The following information is being disseminated by the Radiological Society of North America (RSNA), the American College of Radiology (ACR) and the Society for Imaging Informatics in Medicine (SIIM).

Protecting Patient Information in Medical Presentations, Publications and Products

New Search Engine Capabilities Can Expose Patient Identifiers Thought to Be Anonymized

Recently, it was discovered that search engines can index patient identifiers (see appendix), such as names, in slide presentations and other documents that were believed to have been anonymized.

Background

Advances in web-crawling and content processing technology employed by search engine vendors (e.g. Google, Bing and others) increasingly enable large-scale information extraction from previously stored files. Among other things, this technology can extract source images contained in PowerPoint™ presentations and Adobe® PDF files, and recognize alphanumeric character information that may be embedded in the image pixels. As such, an image with embedded patient information can be indexed by this process. When explicit patient information becomes associated with images in the search engine database, it can be found on subsequent internet searches on the patient's personal information.

For Example:

When a patient searches her name in a search engine, images from a diagnostic imaging study performed 4 years ago appear. When she clicks on the images, she is directed to the website of a professional imaging association which stored an Adobe® PDF file as part of an educational presentation. The association was unaware that the file contained PHI. The author of the file was unaware that PHI had not been sufficiently de-identified prior to creating the original presentation in PowerPoint™ format, and that the saving in Adobe® PDF format also had not preserved privacy.

Solution/Best Practice

Only images without PHI should be included in presentations of any kind. To assure no PHI is included, screen capture software should be used to capture the image pixels for the region of interest only. Or, the user can disable patient information overlays or use an anonymization

algorithm embedded in the PACS before saving a screen or active window representation. Alternatively, the creator of the presentation can use third party image processing software (e.g. Adobe® Photoshop, Irfanview, etc.) to crop out or obscure PHI before inserting the resulting imaging information into a presentation.

Simply cropping out PHI with the image formatting tools available in presentation software (e.g., PowerPoint™, Google Slides™, Keynote®) does NOT permanently remove the PHI. Placing “black bars” or similar visual aids to obscure PHI within the presentation software also does not represent a safe and compliant practice for de-identification.

Specific functions are available in some software to permanently delete cropped, obscured or hidden information in presentation files. As a final quality control check, it is recommended that these "sanitization" functions be run on all presentations prior to being made public.

The following information provides workflow steps to consider and other educational resources for preventing inadvertent inclusion of PHI in medical presentations, publications and products.

Workflow Steps to Consider when Safely Publishing Medical Images for Education and Publication

Exporting Medical Images

The first place to pay attention to potential PHI exposure is the initial workflow step of **exporting images from the PACS or another imaging device or application**.

Optimally, a "region of interest" type screenshot is obtained which has only actual "anatomic" image pixel information in it. Alternatively, the user can disable the DICOM patient info, i.e. use the remove/hide overlays function in PACS first, and then obtain a screen shot.

Every time an image is saved directly from PACS as a file (as opposed to creating a limited screenshot), there is a risk that PHI gets into that file via patient data embedded as pixels within the image itself or in the form of metadata if a DICOM file is saved. Even when images do in fact contain PHI data it can be redacted using appropriate tools and processes. Of note in this example, the original file now resides on the local system drive and should be securely deleted.

Second, some images, for example, can hold data in Exchangeable Image Format tags (additional information stored together with pixel data), much as DICOM stores data in its tag structure. It's possible the PACS will utilize these tags to store metadata that needs to be cleaned.

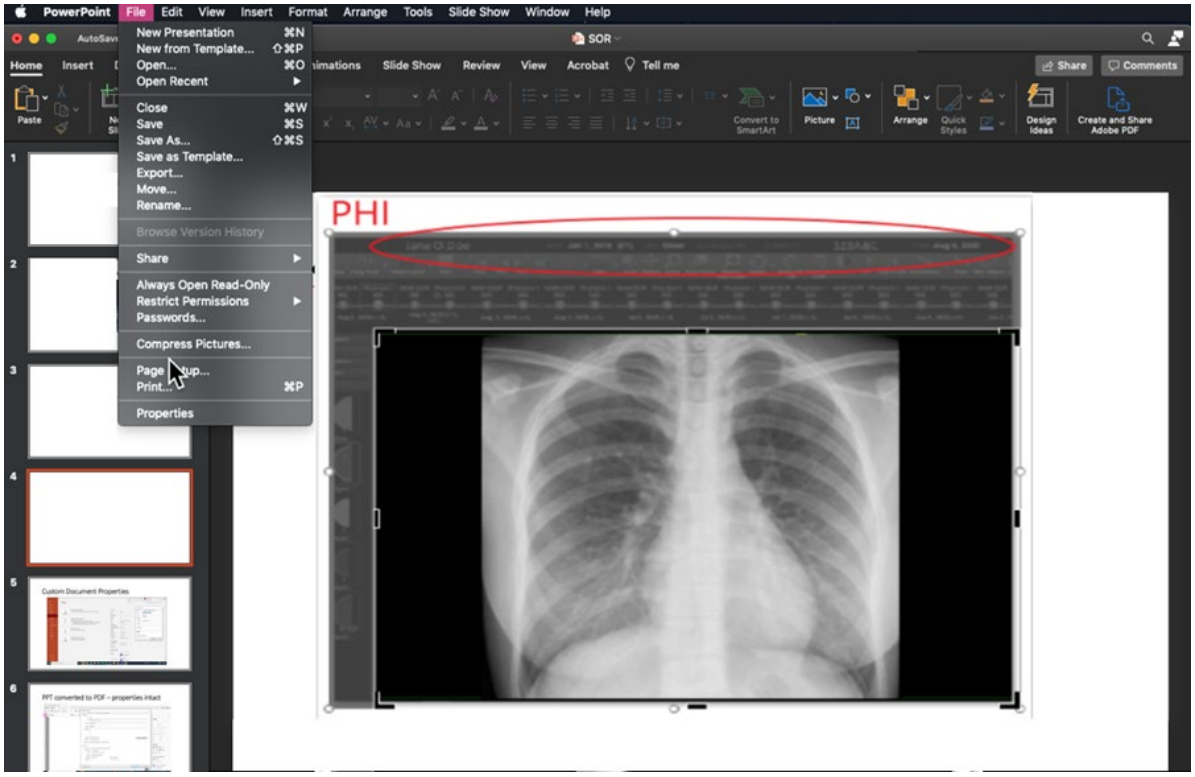
Creating the presentation

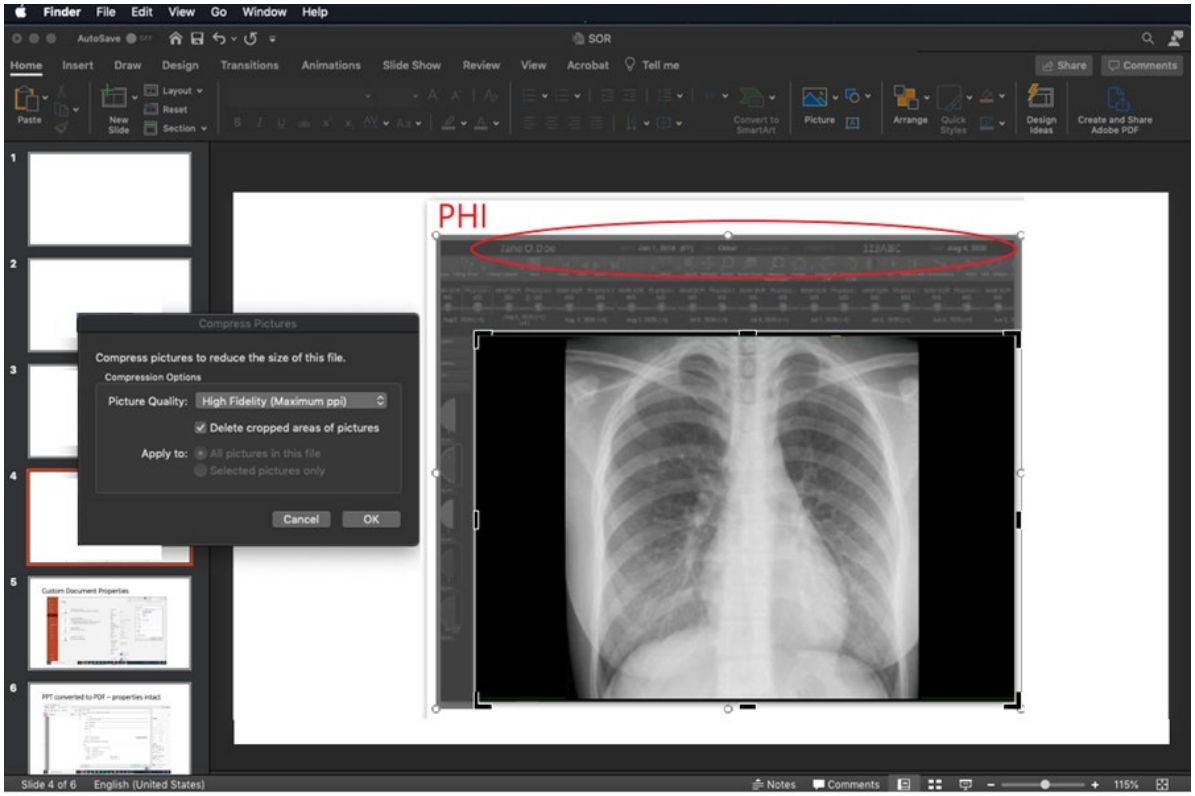
The next place to look for possible PHI exposure is during creation of a document or presentation that utilizes (exported) medical images. PowerPoint™, for example, is a Microsoft product which is frequently used to present images e.g. for medical professional education in a variety of settings.

When medical images are inserted into PowerPoint™ and the user attempts to redact burned in PHI data within the pixel data, they must be careful not to simply "cover up" the PHI by use of a mask. The most commonly used tool is cropping the image with the corresponding PowerPoint™ tool, or changing font color so the text blends into the background. Neither will result in actual removal of the information, rather it is just not visible in the authoring nor the presentation state. The "cropping" can be undone later by another user of the file. Modern search engine technology can automatically identify the content of the original inserted file and index PHI that might have been included. It is important when cropping an image to **explicitly delete** the portion that has been cropped so that it cannot later be uncropped. Microsoft has provided instructions on how to delete the cropped areas of a picture and save the file without them.

Examples in both the Mac and PC versions of PowerPoint™ follow.

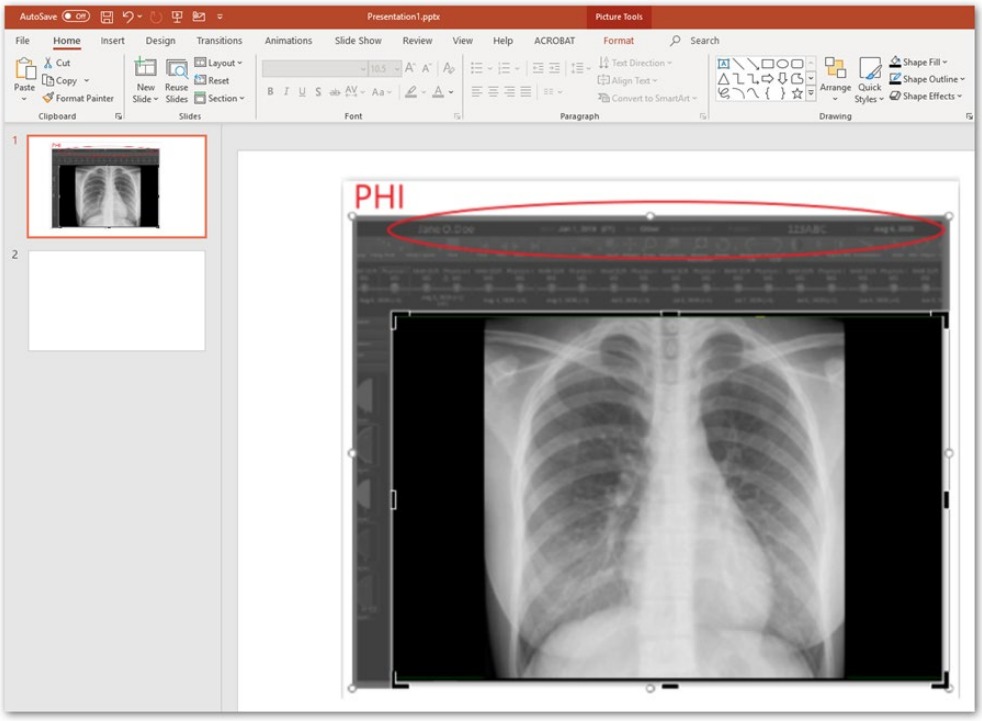
Apple Version of PowerPoint

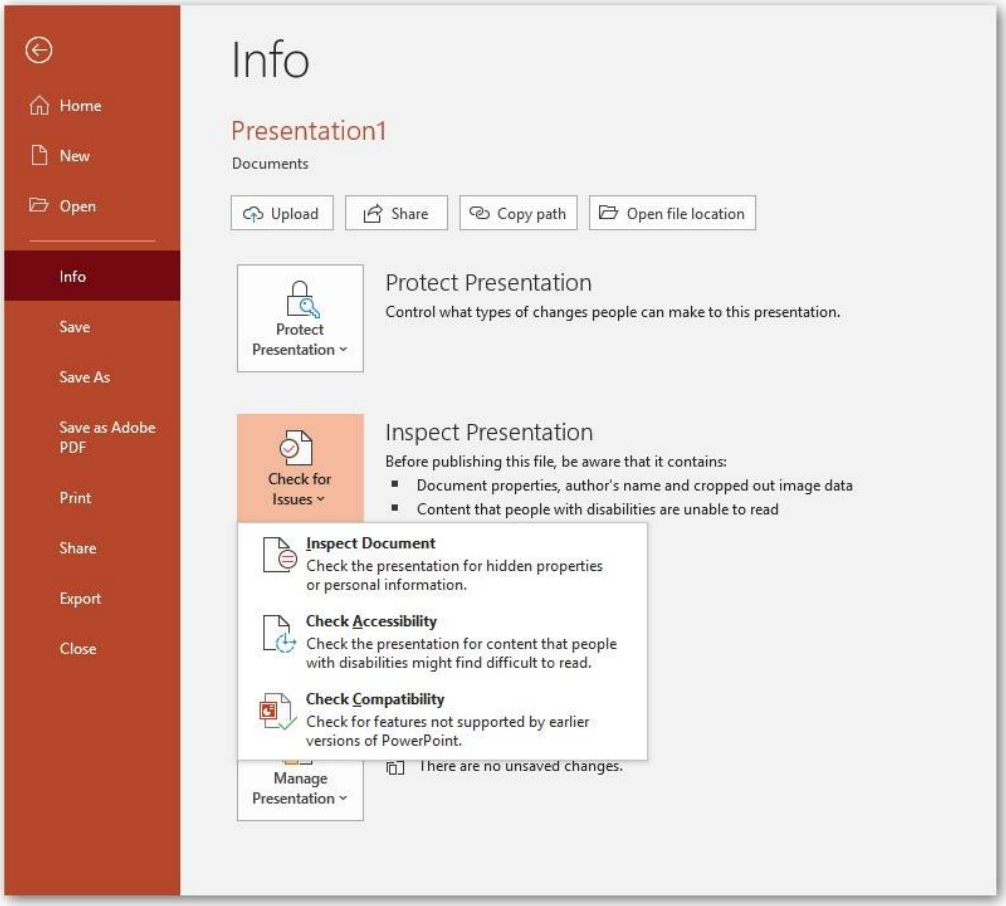


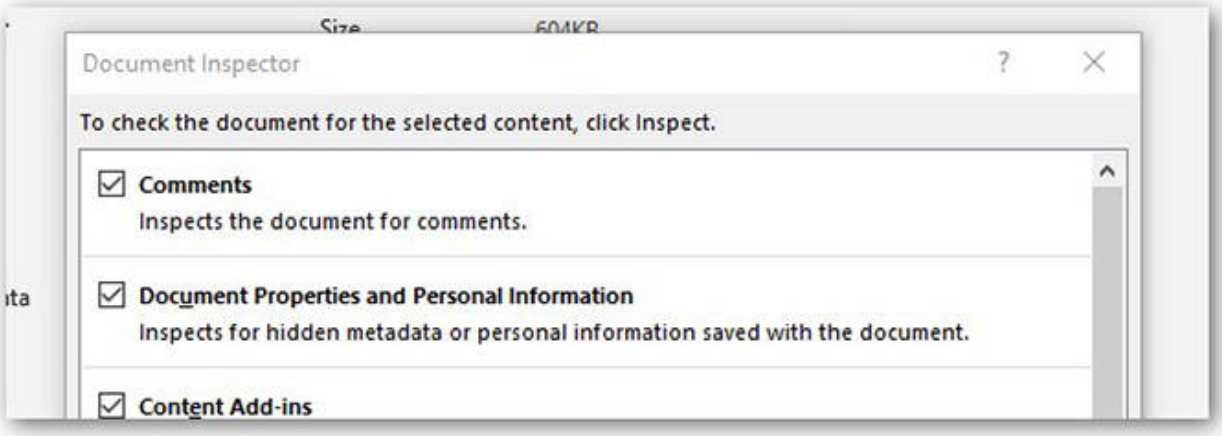


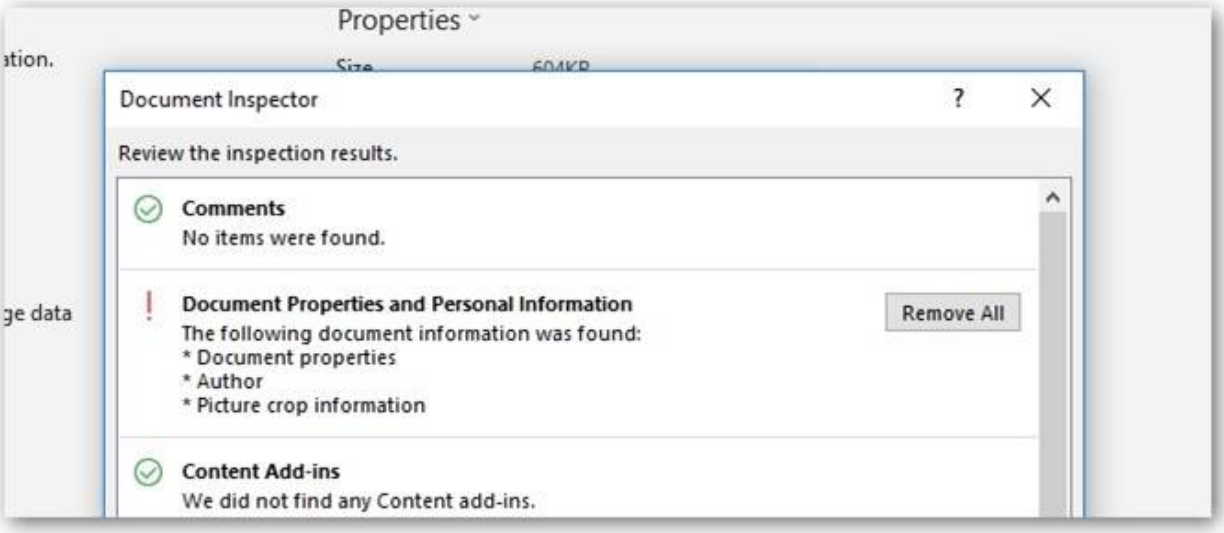
PC Version of PowerPoint

This can be done through the "Inspect Document" function in PowerPoint™ prior to publishing.



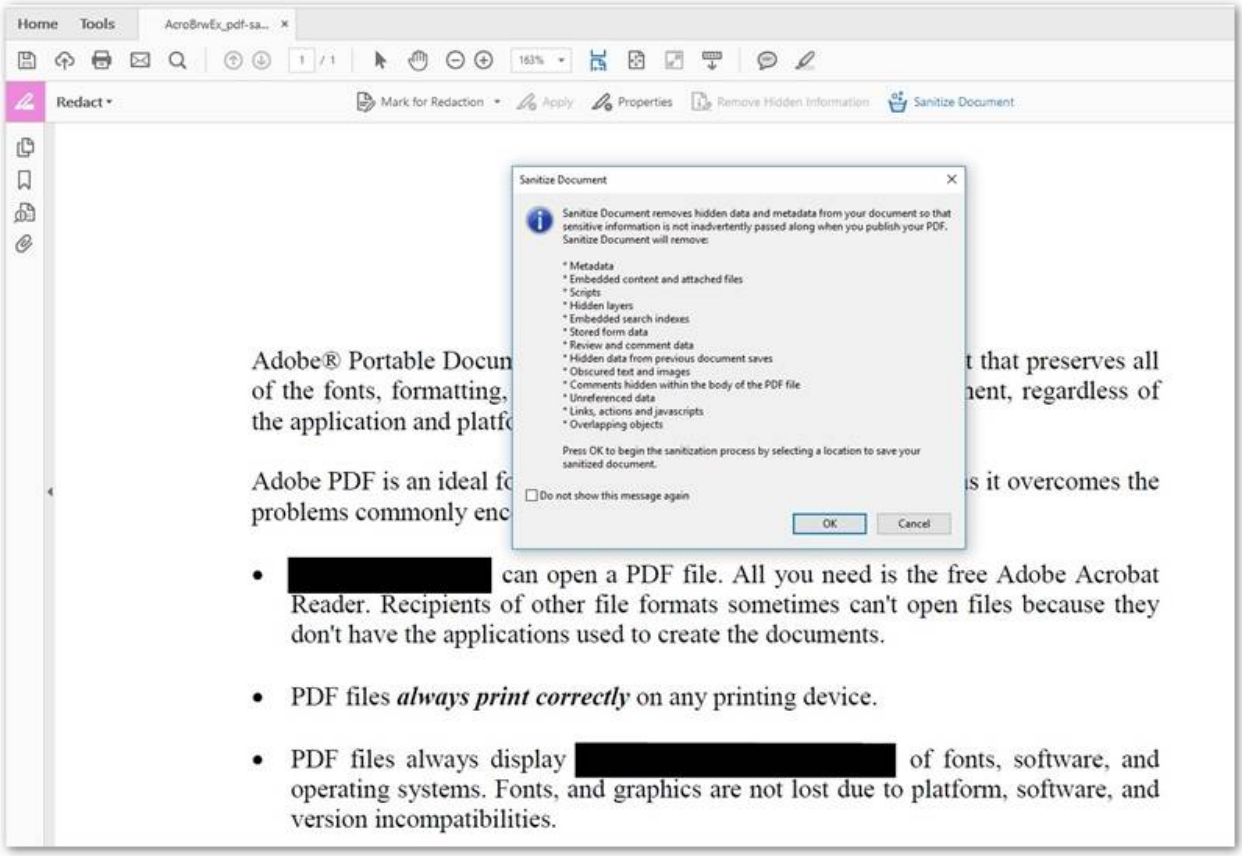


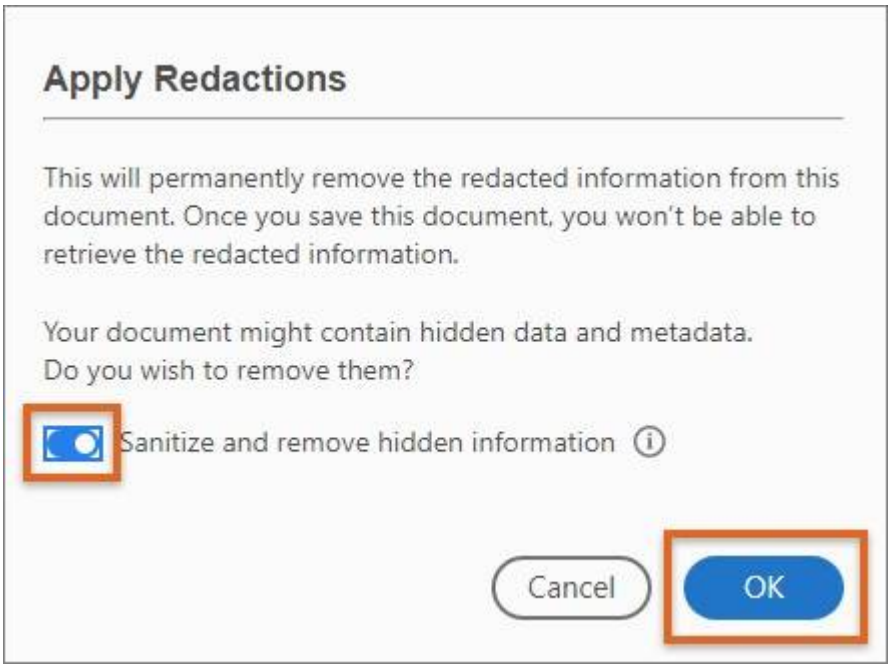




Conversion to PDF Format

Finally, presentations and documents are often converted to PDFs for sharing across the internet. Although you may not see hidden data when simply viewing a PDF through a common viewer, the PDF can contain PHI in hidden objects as well as metadata stored in tags. Adobe has a "Sanitize" function that will help you identify and redact hidden data.


Home
Tools
Redact
Mark for Redaction
Apply
Properties
Remove Hidden Information
Sanitize Document
Sanitize Document
Sanitize Document removes hidden data and metadata from your document so that sensitive information is not inadvertently passed along when you publish your PDF. Sanitize Document will remove:
* Metadata
* Embedded content and attached files
* Scripts
* Hidden layers
* Embedded search indexes
* Stored form data
* Review and comment data
* Hidden data from previous document saves
* Obscured text and images
* Comments hidden within the body of the PDF file
* Unreferenced data
* Links, actions and javascripts
* Overlapping objects
Press OK to begin the sanitization process by selecting a location to save your sanitized document.
Do not show this message again
OK
Cancel
Adobe® Portable Docun
t that preserves all
of the fonts, formatting,
nent, regardless of
the application and platf
Adobe PDF is an ideal f
s it overcomes the
problems commonly enc
[redacted] can open a PDF file. All you need is the free Adobe Acrobat Reader. Recipients of other file formats sometimes can't open files because they don't have the applications used to create the documents.
PDF files *always print correctly* on any printing device.
PDF files always display [redacted] of fonts, software, and operating systems. Fonts, and graphics are not lost due to platform, software, and version incompatibilities.


Apply Redactions
This will permanently remove the redacted information from this document. Once you save this document, you won't be able to retrieve the redacted information.
Your document might contain hidden data and metadata.
Do you wish to remove them?
Sanitize and remove hidden information
Cancel
OK

General Dos and Don'ts

Don't . . .

- Turn the font color the same as the background color
- Put an object over the PHI
- **Crop the image without deleting the cropped portion of the image**

Do . . .

- Capture images without any PHI
- If your image has PHI in the pixel data, consider a third-party image processing software (e.g. IrfanView, Adobe Photoshop or similar) to cut out the PHI and then saving just image data
- Make sure all slides have no PHI data in the cropped areas – use specific presentation software functions designed to permanently remove cropped content if applicable
- Make sure all slides have no PHI data in the notes sections or in areas beyond the displayable slide

Other Regulatory Implications

Additionally, please note that if you reside in a European Union member state or another nation, your use of patient-identifiable information even for educational purposes must comply with that nation's privacy laws and regulations.

This article outlines privacy regulations and considerations for using medical images and related data in the EU and one major U.S. state.

These articles describe important legal differences among PHI, Personally Identifiable Information, or PII, and Personal Data:

- What's the difference between PII and personal data?
- What is personally identifiable information (PII)?

If you or your practice or department remove metadata from medical images, please understand that will strip out alternative text that frequently is used to meet web site accessibility standards. Whether the Americans with Disabilities Act applies to websites and thus requires them to be accessible to individuals with disabilities, though, remains an evolving legal matter. Please consult your practice's or institution's legal counsel for specific guidance.

Modern OCR and Indexing by Search Engines

One of the challenges of publishing objects with hidden data is that it is often possible for programs that crawl the internet to find this hidden data and expose it without the author even knowing that data was distributed. Modern search engines are particularly adept at combing through publicly available files at scale, making it possible to quickly uncover a variety of data previously thought to be absent. Additionally, the ability to use Optical Character Recognition (OCR) at scale allows programs to quickly re-generate explicit PHI that was originally burned into the image pixels. Search engines can then associate ("index") the image with that explicit PHI thereby making it discoverable. As a result, these data can be made available and linked to other text-based information.

What can you do if something gets out? You can ask the search engine company to review and consider removing a link to sensitive information if they agree that this is the appropriate action. Here is an example of how this process works with Google . Similar mechanisms are available for other search engines such as Bing once the content has been removed from the site.

Appendix

What Constitutes PHI?

In Security of Electronic Medical Information and Patient Privacy: What You Need to Know and in various HHS references there are robust discussions of Personal Health Information (PHI) and the importance of protecting patient data. It is the responsibility of the individual sharing the medical case to ensure data has been properly de-identified and that any legal constraints for sharing data have been met.

There are two methods for HIPAA-compliant deidentification, Safe Harbor which identifies specific data elements that need to be removed and Expert Determination, which is used to determine the risk of re-identifying a patient based on statistical expertise. When using medical imaging cases for public consumption via publications or educational activities, the Safe Harbor method should be used. This method identifies the following data elements should not be shared along with the medical image:

- Names
- Geographic subdivisions smaller than a state

- All elements of dates (except year) related to an individual (including admission and discharge dates, birthdate, date of death, all ages over 89 years old, and elements of dates (including year) that are indicative of age)
- Telephone, cellphone, and fax numbers
- Email addresses
- IP addresses
- Social Security numbers
- Medical record numbers
- Health plan beneficiary numbers
- Device identifiers and serial numbers
- Certificate/license numbers
- Account numbers
- Vehicle identifiers and serial numbers including license plates
- Website URLs
- Full face photos and comparable images
- Biometric identifiers (including finger and voice prints)
- Any unique identifying numbers, characteristics, or codes