**DRAFT <“READ-ME” 2 > FILE (Use of ORI Forensic Tools),** December 13, 2011

**EVOLVING MEANS TO INSPECT QUESTIONED IMAGES**

There is not likely to be only one approach that is sufficient for detection of image tampering. One reason is that despite the technological advances in scientific publishing (i.e so-called “pre-flight” aspect of translating a manuscript to the printed or online figure), the authors rarely produced the submitted images by common software. Submitted images are invariably cobbled together using a miscellany of standards, settings, and image file conversions. The amorphous pre-submission environment challenges the use of sophisticated digital forensic software based solely upon computerized detection of statistical anomalies within the bit-mapped image. Even with controlled test images any method is not a guaranteed as a safeguard, as false negatives (and sometimes false positives) occur. The amorphous pre-flight environment encumbers any approach predicated on detection of mathematical anomalies within a single image; the latter also does not appear to be designed to detect re-use of image content occurring between figures in the same document, or in different publications.

Although most allegations still arise from anomalies detected by the human eye, an image manipulated to perfection is not a defense to ultimate detection. Irrespective of their visibility, deceptions the eye or software misses in a good ‘forgery’ may surface later when a laboratory attempts to locate and examine the original raw data files after experiments cannot be replicated. The surest way to preserve confidence is to preserve data.

**Extracting Image Content with Commonly Used Software.**

Due to the enhanced portability in new software, a remarkable degree of information may be automatically embedded in -or ride along- with images that had been inserted into many common programs. The following examples are drawn from DIO’s experience with MS PowerPoint, Adobe Illustrator (for vectorized images), and to Adobe Acrobat (for -.pdf files that had been automatically converted by a file translator). However, the success of all of the following approaches are contingent on how images were originally handled, saved, or transferred between documents, and how the image was created in the translation from data acquisition to its processing for presentation in figure form. (Importantly, those uncertainties do not introduce that content which leads to an allegation.)

MICROSOFT WORD

Evidence for compositing in spliced images that were placed into word processing files may be revealed by the Region of Interest (ROI) that is shown when selecting the image within the document. A ROI that is smaller than the image’s boundary may indicate compositing, or splicing of an object into the image. In this is effect is observed, the different elements can still be separated.

MICROSOFT POWERPOINT

Journals often include the option of downloading a PowerPoint (Ppt) slide of the figure. In Ppt, use of Image Reset after Ungrouping Elements can restore many cropped images to their original size and appearance, revealing their original aspect ratios, and features that had been cropped away for the presentation of the results such as the original labeling or annotations on the source films/blots. The Image Reset may also reveal bands which have been reshaped to appear different when they are used in another image in the same paper, and it provides a tighter rational basis in comparative examinations of the same bands that have been reshaped, rescaled, etc.

ADOBE ACROBAT (Pro 9 or 8)

PDF documents are becoming a convenient way to submit articles for publication, since they provide a (seemingly) uniform standard. They are universally available from online journals, and they are a common way to submit grant applications. If the images in the PDF document have been vectorized, use the “Image Touch up Tool” can reveal spliced or cropped images. As before, one good indication is if the ROI for the selected image is *either larger or smaller* than the image boundaries in the displayed figure. Compositing may be revealed as well as the presence of overlaying splices, which can be moved by dragging to reveal other information that has been covered up. (The latter finding is a good indication that there was a knowing misrepresentation.)

One may need to distinguish between separable components of the image that (rightly or wrongly) have been intentionally composited by the author, and those separable components that in some cases can be “tiled” or “sliced,” an effect introduced by the file conversion to make the pdf document. The key to identifying the latter event or stitching artifacts, is to ask whether there is a pattern to the image tiling/slicing that indicates a lack of purpose with respect to the image’s scientific content: Does the event occur in the same way over the other images and is it only horizontal, being bounded at the same height in neighboring figures?

Possible Recovery of the Original Image: Right clicking (on a PC platform) using the Image Touch Up Tool and selection of “Edit Image” will open up the original image in Photoshop if it was embedded in the PDF file. This will often reset the original size and scaling so that one does not have to resize images by interpolation in Photoshop.[[1]](#endnote-1)

Convenient Extraction of All Images from the Document: One of the more useful tools of Acrobat Pro is the Exporting Images feature. The “Export All Images” is in the “Document Processing” option in the dropdown Advanced Menu. This feature automatically exports all images that have been embedded in the vectorized PDF file, *but any components of images that are smaller than the default setting will be missed*. To guard against this possibility, go to the Settings option when saving, and select “no limit.” The saved images will be uncropped and organized, and saved automatically with a name stipulating the document and its page location from where it was extracted. However, the size and scale of the automatically extracted image may differ slightly relative to that extracted using the Image Touch Up Tool. (In this case, the “Pixel Aspect Ratio” settings in Photoshop’s “View” menu can be examined.) The results can then be examined efficiently using Adobe Bridge (see below).

The features above will not work if the PDF file has been rasterized, or placed into a bit-mapped format. This situation can be detected if selection using the object tool reveals a thin horizontal line rather than a boundary around an object of interest. [[2]](#endnote-2), [[3]](#endnote-3)  The issue with “tiling/slicing” has already been mentioned.

Obviously, the nature of the information available in the online journal <-.pdf > files depends upon the publisher, but it is always worth looking. It is also worth looking to see whether the final prepublication manuscript is provided online, as is now done by some journals. The possibility of extracting images from any supporting supplemental files should not be overlooked.

Important Note: The quality (resolution) of the images extracted by this “Export All Images” feature will depend on the “Settings” that can be accesses at the “Save/Save As” Dialogue. Make sure to inspect those options (see settings for resolution in “Conversion” and the size limit in the “Extraction” options) to determine retrospectively whether the save resolution is sufficient to preserve the details originally present in the embedded image content. The best way to confirm the latter is the slower process of accessing the individual image in the PDF using “Copy Image” feature and then directly transferring the contents of the clipboard into Photoshop. Forensically, it is better to not resample the embedded image. Other factors should be considered.[[4]](#endnote-4) The presentation of these options varies slightly between the Mac and the PC.

ADOBE ILLUSTRATOR

If a suitable version of Adobe Acrobat is not available, similar but slightly different information about object splicing/compositing may be seen if the vectorized <.-pdf> figure is opened in Adobe Illustrator. However, only the thumbnail of the embedded image will be seen. The Illustrator <-.ai> images will also be available directly since many journals require authors to submit their figures as vectorized images to facilitate scale and font conversions used in printing and layout process.

A particularly useful feature in illustrator is that the Link Panel, since it additionally can tell you the magnification and scaling of individual panels within a vectorized figure. This constitutes a useful probe to ask whether the actual scaling of individual bands is consistent with claims that separate elements of a multi-panel figure originated in the same image-scan of a blot.

METADATA

METADATA may show the source of the original files that were used, and on what computer the file was first created. Smart laboratory management in using Photoshop might involve setting up Meta data and turning on the HISTORY LOG. It is a good practice that would help laboratory record keeping, one that potentially might help solve unanticipated problems if/when questions about an image subsequently arise, and also provide means to ‘document’ priority for intellectual property claims. [[5]](#endnote-5) Means for accessing and managing metadata for the some of the other programs are in the footnotes.

A remarkable amount of information, including even the identity of the camera, its settings, the original source file and folder can be recovered from some images that have been downloaded from the internet. Also, some documents can retain specific metadata associated with individual figures, revealing the origin or each irrespective of who created and saved the “hosting” document. Metadata can be accessed at several portals, including the security settings, and “security settings for this document”. However, the availability of this information ultimately depends on the initial settings at the creation of each of the source documents.

**Finding the “Needle-in-the-Haystack” - An Easy Way to Review Many Images**

BRIDGE

Use of an image-based “directory” software such as Adobe Bridge, which is freely bundled with Adobe Photoshop (or the inexpensive Photoshop Elements), can greatly facilitate the rapid visual identification of relevant images in large volumes of image evidence. Images do not have to be opened to be viewed, so they can rapidly be inspected visually at good resolution; images can be grouped together, either manually or automatically sorted by key words, dimensions, date, ratings, etc. As a result, looking at large volumes of image evidence does not have to be mind-numbing nor tedious. Searching can be facilitated by enlarging the thumbnail of an image, selecting it to keep the enlarged image in the Preview panel, and then scrolling through the other enlarged thumbnails in the Content Panel to look for similarities.

Comparison of questioned images after two of more of interest are identified is greatly simplified. Similar images can be assigned a common “rating” that can bring them together in the directory with subsequent sorting for selected properties. A manipulation isolated to one image can easily be detected by alternatively selecting the different thumbnails that will be displayed at large resolution in the Preview Panel. A small difference between two images is easily detected, and co-alignment of objects in microscopic images is an indication the different images had been derived from only one “observation.” Each image can then be inspected individually with ORI’s Forensic tools to see if one has visual evidence of tampering.

Bridge can be used to inspect all images that have been extracted from a PDF document using the “Export All Images” feature described above. The images will appear to be associated with specific pages. Thus, two images of the same result appearing on the same page may be an alert to compositing. Alternatively, sorting the Bridge directory by “dimensions,” by modification date, or “rating” for example, may place two images of the same blot together that are associated with figures on different pages.

The same strategy enables searching for possible reuse of the same image content in different publications. In this case, all extracted images are dumped into a common folder, and sorting by selected criteria (above) may link reuse of image components in different publications. However, when pursuing this approach the directory name of each source publication should be severely truncated *before the extraction process*, so that the auto-labeling does not result excessively long (i.e., hard to interpret) terms in the final directory of the extracted images.

Documentation: The results of the various directory sort strategies can be readily documented by using the Output Window in Bridge, adjusting to a suitable format in the Layout Options, and then saving the Output Panel as a pdf.

**DISCLAIMER**: These draft comments are under continual revision and they offered as exemplars only. Obviously, users assume responsibility for their proper application and interpretation.

Contact for questions/suggestions:

John Krueger, Ph.D.

DIO/ORI/DHHS

240-453-8432

John.krueger@hhs.gov

1. Sometimes Photoshop will not import the selected PDF object image, giving an alert about an “unsupported color space.” In this case, when in the PDF file, simply go to the Touch Up Tool, right click(PC), and select the “properties” menu at the end of the list. At this point, select the submenu for “color” and then select “convert colors.” Generally the default will suffice. With this change to the PDF’s properties setting, Photoshop may be able to import the designated image object. [↑](#endnote-ref-1)
2. Metadata in Adobe Acrobat Pro 9 can be examined by using the “Examine Document” feature in the Document Menu, expanding the results, and then selecting “show preview.” Be aware that this feature can be used to remove unwanted accessory information from the document. [↑](#endnote-ref-2)
3. It may possible that opening a file in a different platform (PC vs Mac) may restore recoverability of separate components of figures. Also, converting figures from one document to a -pdf may restore the separability of objects within a figure, but one needs to test whether this reflects true recovery of the original objects or simply the re-vectorizing of the image into new objects. [↑](#endnote-ref-3)
4. Vectorized images will be converted to bit-mapped images during extraction from pdf documents, and they may be saved at a default of the screen resolution. If the interpretation of the results is potentially affected, it is always better to check these variables by another means. [↑](#endnote-ref-4)
5. Be aware that one of the consequences of setting up a metadata file is that information about the steps of image processing may be propagated unknowingly in other documents into which the image is pasted. If that is a concern check the program. PowerPoint, for example, has an option’s setting for file preferences which can limit propagation of other data when using “Save As.” In MS Word, those options are accessed with the “Tools” button, presented at bottom left in the “Save As” window. Controlling metadata in Acrobat is described in a prior endnote. [↑](#endnote-ref-5)