# **Extracting records from digital carriers**

With the roll out of the Digital Archive, agencies across Queensland Government will be able to transfer their permanent born-digital records to Queensland State Archives (QSA) via the ArchivesGateway portal. In preparation for these transfers, it is important to first distinguish between born-digital records and the media (digital carriers) they are often stored on. As the digital transfer process only covers the content on digital carriers and not the carriers themselves, QSA requires agencies to extract the files – along with their metadata – from any digital carriers so they can be uploaded to ArchivesGateway.

However, migration activities are not without their risks, which is why agencies need to adopt measures that will prevent any harm to the born-digital records during this process. This guide covers the necessary steps to safely move born-digital records from their carriers.

# What are digital carriers?

Digital carriers are the external devices used to store information and data. Examples of digital carriers include (but are not limited to): USBs, hard drives, floppy disks, SD cards, data tapes, and optical disks (CDs and DVDs). Some of these devices require specific hardware or software to access the information and data stored within, whilst others are formatted to only be read by specific operating systems (OS), e.g. Mac formatted floppy disks. This poses quite the challenge from a digital preservation perspective as it increases the risk of information becoming inaccessible to users when technology either changes or becomes obsolete.

# Key steps to protect born-digital records during extraction

Although the act of moving files across devices and systems is often a straightforward process, it comes with a potential risk of important data being unintentionally modified, corrupted, or deleted. To assist agencies with this process, there is a sequence of specific digital preservation tasks (as outlined in the diagram below) designed to safeguard the ongoing integrity of born-digital records as they are extracted off their digital carriers.

A diagram of data processing

Description automatically generated with medium confidence

# Opening digital carriers in a secure manner

A purple floppy disk with white text

Description automatically generated

Digital carriers, just like the digital records they house, are extremely fragile. Merely opening a device can affect the metadata of the files stored on it (such as time and date stamps), which is why it is important to ensure that all digital carriers are opened/read in a secure and controlled manner before any further actions are performed to access the content.

Access measures will vary depending on the type of digital carrier being read. Below is a non-exhaustive list of the more common types of digital carriers agencies will most likely encounter in their collections and the recommended steps to ensure safe access to the records.

|  |  |
| --- | --- |
| Digital Carrier | Proper access measures |
| Optical disk (CD / DVD) | Optical disks can be read on any computer or laptop containing either a built-in disk drive or a USB port that can support a portable CD/DVD reader.  Most optical disks are ‘write once, read many times’, which means that content stored on the carrier cannot be added to or changed when it is being read. However, for a CD/DVD that is read/write, it is recommended that a disk drive without writing capability is used to open the disk so that no data is accidentally altered. |
| 3.5-inch floppy disk | Standard 3.5-inch floppy disks can be read on portable USB Floppy Drives. There are some instances where a floppy disk has been formatted to a specific software that is required to access the files – Mac formatted disks, for example, can only be read on Mac computers.  In terms of securely opening the device, most floppy disks have write-protector tabs on the back that can be physically switched to a lock position. This mechanism blocks any overwrite ability on the disk, meaning that the files cannot be edited or deleted. |
| USB / Hard drive  A close-up of a usb flash drive  Description automatically generated  A black hard drive with a cord  Description automatically generated | USBs and external hard drives are extremely convenient for the temporary storage of born-digital records. Their portability makes copying and moving data between devices easy. However, this simple feature also makes the material stored on these digital carriers extremely susceptible to unintended change or data loss.  For this reason, it is highly recommended that any USB or hard drive is first connected to a computer via a write blocker before it is opened. Write blockers allow information to be read on a drive whilst preventing the computer system from overwriting or altering the content and its metadata.  **Note:** write blockers come in either hardware or software form. Hardware write blockers are primarily recommended over software versions as they operate independently from a computer system and are not affected by OS updates. |

For advice on how to securely access other types of media, please contact the Digital Archives Team at QSA.

### A quick word on virus checks

Whether or not digital carriers in a collection are actively used, the moveable nature of these devices makes them open pathways in an agency’s computer network for potential viruses and malware. It is therefore good practice to always scan digital carriers with antivirus software before interacting with the content.

# Safely extracting files from their carriers

A orange arrow pointing down

Description automatically generated

Any extraction of records from their digital carriers must be done in a manner that mitigates both the potential loss of significant metadata and unintended change or corruption of the files themselves. Fortunately, there are a few methods to safely extract files. Selecting the appropriate one will depend on the state of the carrier and/or the born-digital records.

Below are some recommended extraction approaches for agencies to consider along with further details on when each method is applicable.

### Logical file transfer

A logical file transfer involves the migration of targeted files from one device to another without changing any technical metadata. It is not to be mistaken for simply copying and pasting files as a lot more work happens in the background to protect the files from any alteration. Logical file transfers are useful when only specific records or folders on a digital carrier need to be extracted.

There are a handful of available tools for conducting logical file transfers. Two of the most common ones include:

* **Bagger –** an application designed to package and store groups of files together in digital containers called ‘bags’, which are then safely relocated to a chosen destination (i.e. another computer). Each bag contains a manifest of the packaged files and checksums that help verify the data during transfer and storage. More information about Bagger can be found on the GitHub website (link below).

[GitHub - LibraryOfCongress/bagger: The Bagger application packages data files according to the BagIt specification.](https://github.com/LibraryOfCongress/bagger?tab=readme-ov-file)

* **TeraCopy –** a file transfer tool used for safely copying and moving digital content. TeraCopy can be used to move content at both file and folder level. The tool provides a detailed log of all content that is copied over from on device to another and generates a list of checksums that can be used to validate the files post transfer. More information about TeraCopy can be found here:

[TeraCopy for Windows - Code Sector](https://www.codesector.com/teracopy)

### Disk image

Disk imaging is the process of creating a bit-by-bit replication of a digital carrier and storing it in a single compressed file – often referred to as an image – without making any changes to the original data. The purpose of a disk image is to retain an authentic copy of the entire readable area of a device including all visible files and metadata, but also those that have been deleted or hidden, ensuring that nothing is left behind. This means that the average disk image file is roughly the same size as the carrier being imaged.

Disk imaging is an appropriate extraction method for when a digital carrier is an obsolete media or there are signs of damage to the device. It may also be useful if the operating system or settings on a device are required to enable access to the digital records.

There are a range of disk imaging tools available; some of the more widely used ones include:

* **FTK Imager –** an application that creates forensic images of devices and enables users to preview the contents of disk image files.
* **Mini Tool Partition Wizard –** a free partition manager for Windows which supports disk imaging and space analysis of devices.
* **Clonezilla –** open source disk imaging software.

Once a disk image is created, agencies will need to extract the records from the disk image prior to transfer through ArchivesGateway. At this stage, QSA is only accepting records at an individual file level and not the disk image itself.

# Analysing the files and accompanying metadata

A magnifying glass with a graph in it

Description automatically generated

There are several key things to look for when analysing content that has been extracted from a digital carrier. The most critical is checking the integrity of the files to guarantee they have not been damaged or compromised. This is usually achieved by verifying that a file’s checksum has remained unchanged.

A checksum is a computer-generated string of numbers and letters serving as a digital fingerprint for a file. Even the slightest alteration to a file will result in a complete change in its checksum. Some of the tools mentioned above help streamline integrity checking tasks by combining checksum validation with the extraction process. Another approach is to generate checksums of files first and then validate them after they have been extracted to confirm there is no change.

For more information about checksums and how to generate and validate them, please refer to the [Checksum Requirements Guide](https://www.forgov.qld.gov.au/__data/assets/word_doc/0026/456902/digital-archive-agency-toolkit-checksums-requirements.docx).

Another important step in the analysis phase is to verify that the metadata of the born-digital records has remained intact. This may include checking the creation and last modified dates/times of files, file extensions (.docx, .pdf, .jpeg etc.), titles and creator information. This is also an excellent time to collate the mandatory metadata to complete the transfer list required for transferring permanent records to QSA. For further information about this process, please refer to the [Transfer List and Metadata Requirements Guide](https://www.forgov.qld.gov.au/__data/assets/file/0027/456903/digital-archive-agency-toolkit-transfer-list-and-metadata-requirements.docx).

#### An important reminder:

QSA understands that not every agency will have the resources to safely extract permanent records stored on digital carriers and other platforms. The Digital Archive Team is set up to work with complicated digital records and media found in agency collections and is ready to assist with any migration activities. If in doubt, please contact us first at [digitalarchive@archives.qld.gov.au](mailto:digitalarchive@archives.qld.gov.au)

We are here to help!