

## Electronic Data Shredding

This brief describes the data shredding and system sanitization features included in the Data Domain Retention Lock software option, including the use of this functionality to handle data shredding on Data Domain deduplication storage systems. Developed to support the clearing and sanitization requirements of government and enterprise customers, this is the industry's first inline deduplication storage solution to enable electronic data shredding on a per-file basis.

## Introduction

Unlike other deduplication products, the Data Domain approach to data shredding securely overwrites segments on disk that uniquely belong to deleted files. This way, even small data sets stored inappropriately, for example classified data or social security numbers written to unsecured systems, can be completely eliminated from all disks in those systems. This ensures that the inappropriately stored data cannot be read later even if the disk drive is removed. Other shredding solutions require data to be extracted or migrated, while Data Domain systems sanitize data in-place, in its native, deduplicated state. As required, sanitized deduplicated data can be efficiently migrated to another system, quickly freeing the entire original system for proper disposal.

## Deleting Files Is Not Enough

Deleting a file in most file systems consists of just flagging the file or deleting references to the data on disk, freeing up the physical space to be consumed at a later time. However, this simple action introduces the problem of leaving behind a residual representation of underlying data physically on disks. Deduplicated environments are not immune to this problem. This makes inadvertent disclosure of sensitive information possible, should the data be accidentally placed into the wrong storage environment.

Shredding data in a system implies eliminating the residual representation of that data and thus the possibility that the file may be accessible after it was shredded. There are many standards and guidelines that address accepted levels of shredding or sanitization. This brief focuses on the ones that Data Domain customers need to comply with most often: The US Department of Defense 5220.22-M Clearing and Sanitization Matrix and the National Institute of Systems and Technology (NIST) Special Publication 800-88 Guidelines for Media Sanitization.

## The Complexity of Sanitizing Deduplicated Data

Deduplication storage systems achieve superior cost benefits by extracting common data patterns from files in the system and only storing unique copies of these patterns, referencing all the redundant instances. These data patterns or segments may potentially be

shared among many files in the system. In order to eradicate a file in a deduplicated environment, one cannot simply erase all the segments the file contains because many other files could be affected. The system needs to first determine whether each of the segments of the contaminated file are shared with a clean file and erase only those segments that are not shared, along with any contaminated metadata. The required outcome is to effectively restore the storage device to a state as if the contaminated files never existed in that system. To add to this complexity, storage systems are often optimized for cost and/or performance and as a result, they move data to different locations (different tiers of storage, caches, etc). All of these locations must be taken into account when eradicating contaminated files.

In all cases this is a complex and time-consuming task. Most deduplication solutions in the market require shredding and sanitization to happen in extracted data sets, forcing administrators to extract all the data, sanitize it, and deduplicate it again. At best, this could be a long, resource-intensive process; but in the worst case it could require additional equipment, manual steps, and disruption to daily operations.

## Data Domain Electronic Data Shredding and System Sanitization Capabilities

The patent-pending system sanitization capability introduced in Data Domain Operating System (DD OS) version 4.6 is a special and complete form of clearing that handles every copy of every segment that belongs exclusively to deleted files. The system reclaims and overwrites all of the storage occupied by these segments.

First, the system determines for each segment whether it is still being used by a file. This is accomplished by going through all the files in the system and, for each file, iterating through all of its segments to mark the segments as live. After having gone through all files in the system, any segment that has not been marked live is not part of an existing file. This process ensures that all copies of all contaminated segments uniquely used by previously-deleted files are identified for sanitization.

Second, each and every container in the system is examined to determine if it holds any segment that has not been marked. If

### Classified Message Incident

A classified message incident (CMI) occurs when data at a certain classification level is written to a data storage device that is not approved to store data of that classification. For example, a CMI happens when a user inadvertently sends an email with information classified as 'top secret' to an email system that is only approved for a low-level of clearance.

### Clearing a CMI

When a CMI occurs, the system administrator must take immediate action to completely eradicate the data from the storage device where the data was accidentally written. The required outcome is to effectively restore the storage device to a state as if the CMI never occurred. If the CMI is not rectified before the next backup cycle, the backup device becomes contaminated by the CMI and must be sanitized as well.

so, only the live segments in the container are copied into a new container. After the copy is verified, the old container is marked as ready for sanitization.

Third, for each and every container that has been marked as ready for sanitization, an explicit I/O is issued to zero out the entire container, overwriting any residual representation of the contaminated data.

There is a fourth step for the cases where higher-level sanitization requirements mandate destruction of the disks in the system. Data Domain supports destruction of contaminated disks by migrating clean data to a new system with Data Domain Replicator software. Performing a data migration from a contaminated system to a new clean system immediately after the sanitization command completes will ensure that only clean data is replicated to the new system. Any data migration would implement efficient transfers of deduplicated data using sequential I/O, taking only a fraction of the time. Once the original system is replaced by the new system, its disks can be properly destroyed, thus complying with the strictest standards of data sanitization.

During the entire process, the clean data in the system being sanitized is online and available to users (system is online in read-only mode).

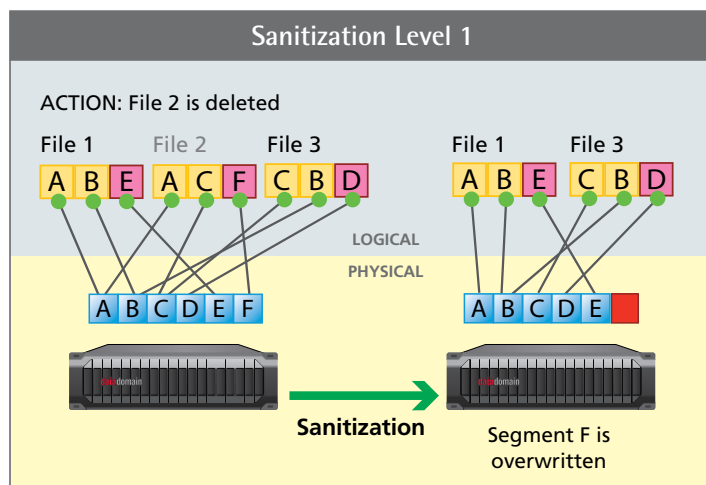
## Steps to Comply with Data Sanitization Requirements

According to the DoD Clearing and Sanitization Matrix and the NIST Guidelines on Media Sanitization, there are essentially two levels of sanitization. The first level requires affected storage to be overwritten once. The second level requires entire storage devices to be destroyed either through degaussing or physical means. Overwriting the storage multiple times used to be acceptable for the second level of sanitization but this is no longer the case.

### Sanitization Level 1: Data Clearing or Data Shredding

The Data Domain system sanitization feature ensures that every copy of every segment that belongs only to erased files is overwritten. This feature provides a basis for handling most data shredding and system sanitization cases. The steps are as follows:

1. Delete the contaminated files or backups through the backup software or corresponding client. In the case of backups, be sure to manage the backup software appropriately to ensure that related files on that image are reconciled, catalog records are managed as required, etc.
2. Run the **system sanitize start** command on the contaminated Data Domain system to cause all previously used space in it to be overwritten once. The system will go into read-only mode. (See Figure 1)



**Figure 1.** Contaminated segments of deleted files are overwritten in the same system while data is available.

3. Wait for the affected system to be sanitized. Sanitization can be monitored by using the **system sanitize watch** command.

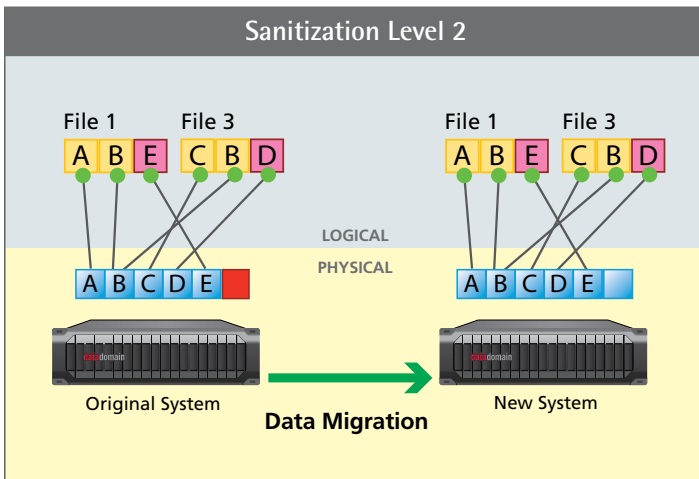
If the affected Data Domain system has replication enabled, all the system containing replicas need to be processed in a similar manner.

Depending on how much data exists in the system and how it is distributed, the system sanitize command could take some time. However, during this time all clean data in the system is available.

### Sanitization Level 2: Full System Sanitization

In those rare situations that require the disk drives to be degaussed or physically destroyed, the system sanitization feature can be used to first compact the “clean” data. The clean data may then be efficiently copied to a secondary system using the Data Domain collection replication feature without bringing along any contaminated segments. This allows the disks in the original system to be processed (e.g. destroyed) in any desired way. The Data Domain collection replication feature operates at the physical level to copy only the clean deduplicated data. With deduplication, the number of disk drives that must be destroyed is also very much reduced. The steps are as follows:

1. Delete the contaminated files or backups through the backup software or corresponding client. In the case of backups, be sure to manage the backup software appropriately to ensure that related files on that image are reconciled, catalog records are managed as required, etc.
2. Run the **system sanitize start** command on the contaminated Data Domain system to cause all previously used space in the system to be overwritten once. The system will go into read-only mode.
3. Wait for the affected system to be sanitized. Sanitization can be monitored by using the **system sanitize watch** command.



**Figure 2.** Clean data migrated to a new system, freeing drives in original system for proper disposal.

4. Initiate the replication process (with collection replication) to copy all data to a secondary system. Only clean data will be copied. (See Figure 2)
5. Run the system power off command on the contaminated Data Domain system.
6. Remove all of the disk drives from the system and destroy as required.

The system sanitization feature works on the Data Domain Appliance, Gateway and DDX Array series. For the Gateway series, the underlying third-party storage must write in place.

## Conclusion: The Data Domain Advantage

Data Domain electronic data shredding is the industry's first solution to enable in-place shredding for inline deduplication storage systems on a per-file basis.

- **In-Place:** The system sanitizes data in its native deduplicated state, avoiding the hassle and expense of setting up additional environments to extract and deduplicate it again or other complex procedures.
- **File-level Control:** Authorized administrators can surgically remove the content of deleted files from the deduplicated environment, ensuring complete removal using a DoD/NIST compliant algorithm and procedures.
- **Data Availability:** During the sanitization process, data is available at all times even if clean data needs to be migrated to a new system, limiting disruptions to daily operations.
- **Supports Full Sanitization:** As required, sanitized deduplicated data can be efficiently migrated to another system, quickly freeing the original system for proper disposal. Any data migration would implement efficient transfers of deduplicated data using sequential I/O, taking only a fraction of the time.

Government agencies and businesses using Data Domain systems can now sanitize confidential data that is accidentally written into an unapproved system or shred any content that is no longer required for internal or external purposes in a way that is simple, secure, and compliant.