

# Image File Management

MARION CRICK

ePic Project, Royal Botanic Gardens, Kew, Kew Green, Richmond, TW9 3AE, United Kingdom. E-mail: [m.crick@rbgkew.org.uk](mailto:m.crick@rbgkew.org.uk)

**Key words:** Digital imaging, image file management.

---

Image file management is an aspect of specimen imaging which runs right through the digitisation process.

There are a number of issues to be taken into account in image file management

- Image file standards
- Image storage
- Image organisation
- Software

## Image file standards

One of the key issues in image file management is ensuring the longevity of the images produced. This is done by producing benchmarks for your image file standards and ensuring that they are adhered to. Benchmarks will change according to the way in which the image will be used.

When imaging you will create archival files, master files and delivery files. Different benchmarks will apply to the different types of file.

## File formats

Image file formats can be divided broadly into lossless and lossy formats and into open and proprietary formats.

Lossless file formats are ones in which the original image data is stored intact or compressed in such a way that the data is unchanged when the image is later decompressed. As a result, the image files are more faithful to the original. They are also very large. Examples of lossless file formats are TIFF (uncompressed in its plainest form) and PNG (uses lossless compression). Lossy file formats use compression within the image file. The compression works by removing image data. As a result the image file is very small when compared to the lossless format. The most well known lossy format is JPEG. One of the more recent formats available is JPEG200 which comes in both lossy and lossless versions.

Open formats are those which are the source code for the format is available to software developers. It means that the file can be read in a variety of formats and that the format is likely to be sustainable in the long term. Examples of open formats are TIFF (this is not fully open but has been so well documented that it is widely supported), PNG, JPEG and JPEG2000. Proprietary formats are those which can only be read in a limited number of applications. Their maintenance is dependent on the company which developed them. Examples of proprietary formats are MrSID and PSD.

## **File compression**

Images stored in lossless formats tend to be large and require large amounts of storage. For this reason, many organisations prefer to compress images to reduce the storage requirement. There are two ways in which files can be compressed.

***Internal compression:*** These are compression algorithms which are specific to the image file format and may be lossy or lossless. PNG files use lossless compression within the format. Both lossy (jpeg) and lossless (LZW) compression is available for colour TIFF files.

***External compression:*** Images can also be compressed using an external data compression tool to create a “zipped” file. When using external compression, it is advisable to use one which is open source (ie the source code for the algorithm is available to you so that it can be maintained). There are two open source compression tools which perform well – gzip and bzip2. Of the two, bzip2

produces a slightly higher rate of compression and halves the storage requirement for images.

If any compression is used, it is essential to note within the image metadata which compression type has been used and to maintain information about how to decompress the image within your organisation.

**Archival files:** These are files which are as close to the original captured image as possible. Where possible they should be stored in the original capture format. These files are very rarely accessed. If possible you should store the archival file in addition to the master file.

**Master files:** These are files from which all delivery images are created. For this reason the files should be high quality, high resolution files. Master files should be stored in a lossless format such as TIFF, PNG or the lossless form of JPEG2000. If compression is used then the form of compression should be noted.

**Delivery files:** These are files which are created for fast delivery to users. For this reason size is the most important issue. These files should be stored in a compressed or tiled (eg MrSID) format. Using open source formats will allow for easier access for users. These files can be processed from the master file at any time – they do not need to be created at the capture stage.

## Image organisation

When you build up a store of digital images it will become necessary to organise and document the images in such a way that they are easily accessible. There are three elements to this.

**Naming the image** – make sure the image name will be unique - do not duplicate an image name. Always use the appropriate 3 letter file extension - TIF for a TIFF file, JPG for a JPEG file.

**Organisation of images** – If you have more than a very few images it is advisable to create a series of folders for the images. Placing too many images in a single folder can affect the performance of your

computer. No metadata should be recorded in the folder names, unless it is also recorded in the image metadata.

**Metadata** – All images should be accompanied by metadata. All image file metadata should be contained in an image management system, specimen metadata can be contained in the same database or else in a specimen database linked to the image database.

## **Image storage**

There are a number of image storage methods which can be considered. It is essential that, when you select a method for image storage, you consult closely with the IT department within your organisation. They can advise best on current costs and on solutions that fit within your organisation's existing infrastructure.

### **CD/DVD:**

Specimen imaging produces large image files. This makes storage on CD impractical as CDs hold up to 700mb of data. DVD-R stores up to 4.7GB of data, which makes it more viable for a digitisation project. If storing images on CD or DVD is it essential to follow the following guidelines.

1. Use archival quality discs
2. Do not burn the disc at top speed
3. If you intend to use the images on a regular basis, create a second "working" copy of each CD.
4. Check each disc after it has been written.
5. Store the discs in suitable conditions – a cool, dark, dry place.
6. Re-check the discs on a regular basis and replace the discs periodically (approximately every 5 years).
7. Create a filing system for the discs and ensure that the location of each image is recorded in its metadata.

### **Local hard drive:**

Images can be stored on a PC hard drive. This method of storage means that images can be quickly accessed. However, it does mean

that access to the images is limited and the storage capacity is limited by the capacity of the hard drive. There is an additional problem in that the images are vulnerable as they could be lost if the PC breaks. Images stored on local hard drives must be backed up to another format on a regular basis.

### **Networked storage:**

There are a number of ways of approaching networked storage.

1. Storage on a file server.
2. Network addressed storage.
3. Content addressed storage.
4. Storage area network.

The most appropriate of these types of storage will depend on your organisation's infrastructure. Selection of networked storage in particular should be discussed with your own IT department.

The advantages of these types of storage are that you can store far larger numbers of images as there is capacity for extending the storage, the images can be shared easily within the organisation, the images are held in a central location and the images can be backed up easily.

### **Image management and serving software**

Although it is possible to manage a small number of images without additional software, if you are building a large image collection, it is advisable to use some sort of image management software to do so.

Before approaching any image management software provider, it is essential that you determine your image management requirements and consult with your organisational IT department about your technical infrastructure.

TASI<sup>1</sup> identifies four different categories of image management application.

**Folder views** – these do not really manage the images in any way but provide an access route to the images either through browsing or a very simple search. Some allow very basic metadata to be captured – but this metadata is contained within the image file rather than held in a separate database. These systems are very cheap, often available as shareware.

**Simple image management systems** – these systems are often used by image users with small collections rather than by image librarians. They create basic flat file databases for the image metadata. They are generally suitable for small to medium collections of images but sometimes have server editions which allow larger databases to be built. Because the underlying database is generally in a proprietary format they are often not interoperable with other databases. Development of the database is often limited but some do offer scripting solutions. Some applications allow basic image processing – e.g. creation of thumbnail and screen resolution images, cropping and rotation of images. These systems can be created fairly cheaply. Always find out whether free systems are proprietary or open-source. In the former case, you might find that what is free now might not be free a few years later. In general, users of proprietary software (whether free or not) need to be concerned about being locked in by the use of proprietary exchange standards. Thus, if a management system has a good export story for metadata (e.g. plain text, XML, spread sheets, etc.) then users have some insulation against loss of their metadata should they have to migrate to another system.

**Full image management systems** – such image management systems have are similar to simple image management systems in overall appearance but very different in underlying structure. Some applications allow basic image processing – e.g. creation

---

<sup>1</sup> <http://www.tasi.ac.uk/advice/delivering/comparison-ims.html>

of thumbnail and preview images, cropping and rotation of images. The databases offer greater extensibility and interoperability and the development allows for a more complex database structure. The disadvantage is that these systems are somewhat more expensive than a simple content management system.

***Bespoke image management systems*** – many organisations have created their own image management system in the absence of commercial software applications which fulfilled all their needs. If outsourced this is an extremely expensive approach to take, but does allow for a precise specification to be met. As Kew has an applications department the expense is not such a great factor but the specification would need to be carefully planned.

Within these categories, the boundaries are becoming blurred – as simple image management systems become more advanced the increased functionality and ability to user server based additions means that the difference between the simple and advanced image management systems has become much smaller.

### **Collection/Digital asset management systems**

In addition to the systems described by TASI, a number of applications are broadening their scope to handle a wider range of digital files. There seems to be a trend towards complete asset management rather than purely handling image management.

### **Image serving software**

At the simplest level, an image serving application is one which makes images available for other applications, specifically one which allows images to be viewed through a web browser. This function is supplied by some image management applications through web interfaces.

There are also specific image serving solutions which can be used in conjunction with full/bespoke image management systems to improve delivery of images.

**Image authoring:** Some image applications offer image authoring tools which allow images to be saved in proprietary formats so that they can be fully utilised in the image serving application.

**File formats:** Many image serving applications will support a variety of image file formats and make them available to web browsers without having to store an additional set of JPEG or PNG surrogate files.

**Scalable image delivery:** Some applications will use “tiled” image formats such as SID or Flashpix to deliver scalable images through web browsers. It is only necessary to store a single copy of the image and the software automatically generates images on the fly from the original. This means that the software reads only as much of the image as is required for the specific viewing request and delivers only the portion of the image required. This may be the entire image at different resolutions or parts of the images. This feature means that network traffic is greatly reduced and that higher quality image files can be delivered across limited Internet bandwidth.

**Zoomable images:** Applications often have features which allow the user to zoom into an image to see details at higher resolution.

**Navigation within images:** Applications often allow the user to navigate within the image – a small portion of the image is shown inside a frame and the user can use tools to view adjacent portions of the image.

**Choice of ways to access the images:** The images served through the applications may be viewed either through a free browser plug-in, which has to be downloaded by the user, Java applets or even through plain html pages.

**Image security:** Some software applications offer the facility to give the appearance of a watermark on the image file when the images are served. Some actually watermark the image.

## Examples of image management applications

### Image authoring and editing software

**Product:** Zoomifyer

**Manufacturer:** Zoomify

**Website:** <http://www.zoomify.com/zoomifyer/competitive.asp>

**Features:** Zoomifyer is an authoring tool rather than a server application. Images can be viewed by end users using plug-in (Netscape and Mac users), ActiveX controls (IE users), Java applet, clientless Javascript option all of which are generated by the authoring tool. Generates images in proprietary PFF format. The next release of Enterprise should directly support JPEG2000. Compatible with any web server. Visible watermark can be placed in the viewing window. Zoomifyer Enterprise can work as a desktop application or as an automated application on a network server (Mac or Windows, but not Unix).

**Product:** ImageZoom

**Manufacturer:** Scalado

**Website:** <http://www.scalado.com/index2.html>

**Features:** Allows you to create “zoomable” images – where sections of the image can be enlarged. The images cannot be panned across. Creates multiple image files – a large image and several higher resolution zoomed sections of image - rather than a single image file. This works as a desktop application which runs on either Windows or Mac platforms.

**Product:** GeoExpress

**Manufacturer:** Lizardtech

**Website:** <http://www.lizardtech.com/>

**Features:** Writes images to the SID format, which is a multi-resolutional, tiled format, allowing portions of very high resolution images to be viewed selectively. Compresses the image file. Supports image file sizes of over a gigabyte.

**Product:** ImageMagick

**Manufacturer:** ImageMagick

**Website:** [www.imagemagick.com](http://www.imagemagick.com)

**Features:** Open source desktop or server based imaging application. The software can be run as a script or used through a variety of interfaces, including Java. The application supports the majority of file formats, including JPEG2000 and also supports some vector graphic formats. It has a number of different utilities which perform different functions. Convert and mogrify do one time image conversion, display offers on the fly conversion of image files, montage automatically builds thumbnail galleries.

### Image serving software

**Product:** eyespy

**Manufacturer:** AXS technologies

**Website:** <http://www.axs-tech.com/html/products/eyespy/index.html>

**Features:** Creates tiled images files for web delivery – images can be zoomed into and panned. Supports 2 major file formats and plans to support JPEG2000. Allows different interfaces for internal and external users.

**Product:** MediaRich.

**Manufacturer:** Equilibrium

**Website:**

<http://www.equilibrium.com/Internet/Equil/Products/MediaRich/Product+Tour/index.htm>

**Features:** Image creation and processing tools, can be easily integrated with open standard databases, supports wide range of file formats, delivers an appropriately sized image for the viewing device, allows image files to be repurposed for many uses from a single source image, creates images which can be zoomed into and panned across. MediaScript creates the javascript code which allows the images to be viewed through a web interface. Requires Microsoft or Solaris operating systems and IIS or Apache web server.

**Product:** Zoom Image Server

**Manufacturer:** iSeeMedia

**Website:** <http://www.iseemedia.com/products/zoom/index.html>

**Features:** Scalable image delivery from a single image, zooming and navigation within images. Images can be viewed with a plug-in

or a Java applet, supports Flashpix and IVUE file formats. Can be used on Windows, Linux and Solaris web servers. Enterprise edition allows images to be watermarked.

**Product:** Web Image Server

**Manufacturer:** Vimas Technologies

**Website:** [http://www.vimas.com/ve\\_image\\_server.htm](http://www.vimas.com/ve_image_server.htm)

**Features:** Supports TIFF, JPEG, PNG, TGA file formats. Application will only run on a Windows NT or Windows 2000 server. Allows images to be edited through a web interface and allows the user to resize the displayed image.

**Product:** Express Server

**Manufacturer:** Lizardtech

**Website:** <http://www.lizardtech.com/>

**Features:** ExpressServer allows SID images, as well as high resolution JPEGs, TIFFs and other image formats to be delivered as scaled JPEG files through a web browser, either with our without a web plug in.

**Product:** Image Web Server

**Manufacturer:** Earth Resource Mapping Pty Ltd

**Website:** <http://www.earthetc.com/>

**Features:** Images available to web browsers and to other applications, mainly GIS applications. Saves images in compressed ECW format. Images can be viewed through a web browser using an automatically downloaded plug-in. The plug-in allows images the user to define the extent of the image file that they wish to download in each instance (resolution, portion of image). The application is only suitable for Windows servers and works as an extension to Microsoft IIS.

## **Image management applications**

**Product:** iBase Manager

**Manufacturer:** iBase

**Website:** <http://www.ibase.com/site/software/index.shtml>

**Features:** Allows images to be catalogued and batch processed to create smaller image files – the format of the batch-processed files is not specified. Additional modules are available to create a touch-screen browser and a web interface. The application has cross

platform client capability but will only run on a Microsoft Windows Server.

**Product:** Fotostation/ Index Manager

**Manufacturer:** Fotoware

**Website:** [www.fotoware.com](http://www.fotoware.com)

**Features:** Software developed mainly for use in publishing, specifically newspaper, industries. Two applications are used in conjunction – Fotostation allows image creation and manipulation workflow to be streamlined and provides the thumbnail interface, Index Manager created proprietary format databases in “open industry standard” and has a search engine. Index uses IPTC standard fields and allows 20 custom fields. Windows NT based.

**Product:** IDA II (Professional edition)

**Manufacturer:** Aequitas

**Website:** [www.aequitas.co.uk](http://www.aequitas.co.uk)

**Features:** Image management application that handles imaging workflow and supports relational databases held in SQL server and Oracle. The database is customisable and the application supports JPEG, BMP, TIFF and PCD images. Both the server and client software only run on a Windows platform.

**Product:** Picassa

**Manufacturer:** Google

**Website:** <http://picasa.google.com/>

**Features:** Proprietary free system for Windows. Designed for home use, but suitable for many sizes of collections. Application has a very flexible album structure, captions and other metadata said to be stored in IPTC standard (see MORRIS, in this book, pp. ##-## on Metadata Standards). There is an active discussion group about Picassa on Google Groups (<http://groups.google.com/group/picasa/>) and some complaint that the IPTC data is not fully portable to other IPTC aware systems, though the consensus is that exchange with many important image annotation systems, including Adobe Photoshop and Menalto Gallery (see below), is generally satisfactory. A documented CSS template-based export

### **Collections/Digital asset management systems**

**Product:** The Museum System (also TMS light and embARK)

**Manufacturer:** Gallery Systems

**Website:** <http://www.gallerysystems.com/>

**Features:** This is a museum administration database rather than an image management system. Image files are handled in the “media” module but can be catalogued in the collections module. This is a complex system which has a number of sophisticated data entry options and allows the user to link to an integrated thesaurus. The database can be held in SQL server or Oracle. A web interface is provided through the e-museum application which needs to be purchased separately. The applications are heavily reliant on Microsoft and optimum performance is only gained with a Microsoft database server.

**Product:** Index+

**Manufacturer:** System Simulation

**Website:** [www.ssl.co.uk](http://www.ssl.co.uk)

**Features:** This is a full digital asset management system which can handle resources other than images. This is a bespoke system which is tailored to each organisation’s needs. The metadata is held in an object-oriented database which is fully interoperable. System Simulation also builds interfaces which allow Index+ to interact with existing organisational databases.

**Product:** Portfolio

**Manufacturer:** Extensis

**Website:** [www.extensis.com/portfolio](http://www.extensis.com/portfolio)

**Features:** A simple image management system. It allows you to build a customised set of database fields, easy import and export of data as a delimited text file and can be used to generate preview files for images. It also automatically captures technical image metadata. The data is held in a proprietary file format but can be exported to a delimited text file. Images and metadata can also be exported to web based galleries. There are opportunities to develop the software interface further through scripting in Visual Basic or Applescript. The software also allows for records to be created for hard copy images as well as digital images. This application is being used as an interim image management system until the image server has been created. Cross Platform capability. Runs as standalone or on Windows or Mac servers.

**Product:** Canto Cumulus

**Manufacturer:** Canto

**Website:** [www.canto.com](http://www.canto.com)

**Features:** This application is described as a digital asset management system rather than an image management system, as it handles a large range of digital files. The underlying database is held in a proprietary format which is ODBC compatible and the field structure can be customised. Previously the data could only be exported in a proprietary data format but the latest release allows data to be exported to XML. There is a 4GB limit on the database size. There are a number of additional applications which Canto provide to enhance the software. Of chief interest are the database connector application, which allows you to connect the application interface to a variety of database types, and the web publishing application which serves the images and database through a web browser. Neither are integral to Cumulus but are designed to work alongside it. Canto provide a software development kit to customise the client application using visual basic or applescript. Both PC and Mac clients. The workgroup and enterprise editions are compatible with a Unix server.

### Hybrid applications

In addition to the applications which offer elements of the image server, there are some applications which have all of these features – offering some image management, manipulation and serving functions.

**Product:** KE Emu

**Manufacturer:** KE Software

**Website:** <http://www.kesoftware.com/emu/>

**Features:** Collections management system with an images module, which allows images to be connected to other databases within the system. It is server based and runs on Windows PCs, connected to either Windows or Unix servers. Data follows established museum standards. Uses KE Texpress database, which is an open systems database. Web serving software is available.

**Product:** Toadview and ToadHMS server editions

**Manufacturer:** Oxford ArchDigital

**Website:** <http://www.oxarchdigital.com/toadview/>

**Features:** This is a server based image database based on open source applications (MySQL, imagemagick). The application has a web browser interface so there is no need to install client software

and both PC and Mac use is supported. The application supports several image file formats and creates thumbnail and JPEG preview images when the image files are loaded into the system. Allows basic image manipulation. Also acts as an image server – resizes and watermarks original images for web delivery. Metadata can be stored in a variety of database formats including MySQL. Adaptable metadata framework. Has advanced image security and access features.

**Product:** Insight

**Manufacturer:** Luna Imaging

**Website:** <http://www.lunaimaging.com/insight/index.html>

**Features:** Java based application. The application is used through a web browser. This application combines an image management application with an image serving application. The Insight image server incorporates LizardTech's MrSID Photo code. The server side of the application allows scalable image delivery from a single file, zooming and navigation within the image files. No plug-in is required to view images. The management application can be based around a variety of databases including Oracle, MS SQL and Sybase. Has a number of thesauri loaded. Has workflow management features.

**Product:** Menalto Gallery

**Manufacturer:** Menalto.com (open source)

**Website:** <http://gallery.menalto.com/>

**Features:** Mature open source, web-based image server software that manages albums, automates thumbnailing, and provides for limited collaborative image annotation and web search against that annotation. A companion program, GalleryRemote provides for uploading from Windows and Macintosh machines directly to the server. The system supports mirroring, whereby copies of images can reside on multiple servers with the web client able to select which mirror to use for best service. It can be embedded in a number of content management systems, including the widely used open source Mambo system (<http://mamboforge.net/>) Gallery2 as completed two release candidates and is nearing its final release at this writing. While version 1 uses its own database, Gallery2 is SQL based and includes support for MySQL, Oracle, and Postgres databases.