erasure)	4053
S.M.A.R.T short self-test failed	1090
Grown defects value exceeds threshold	980
S.M.A.R.T status is not OK/Failed after erasure	752
Health value exceeds threshold	623
Unexpected data in read sample	564
Test unit ready failed	521
Check condition failed. Data channel impending failure data error rate too high	474
S.M.A.R.T status is not OK/failed	390
Uncorrectable errors value exceeds threshold	383
Device is not ready	192
I/O Error during stress test	183
Device is TCG locked. PSID revert failed	133
Sequential read test failed	122

or glitches in the firmware can result in communication errors between the device and the erasure software, leading to I/O errors.

5.3. *S.M.A.R.T Short Self-Test Failure: 1190 instances*

Identifies potential drive health issues detected during self-testing. Failures in S.M.A.R.T tests serve as early indicators of impending drive failure, emphasizing the importance of proactive monitoring and maintenance.

5.4. *Grown Defects Value Exceeds Threshold: 980 instances*

Exceeding the threshold for grown defects highlights the progressive degradation of drive performance, posing challenges to erasure operations and data integrity.

5.5. *S.M.A.R.T Status is not OK/Failed after Erasure: 752 instances*

This indicates post-erasure issues with drive health or reliability.

5.6. *Health Value Exceeds Threshold: 623 instances*

This underscores deteriorating drive conditions that impede erasure processes or indicate imminent drive failure.

5.7. *Unexpected Data in Read Sample: 564 instances*

This indicates deviations from the anticipated pattern, such as 0x00 or 0xFF. These deviations suggest potential irregularities in the erasure process or data corruption issues.

5.8. *Test Unit Ready Failure: 521 instances*

This indicates issues with drive readiness during testing or erasure procedures, potentially stemming from hardware malfunctions or operational irregularities, such as Medium Format corruptions, or Critical Status.

5.9. *Check Condition Failed. Data Channel Impending Failure Data Error Rate Too High: 474 instances*

These are failures related to an increasing Data Error rate points to a potential mechanical or electrical errors, this is a critical failure.

5.10. *S.M.A.R.T Status is not OK/Failed: 390 instances*

These are failures in S.M.A.R.T status checks underscore potential drive health issues that may compromise data integrity and erase reliability.

5.11. *Uncorrectable Errors Value Exceeds Threshold: 383 instances*

Issues in maintaining data integrity during read or write operations, requiring attention to drive health and performance metrics.

5.12. *Device is Not Ready: 192 instances*

Issues with device readiness pose operational challenges during erasure procedures usually linked to hardware malfunctions or environmental influences.

5.13. *I/O Error during Stress Test: 183 instances*

Encountering I/O errors under stress conditions highlights potential vulnerabilities in drive performance and reliability.

5.14. *Device is TCG Locked. PSID Revert Failed: 133 instances*

These are security-related impediments to erasure processes and potential firmware anomalies.

5.15. *Sequential Read Test Failure: 122 instances*

Failures in sequential read tests may indicate underlying issues with drive performance or data integrity, linked to drive health and reliability.

6. CONCLUSION

Over 85% of the failures recorded during the data sanitization study were related to drive wearing, health or use impacts on the drives. These results are consistent with current research on how storage media can fail over time. With almost 87% of drives rendered suitable for reuse, there is significant potential for extending the lifecycle of storage media, thereby supporting the principles of the circular economy.

Despite the high success rate of data sanitization, the health and quality metrics of the erased drives exhibited considerable variability. This underscores the challenge posed by the absence of standardized benchmarks for assessing the quality and health of storage devices intended for secondary use. If standards were defined, and tolerances established, the number of drives failing for health-based reasons is likely to be higher.

Inconsistent quality metrics can reduce customer confidence in refurbished drives, limiting their acceptance and widespread adoption in the market. Addressing this challenge requires the development and implementation of comprehensive standards and benchmarks for evaluating the quality and health of refurbished drives.

Establishing such standards would enhance customer trust and increase the adoption of refurbished drives, thereby maximizing the reuse of storage media. Aligning with these standards is essential for fostering a robust secondary market for storage devices, contributing to more sustainable practices within the industry. This approach not only supports environmental sustainability but also promotes economic efficiency by reducing waste and optimizing resource use.

CONFLICT OF INTEREST

Authors declare that they do not have any conflict of interest.

REFERENCES

- [1] Global Standards for Microelectronics Industry. Solid-state drive (SSD) Endurance Workloads. 2023. [accessed 10.06.24]. Available from: <https://www.jedec.org/>.
- [2] Maneas S, Mahdavian K, Emami T, Schroeder B. Operational characteristics of SSDs in enterprise storage systems: a large-scale

- field study. *20th USENIX Conference on File and Storage Technologies*, 2022.
- [3] Western Digital. 2023. SSD endurance and HDD workloads. [Accessed 12.05.2024]. Available from: https://documents.western-digital.com/content/dam/doc-library/en_us/assets/public/western-digital/collateral/white-paper/white-paper-ssd-endurance-and-hdd-workloads.pdf.
 - [4] Miller Z, Medaiyese S, Lin F, Beatty A, Ravi M. Hard drive disk failure analysis and prediction, META platforms. 2024. [Accessed 07.06.2024]. Available from: https://www.youtube.com/watch?v=Fhie_8GECHU.
 - [5] Maneas S, Mahdavian K, Emani T, Schroeder B. Operational characteristics of SSDs in enterprise storage systems: a large-scale field study. 2022. [Accessed 09.05.2024]. Available from: <https://www.usenix.org/system/files/fast22-maneas.pdf>.
 - [6] Kadekodi S, Maturana F, Athlur S, Merchant A, Rashmi K, Ganger G, et al. Tiger: disk-adaptive redundancy without placement restrictions. 2022. [Accessed 09.05.2024]. Available from: <https://www.usenix.org/system/files/osdi22-kadekodi.pdf>.
 - [7] Hands J, Coughlin T. New IEEE media sanitization specification enables circular economy for storage. *Computer*. 2023;56(1):111–6. doi: 10.1109/mc.2022.3218364.
 - [8] Kissel R, Regenscheid A, Scholl M, Stine K. IEEE standard for sanitizing storage. 2014. [Accessed 05.05.2024]. Available from: <https://standards.ieee.org/ieee/2883/10277/>.
 - [9] National Institute of Standards and Technology (NIST). NIST special publication 800-88 revision 1: guidelines for media sanitization. 2014. Available from: <https://csrc.nist.gov/pubs/sp/800/88/r1/final>.