

REDACTION TOOL

Redacting Electronic Records

Redaction is the permanent removal of sensitive information so unauthorized users can't access it (it is best practice to redact a copy of a record, never the original). Paper records might be redacted by using opaque tape or ink to cover up text before reproducing the redacted record to ensure that the information is not present when held up to a light source. However, with electronic records, obscuring the information is not enough—it must be completely removed. As long as the record is digital, the internal bitstream of a file can still reveal or provide clues about redacted content.

Redaction Methods

Word processing applications, such as Microsoft Office, create and store digital documents in proprietary, binary-encoded formats. These formats may contain hidden information, such as change histories, embedded metadata, stored form data, or hidden comments, that allow recovery of deleted information. Redaction might be done through conversion to a format which contains displayable information only and eliminates hidden information from the original bitstream, such as a plain text file. For plain text files or other formats with a one to one relationship between the bytes and the displayable characters, simply deleting characters redacts information. Other format options that may be useful in redacting digital records include PDF, CSV, or Windows BMP, but keep in mind: redaction means unrecoverable deletion.

Example Tools

When dealing with large quantities of information, how can you accomplish redaction efficiently and accurately? Software exists that assists in automating digital redaction by using techniques to scan and locate patterns of information. For example, the tool Bulk Extractor uses regular expressions to find information like SSNs or credit card numbers. There are other tools that integrate with Adobe Acrobat, such as Redax, which can eliminate both text and graphics from a PDF. Considerations for selecting an application might include:

- + Cost
- + Applicable file types
- + Ability to integrate with existing applications
- + Full automation vs. review ability
- + Developer configuration vs. a ready-made interface
- + Level of customization for search expressions
- + Ease of testing



Pennsylvania State Archives, MG-526.1

MATRIX

	TEXT	AUDIO	VISUAL (STILL IMAGE)	VISUAL (MOVING IMAGE)	STRUCTURED DATA/METADATA
Redaction Tool Matrix	Bulk Extractor	Audacity	Redact-It	Free Video Cutter Joiner	FolderMatch
	Adobe Acrobat Pro	Adobe Premier	Redax	Adobe Premier	MRU-Blaster
	RapidRedact		Rapid Redact	Final Cut Pro	
	Redact-It			VIDIZMO	
	Redax			Caseguard	
	PDF Element			Sighthound	

eRECORDS DAY 2022

Best Practices Checklist

- + Search thoroughly for PII, PHI, Social Security numbers, accounts, and other sensitive data.
- + Always verify redaction methods.
- + Make sure you use bit-level redaction, and confidential information is not recoverable.
- + Verify that compromising metadata is wiped.
- + Retain original records as preservation files. Only perform redactions on derivatives of originals.

Redacting Non-textual Formats

Audiovisual materials may also require redaction—whether on-demand or in a large scale project. The ability to blur or black out video content or online images may be necessary to provide access. Oral histories or interviews may require editing, cutting and trimming, or redaction if the consent form notes information is private or confidential.

SOURCES

- <https://www.digipres.org/tools/by-function/>
- https://cdn.nationalarchives.gov.uk/documents/information-management/redaction_toolkit.pdf
- <https://www.dpconline.org/docs/digital-preservation-handbook2/1550-dp-handbook-technical-solutions-and-tools/file>



CoSA
Council of State Archivists

www.statearchivists.org
info@statearchivists.org