# The Three Essentials of Digital Preservation Part 2: File Integrity

This document provides an introduction to three essential concepts in digital preservation: **File Storage**, **File Integrity**, and **File Access,** with a detailed focus on file integrity. The "Three Essentials of Digital Preservation Pyramid" below identifies and describes the most important pieces of digital preservation for smaller institutions. This document focuses on file integrity. Establishing a basic understanding of concepts related to file integrity will help identify what your institution already has in place, and areas where preservation plans and policies need to be expanded.

This document begins with an **Introduction to the Three Essentials of Digital Preservation Pyramid**, providing a brief introduction to the concepts. After that, File Integrity is divided into an **Introduction to File Integrity** section, **Important Terms Related to File Integrity** section, and section of **Questions to Ask Your Institution About File Integrity**. Important related terms are in bold in each Introduction section.

For more information about digital preservation concepts, tools, and policies, view related items connected to this resource on the Sustainable Heritage Network in the "Digital Preservation" category.

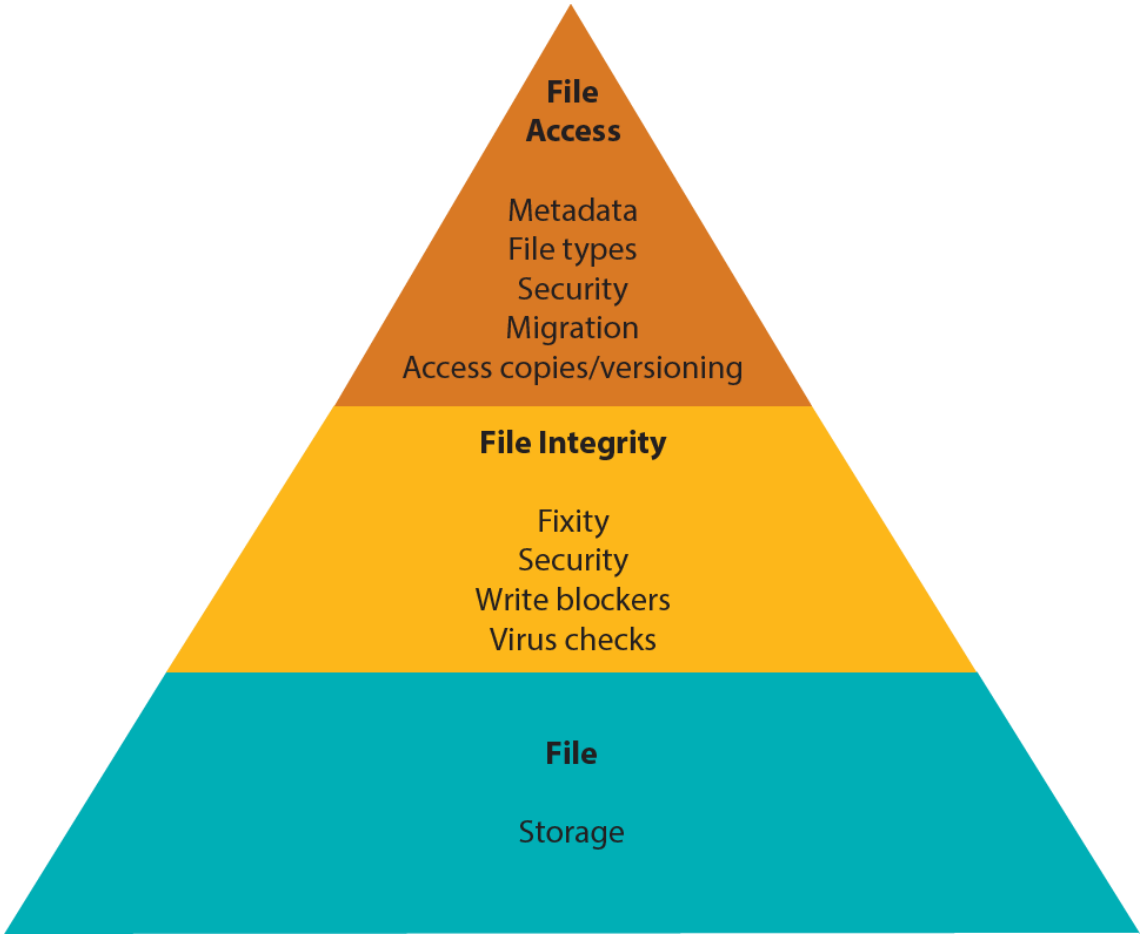
- The Three Essentials of Digital Preservation Part 1: File Storage
- The Three Essentials of Digital Preservation Part 3: File Access
- Levels of Digital Preservation Preparedness
- Activities to Include in a Digital Preservation Plan
- Digital Preservation Glossary
- Developing a Digital Preservation Policy

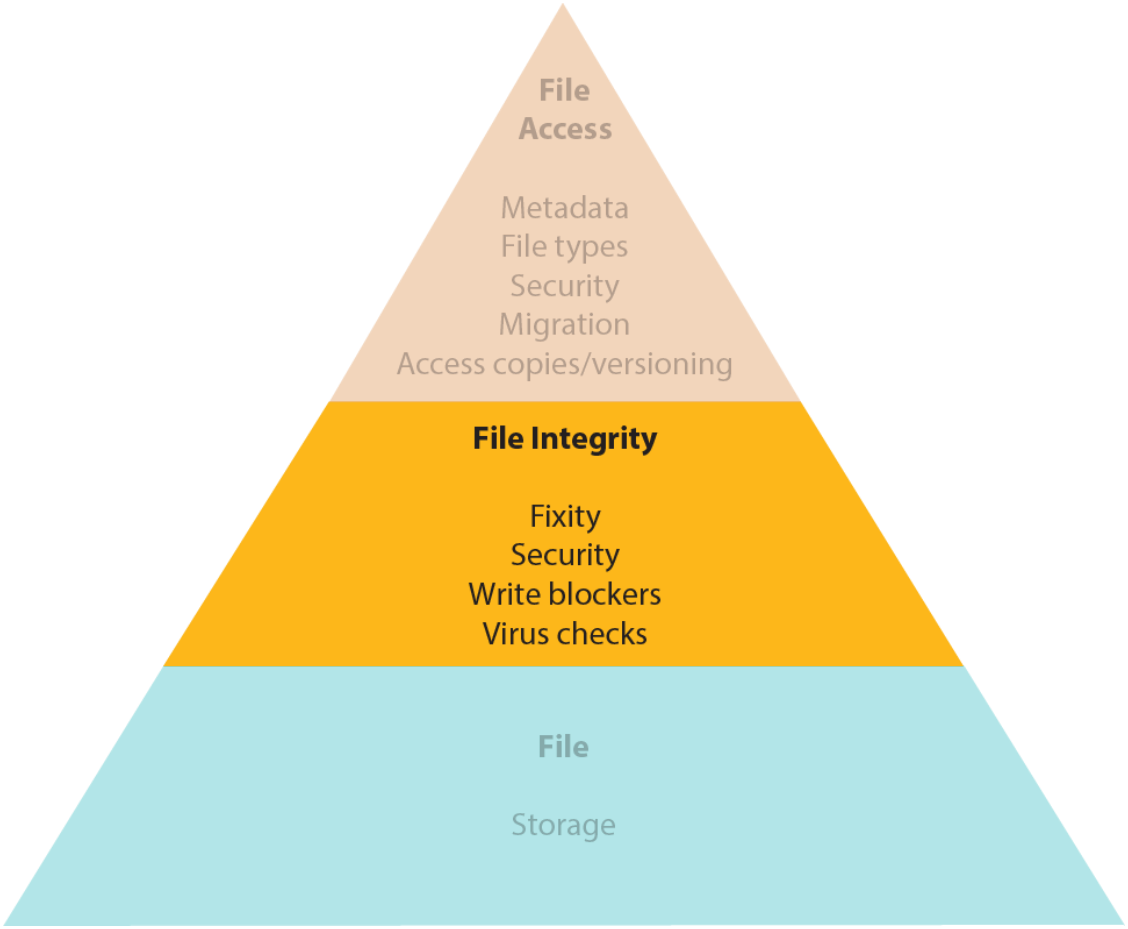
# Introduction to the Three Essentials of Digital Preservation Pyramid

**File Storage:** *Ensuring that digital content chosen for long term preservation is stored safely and securely.* File storage addresses physical storage systems, location of storage, and use of multiple physical storage locations to prevent or minimize data loss due to storage device failure or natural disaster.

**File Integrity:** *Ensuring the stability of digital content over time.* File integrity addresses stability of data, concerns about data corruption and alteration, as well as prevention, detection, and recovery of changed data.

**File Access:** *Organizing and describing digital files so that all staff (now and in the future) will be able to find, access, understand, and use digital content.* File access addresses security of data, documentation of data, file formats, data structures and naming conventions.






## Introduction to File Integrity

File integrity ensures the stability and usability of digital files over time. It is important to be able to track and document that digital files remain unchanged over time. Digital files can be altered or damaged by things like viruses or human error, but digital files can also change and deteriorate at the bit level *without* warning in a process known as **bit rot**.

Due to these risks, it is important to have tools and processes in place at your institution that help manage digital files and maintain file integrity. Having a plan to check and verify **fixity** of digital files, and ways to document this information, is an important step. There are several ways to monitor and confirm fixity, including using **checksums**. There are both both free and for fee software to create and verify checksums. Checksums allow you to compare the data between two versions of a file to

ensure they are the same. Some checksum softwares are standalone, and some come bundled with other digital preservation or management tools.

**Security** of files is important to file integrity. Ensure files are secure, with proper permissions applied to make files accessible to only appropriate staff, so that staff do not accidentally alter or delete files. Files coming into your institution from donations or or other departments can pose a risk. Implement tools and strategies for managing new donations of digital files. Steps can include **virus scans** to protect files and systems, a **non-networked computer** for processing new accessions if possible, and a **write blocker** to ensure authenticity of files.

Ensuring file integrity may involve with other departments such as: Information Technology, Administration, or others. When working with others, you may need to explain certain concepts and risks that are unique to digital archives and to your digital collections. For example: the need for saving high resolution scans, the need for metadata, or the large file sizes of video formats.

## Important Terms Related to File Integrity

**Bit rot:** Also known as data decay or data degradation. Deterioration of files due to changes at the bit level.

**Fixity:** Ensuring a file remains unchanged over time. A **fixity check** is a test to ensure a file has not been changed, using checksum or digital signature.

**Checksum:** A string of characters, like a digital fingerprint, that can be generated and checked to ensure the integrity of a file after it has been transmitted from one storage device to another. Create checksums when a file is created, during a transfer or recovery, and at regular intervals. MD5 and SHA-1 are two types of checksums.

**Security:** Techniques for ensuring that data stored in a computer cannot be read or compromised by any individuals without authorization.

**Write Blockers:** Write Blockers allow you to transfer files from another storage media onto your computer without being able to *write* to the original storage media. There are both hardware and software write blockers.

**Virus:** A program or piece of code that is loaded onto your computer without your knowledge and runs against your wishes.

**Virus Scans:** A utility that searches a hard disk for viruses and removes any that are found. A type of antivirus program that searches a system for virus signatures that have attached to executable programs and applications.

**Ingest:** The process through which digital objects are added into a managed environment.

**Non-networked computer:** A standalone computer that is separated and not connected to other computers at your institution. A non-networked computer can be a good way to process digital files from storage media that may be carrying viruses (like flash drives).

**Quality control:** Quality control or quality assurance is a systematic review of digital files and processes to ensure files meet the standard of quality, accuracy, and consistency that is needed at your institution. Quality control may be something that you implement during digitization workflows, but it will be helpful to digital preservation as well.

## Questions to Ask Your Institution About File Integrity

- Do you have any tools or processes in place to check for file fixity of digital files?
    - If there is nothing already set up, then how and when should you check for file fixity? Where will fixity information be stored?
- Who are the people who have access to your files?
    - What actions they can take? For example: viewing, editing, deleting, etc.?
    - Do you need to restrict access?
- Does IT keep logs for who accesses files and what actions they take?
    - If not, is this something you want to implement?
- When do virus checks currently happen?
    - Are there regularly scheduled virus checks? If not, can you start doing them?
    - Do you run virus checks on new digital files that are donated?
    - If a virus is found, who is responsible for fixing the affected computers and files?

- How do you currently accept donations of digital files?
    - You might accept donations of digital files on hard drives, flash drives, compact discs, via digital transfer, or on other media.
    - Are there any risks present in the way we currently accept digital files?
    - Is there is a non-networked station you can use for transferring new digital collections? Transferring digital files to a networked computer can pose a risk to all computers on the network if there is a virus present.
    - If you are not yet accepting donations of digital files, consider if you will in the future, and what policies and procedures you should create.
- Do we use write blockers for new donations of digital files?
    - If you want to maintain and document authenticity of donated digital files, a write blocker may be an important tool to consider.

## Conclusion

Digital preservation is a unique challenge for every institution. The three concepts of the Three Essentials of Digital Preservation Pyramid: **File Storage**, **File Integrity**, and **File Access,** provide a framework and ideas for creating a structured system of digital preservation. This document provides file integrity background, important terminology, and questions to help with planning steps. Decide the best path forward for digital preservation based on resources available, including staff time, funding, and technology support. Consider the questions in this document, bring others into the conversation, start creating digital preservation policies and other documentation, and continue learning about the need for digital preservation through resources on the Sustainable Heritage Network.