# **Checksum Requirements**

# What is a Checksum?

A checksum is a computer-generated string of numbers and letters serving as a digital fingerprint for a digital object. Even the slightest alteration to a digital object will result in a complete change in its checksum. For instance, consider the following checksum for a digital object:



Original checksum: **f75d91cdd36b85cc4a8dfeca4f24fa14**

Here’s the checksum for the same digital object after some corruption:



New Checksum: **7aca5ec618f7317328dcd7014cf9bdcf**

## Why are they used?

Checksums play a crucial role in ensuring the integrity of digital objects. Any change in a digital object’s checksum signals a corresponding change in the object’s data—be it alterations in content, data loss, or corruption. An unchanged checksum indicates that the object’s data remains unaltered since the checksum's creation.

## What can they be used for?

Checksums facilitate the establishment of a ‘chain of custody’ between those involved in creating, preserving, and accessing digital objects. In data management, checksums serve at least three primary purposes:

1. **Confirmation of Successful Transfer:**
   * To verify that a digital object has been transferred and received without any alterations.
2. **Integrity Maintenance in Storage:**
   * To confirm that the integrity of a digital object has been preserved whilst in storage, safeguarding it from changes or corruption.
3. **User Confirmation:**
   * To provide users with confirmation that the digital object they access has been retrieved, stored, and delivered without any data changes.

## When are checksums required?

Archives require metadata accompanying physical or digital record transfers, including descriptive and fixity metadata. Fixity, ensuring record integrity, is confirmed through checksums, algorithms, and file paths.

Agencies are required to provide checksums when transferring digital records, to ensure their continued unaltered state. The ability to produce checksums is a crucial readiness characteristic for digital transfer.

# Generating and validating checksums

Checksums can be generated and validated using various software tools, some of which are freely available. Examples include FileList Creator and HashMyFiles. DROID (Digital Record Object Identification) is another tool for generating checksums, provided by The National Archives UK. More information is provided below, in the “How to generate checksums” section.

Archives can work with various checksum algorithms; some of the more widely used include MD5 and SHA-256. The ‘stronger’ the algorithm, the harder it is to deliberately change a file in a way that goes undetected. For example, SHA-256 checksums have longer hash lengths that make them more robust from a cybersecurity perspective. This can be important for applications where there is a need to demonstrate resistance to malicious corruption or alteration of digital objects, especially when this information is required as legal evidence.

In a digital preservation context, checksums are primarily used to detect accidental loss or damage to files (i.e. due to a storage or migration failure), which makes the MD5 algorithm more than sufficient. MD5 checksums also have the advantage of being well supported in most tools and are quick to calculate. For these reasons, Queensland State Archives (QSA) has chosen to work with MD5 checksums when Agencies transfer their digital records.

It is important to note that while checksum algorithms detect change, they do not specify where or what changes have occurred.

## How QSA use checksums

When transferring permanent born-digital records to QSA, Agencies generate MD5 checksums before transfer and QSA use these values to validate each digital record, ensuring successful transfer without alterations.

Checksums are also employed in the QSA Digital Archive, where Archivematica, the long-term digital preservation system, continuously creates and monitors SHA-256 checksums for every digital record, ensuring the perpetual integrity of the digital archives.

## How to generate checksums

QSA is not prescriptive on the specific tool agencies use to generate and validate checksums for their born-digital records, so long as they support the MD5 format. If you are having difficulty selecting the right tool for your agency, some things to consider include:

* The platforms you need to run the software on i.e., Windows, Mac, or Linux
* The volume of checksums you need to generate at any one time, and how often
* How you plan to export and share checksums in a standardised way
* The IT set up within your agency and whether you require permission to install new applications.

Here are some widely used applications for generating checksums that we can recommend for your convenience:

| Tool | Operating System | Availability | About the tool |
| --- | --- | --- | --- |
| HashMyFiles | Windows | Free to download here: <https://www.nirsoft.net/utils/hash_my_files.html> | HashMyFile is a small, efficient utility that allows users to calculate MD5 hashes (checksums) and identify duplicate files. The hash lists can then be easily copied into a clipboard and saved into a text file. |
| FileList Creator | Windows/Mac/Linux | Free to download here: <https://www.sttmedia.com/filelistcreator-download> | FileList Creator is an easy-to-use tool for producing directories/manifests that list the files (including their file paths and other technical metadata) within a specific folder structure. The tool also creates checksums for each file that can be included in the file list. These manifests can be exported to an .xlsx file. |
| ExactFile | Windows | Free to download here: <https://www.exactfile.com/> | ExactFile has multiple functions. The tool enables users to run checksums over individual files or groups of files. It can also generate associate checksums with files copied to and from media carriers i.e., optical discs. Checksum reports are saved as text files which can be read by other tools and text readers. |
| DROID | Windows/Mac | Free to download here:  <https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/> | DROID is predominately used as a file format identification tool (which is still a useful preservation practice to take up). However, during the identification process, technical metadata is captured along with the file’s checksum which can then be exported via a .csv report. |

The above is a non-exhaustive list of the tools that are available. There are also applications within the Microsoft suite, such as PowerShell, that can also create checksums. Agencies may wish to consult their IT departments to confirm whether preinstalled software supports this feature.

Remember, QSA is always able and willing to provide assistance on how best to use these applications.

### Further resources

[**Digital Preservation Coalition – Fixity and Checksums (External link)**](http://www.dpconline.org/handbook/technical-solutions-and-tools/fixity-and-checksums).

Contains further reading and links to other tools.

[**Digital Preservation Coalition - Which checksum algorithm should I use? (External link)**](http://doi.org/10.7207/twgn20-12)

(PDF 468 KB). Further reading on what checksum algorithm you should use, depending on a number of factors.