File Formats:

Brief Overview of Some File Formats

Remember, the "digitization" of a document is initially just taking a photograph of the page. As a result, the file formats involved are the same as those involved in photography. Here is some info from Wikipedia and other sites.

RAW

v   Raw image files are sometimes called digital negatives, as they fulfill the same role as negatives in film photography: that is, the negative is not directly usable as an image, but has all of the information needed to create an image.
v   Like a photographic negative, a raw digital image may have a wider dynamic range or colour gamut than the eventual final image format, and it preserves most of the information of the captured image. The purpose of raw image formats is to save, with minimum loss of information, data obtained from the sensor, and the conditions surrounding the capturing of the image (the metadata).
v   There is no single raw format; formats can be similar or radically different. Different manufacturers use their own proprietary and typically undocumented formats, which are collectively known as raw format.
v   Raw files contain, by necessity, the information required to produce a viewable image from the camera's sensor data. The structure of raw files, including the ISO standard raw image format ISO 12234-2, TIFF/EP, often follows a common pattern, that is:
  ·   A short file header which typically contains an indicator of the byte-ordering of the file, a file identifier and an offset into the main file data
  ·   Camera sensor metadata which is required to interpret the sensor image data. This includes the size of the sensor, the attributes of the CFA and its colour profile
  ·   Image metadata which is required for inclusion in any CMS environment or database. These include the exposure settings, camera/scanner/lens model, date (and, optionally, place) of shoot/scan, authoring information and other information.
  ·   An image thumbnail
  ·   Optionally a reduced-size image in JPEG format, which can be used for a quick and less computing-intensive preview.
  ·   In the case of motion picture film scans, either the timecode, keycode or frame number in the file sequence which represents the frame sequence in a scanned reel; this is the most important metadata item, because it allows the file to be ordered in a frame sequence (without relying on its filename).
  ·   The sensor image data.

TIFF

- Tagged Image File Format (abbreviated TIFF) is a file format for storing images, popular among Apple Macintosh owners, graphic artists, the publishing industry, and both amateur and professional photographers in general. As of 2009, it is under the control of Adobe Systems.
- The TIFF format is widely supported by image-manipulation applications, by publishing and page layout applications, by scanning, faxing, word processing, optical character recognition and other applications.
- TIFF is a flexible, adaptable file format for handling images and data within a single file, by including the header tags (size, definition, image-data arrangement, applied image compression) defining the image's geometry.
- The TIFF format is the standard in document imaging and document management systems and can save multi-page documents to a single TIFF file rather than a series of files for each scanned page.
- TIFF offers the option of using a lossless data-compression technique for reducing a file's size [ie. No data is lost].

JPEG

- JPEG stands for "Joint Photographic Expert Group" and, as its name suggests, was specifically developed for storing photographic images. It has also become a standard format for storing images in digital cameras and displaying photographic images on internet web pages.
- JPEG files are significantly smaller than those saved as TIFF, however this comes at a cost since JPEG employs lossy compression [a method which discards (loses) some of the data]. A great thing about JPEG files is their flexibility.
- JPEG files achieve a smaller file size by compressing the image in a way that retains detail which matters most, while discarding details deemed to be less visually impactful. JPEG does this by taking advantage of the fact that the human eye notices slight differences in brightness more than slight differences in colour. The amount of compression achieved is therefore highly dependent on the image content; images with lots of detail will not be as easily compressed.
- JPEGs are popular for images on the web as their smaller size allows them to load faster.

Library and Archives Canada Internal Digital Standards:
http://www.collectionscanada.gc.ca/digital-initiatives/012018-1200-e.html