**Report: Redaction for Digital Records**

RAC – Digital Records Committee

Lucas Buresch – January 2014

**What is Redaction?**

“Redaction is the separation of disclosable from non-disclosable information by blocking out individual words, sentences or paragraphs or the removal of whole pages or sections prior to the release of the document. In the paper environment some organizations will know redaction as extracts when whole pages are removed, or deletions where only a section of text is affected.”[[1]](#footnote-1)

**Building a redaction program”**

* Redact a copy, not the original - “Always carry out redaction on a copy of the original record, whether paper or electronic, never on the record itself. This ensures that while the redacted information is permanently removed from the copy of the record (which can then be made accessible) the original text remains in the original record. Redaction should never result in the complete removal of text or information from a record.”[[2]](#footnote-2)
* Criteria for identifying information to be redacted
* Identify redacted records (metadata)
* Storage & tracking of retained extracts (un-redacted originals)

**Methods of redaction for digital records:**

Issues:

* “The simplest type of electronic record to redact is a plain text file, in which there is a one to one correspondence between bytes and displayable characters. Because of this direct Redaction toolkit for paper and electronic documents correspondence, redacting these formats is simply a matter of deleting the displayed information - once the file is saved, the deleted information cannot be recovered.”[[3]](#footnote-3)
* “However, the majority of electronic records created using office systems, such as Microsoft Office, are stored in proprietary, binary-encoded formats. Binary formats do not have this simple and direct correlation, and may contain significant information which is not displayed to the user, and the presence of which may therefore not be apparent. They may incorporate change histories, audit trails, or embedded metadata, by means of which deleted information can be recovered or simple redaction processes otherwise circumvented.”[[4]](#footnote-4)
* “It is therefore essential that any redaction technique is secured to eliminate the possibility of redacted information being recovered.”[[5]](#footnote-5)

Approaches:

* Traditional – print digital records, manually redact info (black-out), re-digitize the record (scan into PDF)
* Conversion – convert digital records to ASCII text files (eliminate any information that may be hidden in non-displayable portions of a bit stream)
* Round-trip redaction – “conversion of the record to another format, followed by conversion back to the original format, such that the conversion process removes all evidence of the redacted information”. i.e. PDF  Windows Bitmap (BMP)  PDF.

**Software Options:**

Adobe Acrobat XI Pro - <http://www.adobe.com/products/acrobatpro.html>

* “Does offer functionality for redaction, it’s not the product’s focus, so its features for redaction projects are somewhat limited”.[[6]](#footnote-6)
* Expensive; $20 per month subscription ($240 annually)
* Acrobat XI Standard does not offer redaction functionality

Appligent Redax 5.5 - <http://www.appligent.com/desktop-software/redax/>

* Plug-in for Acrobat (7 or higher)
* “Appligent Redax 4.6 has excellent features, a focus on redaction workflow, a mature product, and opportunities for integration into a more complete high-end redaction suite if the project scope grows to demand it.”[[7]](#footnote-7)
* $250 (single user)

Informative Graphics Redact-It - <http://www.redact.com/>

* Stand-alone program, does not require Acrobat
* “Unfortunately, their proprietary encrypted CSF format is not helpful for current library projects, which need full access to redaction data. They support output to PDF, but it’s not clear from their documentation if those PDFs include searchable text, or if they’ll require additional workflow to be searchable.”[[8]](#footnote-8)
* $195

RapidRedact - <http://rapidredact.com/>

* Stand-alone program, does not require Acrobat
* “Integrates into Microsoft Word to simplify workflow for redacting Word documents”[[9]](#footnote-9)
* “RapidRedact offers no pricing online for even their basic product. They sell primarily through resellers and offer only long-distance phone numbers and email for sales contact”.[[10]](#footnote-10)

**An RAC policy for redaction of digital records?**

* Does the RAC have an existing redaction policy for paper records? Is this found in the RAC’s Processing Manual?
* Should a separate policy be written into a ‘Processing Manual for Digital Records’?
* Is redaction of digital records feasible at the RAC given the move away from intensive processing towards MPLP?

1. http://www.nationalarchives.gov.uk/documents/information-management/redaction\_toolkit.pdf [↑](#footnote-ref-1)
2. Ibid. [↑](#footnote-ref-2)
3. Ibid. [↑](#footnote-ref-3)
4. Ibid. [↑](#footnote-ref-4)
5. Ibid. [↑](#footnote-ref-5)
6. <https://techknowhow.library.emory.edu/blogs/branker/2009/06/03/redaction-software-recommendation-marbl-digital-archiving> [↑](#footnote-ref-6)
7. Ibid. [↑](#footnote-ref-7)
8. Ibid. [↑](#footnote-ref-8)
9. Ibid. [↑](#footnote-ref-9)
10. Ibid. [↑](#footnote-ref-10)