Methods for the Redaction of Social Security Numbers from Electronic Land Record Documents Report

Version 1.0, October 2006

Prepared for:
The Joint Study Committee Studying Operations of the Circuit Court Clerks' Offices
Commonwealth of Virginia

By:
Virginia Information Technologies Agency
Redaction Workgroup

(This page is blank, intentionally.)

Workgroup Members

David E. Anderson	Virginia Real Estate Attorneys League
Ralph L. Axselle, Jr.	Williams Mullen, LLC
Steve Baugher	Virginia Association of Mortgage Brokers
David Bell	Arlington County Circuit Court Clerk
Lisa Carson	Compensation Board
Randy Carter	Suffolk City Circuit Court Clerk
Gary Clemens	Loudoun County Circuit Court Clerk
Ron Cornelison	AmCad, LLC
Dave Croson	Stonewall Title and Escrow
Paul DeLosh	Executive Secretary of the Supreme Court
John G. "Chip" Dicks	Virginia Court Clerks' Association
Ronald Donn	LandAmerica Financial Group, Inc.
John Frey	Fairfax County Circuit Court Clerk
Frank Hargrove, Jr.	Hanover County Circuit Court Clerk
Jonathan Haupt,	International Land Systems, Inc.
Larry Herman	AmCad, LLC
Wade Jewel	Virginia State Compensation Board
Anne Leigh Kerr	Virginia Bar Association
Daniel McBryde	Executive Secretary of the Supreme Court
Paul M. Metzger	Spotsylvania County Circuit Court Clerk
David Michie	LexisNexis ®
Faye Packet	Future Law, LLC
Eric B. Perkins	Virginia Information Technologies Agency
Craig Sanders	Logan Systems, Inc.
Visagar Shyamsundar	AmCad, LLC
Tina Sinnen	City of Virginia Beach Circuit Court Clerk
Bob Smith	Executive Secretary of the Supreme Court
Glenn Smith	Library of Virginia
Yvonne Smith	Henrico County Circuit Court Clerk
Joseph E. Spruill, III	Virginia Bankers Association
Pia Trigiani	Virginia Bar Association
Chuck Tyger.	Virginia Information Technologies Agency
Chris Whyte	Vectre Corporation
Judy Worthington	Chesterfield County Circuit Court Clerk

(This page is blank, intentionally.)

Table of Contents

(This page is blank, intentionally.)

Executive Summary

The automation of land records was authorized by the General Assembly in legislation in 1996 and continued in various legislative sessions through the 2006 General Assembly. In 2006, the General Assembly requested the Virginia Information Technologies Agency (VITA) to develop this report: **Methods for the Redaction of Social Security Numbers from Electronic Land Records**.

Redaction is the removal of non-disclosible information from a document. For the purpose of this Report "redaction" relates only to the automated removal of social security numbers (SSNs) from electronic copies of land records maintained by circuit court clerks made available via secure remote access. Original records of the circuit court clerk will not be altered.

The redaction of SSNs from land records is accomplished by employing directly a software redaction system designed for this purpose. Such a system is designed to defeat any electronic attempt to reinstall redacted information by physically removing SSNs permanently from the document.

This Report sets out principles and functionalities for software redaction systems and three recommendations for the redaction of SSNs from electronic land records.

Redaction Principles:

- Redaction should always be carried out on copies and never result in the removal of text from an original document.
- Redaction should be limited to SSNs and performed on a "day-forward" and on a "back-file" basis, as defined in the Glossary to this Report.
- Redaction should be performed by persons who are knowledgeable and trained in redaction; and the circuit court clerks may choose to engage vendors to perform these functions.
- Decisions regarding redactions need to be documented and stored. Most software redaction systems automate this process.

Functionality of Software Redaction System:

- Compatibility with the circuit court clerks' electronic recording (eRecording) and secure remote access systems.

- Employ Optical Character Recognition or other software applications which are capable of translating images of typewritten text into machine-editable text.
- Ability to accurately store and process date/time data between the nineteenth through the twenty-second centuries, with leap year calculations.
- Retain original electronic files in their native formats.
- Perform automatic data extraction by file property/attribute and document property.
- Automated redaction process workflow should include, but is not limited to, the following steps: Input (image scanning); Field Identification (locate and read SSNs); Validation (test the accuracy of data recognition); Quality Control (ensure data recognition is accurate); Redaction (permanently removes SSNs and documents the decision); Output (publish the redacted documents).
- Perform "back-file" redaction from 1935 to the present as one project and, at least annually, perform "day-forward" redaction.

Recommendations:

Recommendation 1. Perform redaction of SSNs on both a "day-forward" and "back-file" basis from electronic copies of previously recorded land records through the use of a software redaction system.

Recommendation 2. Require vendors adhere to the Guidelines for Functionality of Software Redaction Systems and the Redaction Principles set out in this Report.

Recommendation 3. Estimate the number of images in their land records database for the period January 1, 1935, to June 30, 2007, and provide that information so that it can be included as a supplement to this Report.

Recommendation 4. Plan for an estimated average redaction cost of $0.04 per image for "back-file" redaction and a similar cost for "day-forward" redaction. The estimated average redaction cost is based on information VITA received from several vendors regarding their pricing structure.

(This page is blank, intentionally.)

I. Background

The following is an overview of the background of the automation of the land records project and a brief legislative history from 1996 to the present, including the legislation authorizing VITA to develop *Methods for the Redaction of Social Security Numbers from Electronic Land Records*.

- In 1996, the General Assembly established the "technology trust fund" for the purpose of automating land records and making them available online "on a statewide basis" (§ 17.1-279 of the *Code of Virginia*). At the time, the Internet was in its formative stages and was not widely used in commercial, governmental or personal transactions. As the Internet and technology matured and the General Assembly determined, as a matter of public policy, that the online information should be secure and available only to subscribers, the automated land records project evolved into providing "secure remote access" to land records. A number of clerks began automating their land records using the technology trust fund money and, in many cases, money from the localities.

- In 2003 the General Assembly enacted HB 2426 (§ 2.2-3802.2 of the *Code of Virginia*) which provided that beginning January 1, 2004, no court clerk shall post on a court-controlled website any document that contains the following information: (i) an actual signature; (ii) a social security number; (iii) a date of birth identified with a particular person; (iv) the maiden name of a person's parent so as to be identified with a particular person; (v) any financial account number or numbers; or (vi) the name and age of any minor child. The bill also provided an exception for circuit court clerks providing secure remote access to their records if their network or system used to provide the access has been certified by the Department of Technology Planning (now the Virginia Information Technologies Agency [VITA]). VITA developed standards for secure remote access, which have become the "Secure Remote Access Standards" (SEC503-02) and has updated those standards annually as the General Assembly has made amendments to the legal framework for automation of land records and their availability by secure remote access to subscribers.

- With the passage of SB 241 in 2004, the General Assembly provided further direction for the records automation project,

increased the recordation fee from $3 to $5 and set the goal of making secure remote access to land records available on a statewide basis by July 1, 2006. In 2004 and 2005, the General Assembly began to address the concerns about the existence of social security numbers in some limited number of records contained in the circuit court clerk's paper and electronic databases. Legislation was enacted requiring the preparer of a document to be recorded to redact the social security number prior to recordation. Further, the concept of a "separate addendum" was placed in the *Code of Virginia*, specifically in divorce cases, to prevent the placement online of sensitive personal information even though it would be made available only through secure remote access to subscribers. Finally, the General Assembly narrowed the universe of documents to "land records", with the goal of limiting the scope of documents available to subscribers through secure remote access.

- In 2005 the General Assembly passed the Virginia version of the Uniform Residential Mortgage Satisfaction Act and the Uniform Electronic Recording Act, with a provision that VITA develop standards for electronic recording and that the legislation be re-enacted in the 2006 General Assembly before it would become effective. Those standards, known as the Electronic Recording (eRecording) Standards, have been largely developed and will be finalized in 2006. In the 2006 session, the General Assembly enacted both "Uniform Acts", which establish the legal framework for recordation of electronic land records in Virginia, similar to versions of the "Uniform Acts" to be adopted in all other states.

- Also, in the 2006 session, the General Assembly enacted HB 563 which, among other things, removed the "goal" that all clerks will provide secure remote access by July 1, 2006, and provided instead that clerks certify they will provide secure remote access to land records on or before July 1, 2007. The bill requested VITA to develop methods for the redaction of social security numbers from land record documents maintained in electronic form by offices of circuit court clerks and made available via remote secure access.

II. Assignment

The Virginia Information Technologies Agency (VITA) is directed by the General Assembly through HB 563 to develop methods for the redaction of social security numbers (SSNs) from electronic land record documents for Circuit Court Clerks.

Redaction is the separation of disclosible from non-disclosible information by blocking out individual words, sentences or paragraphs or by removal of whole pages prior to the release of the document. For the purpose of this report "redaction" relates only to the removal of SSNs from land records available via secure remote access.

Prior to development of software programs, the only way to redact a SSN from a land record document was to do so manually, and then rescan the document. Today, there are software redaction systems designed for this purpose and available for application to land records maintained by circuit court clerks. VITA has reviewed several of these software redaction systems as part of its work in preparation of this Report.

The software redaction systems reviewed by VITA are designed to defeat any electronic attempt to reinstall redacted information. In reviewing these software redaction systems, VITA recommends that the electronic redaction methods physically remove SSNs permanently from the documents and that word processing software functions, where redactions can easily be restored should not be used. Further, VITA recommends that setting the background behind text to black or white out should not be used, nor should software be used that attempts to make images of SSNs unreadable by reducing their size.

III. Redaction Principles

1. Redaction should never result in permanent removal of text from an original document. Redaction should always be carried out on copies of an original document. This preserves the original document as a public record, as defined in the *Code of Virginia*.

2. As directed by the General Assembly and for the purposes of this Report, redaction should be limited to SSNs in land records documents available via secure remote access and performed on a "day-forward" and on a "back-file" basis.

3. When photocopying an original document prior to redaction, reviewers should consider whether any other factors are important for the understanding of the material. For example, if color clarifies a meaning clear (for instance on a color map), a redacted color copy should be released.

4. Redaction should be performed by persons who are knowledgeable and trained in redaction of SSNs. The circuit court clerks may engage their vendors to perform these functions so that the day-to-day responsibilities of the circuit court clerks to the circuit courts and the citizens can continue to be provided without interruption.

5. No method, including the use of a software redaction system, should be regarded as secure unless it has been tested to ensure that it provides for the irrecoverable erasure of redacted information to a level of quality assurance set out in the guidelines.

6. Decisions regarding redactions need to be documented and stored. Most software redaction systems automate this process. The identifying information should include, at the minimum,

    a. the name of the person who was responsible for the software redaction system that performed the redaction of SSNs;
    b. an identifying document reference and a book and page number or instrument number for each redaction indicating where the redaction(s) occurred;
    c. the date and time when the redaction was done; and
    d. the identifier which enables easy identification and retrieval of the original document and chosen, consistent format.

7. Sample forms for recording decisions can be found at Appendix A.

IV. Guidelines for Functionality of Software Redaction System

The software redaction system used by the circuit court clerks to comply with the *Code of Virginia* should meet the following guidelines

1. Compatibility with the circuit court clerks' electronic recording (eRecording) and secure remote access systems.

2. Optical Character Recognition (OCR) or other software applications capable of translating images of typewritten text into machine-editable text.

3. Ability to accurately store and process date/time data accurately in four (4) digit fields (including, but not limited to, calculating, comparing, interfacing and sequencing) from, into, and between the nineteenth through the twenty-second centuries, and leap year calculations.

4. Ability to retain original electronic files in their native formats.

5. Automatic data extraction by file property/attribute (e.g., filenames, file size, creation date, etc.) and document property (e.g., title, author, etc.).

6. Automated redaction process workflow that includes, but is not limited to, the following:

   a. Input: Imaging scanning technologies, such as: WIA (Windows Image Acquisition); TWAIN; ISIS (Image and Scanner Interface Specification); SANE (Scanner Access Now Easy), etc., that provide access to technical applications such as flatbed scanners, handheld scanners, video- and still-cameras, frame grabbers, etc.

   b. Field Identification: Automated field identification rules locate and read SSNs in nine (9) digit fields, with or without hyphens, printed on either one line or split over a line brake.

   c. Validation: Test the accuracy of data recognition through manual and automated validation. Use manual validation processing by an operator, if so designated, or an automatic process in the background using operator preset confidence levels.

   d. Quality Control: Apply automatic quality control (QC) to ensure data recognition is accurate (e.g., scoring below an operator designated level would flag the image for additional manual

validation processing and send it to a selected queue). The QC process should take into consideration:

    i. the condition of the document and the quality of the image to be redacted, including the sharpness of the image as determined by dots per inch (DPI); and

    ii. the ability to achieve a level of accuracy of at least 95%, however the level of accuracy may be negatively affected by the quality of the image from which the SSN is being redacted.

e. <u>Redaction</u>: Use secure effective redaction software that alters the copy (**not** the original) by removing the original text containing the SSN and overlaying the space previously occupied by the SSN with a graphical element such as a black rectangle. The original text remains in the original file and **not** in the copy, therefore, it can not be uncovered by simply deleting the overlaying graphics; and, the redaction software documents and stores the redaction decision.

f. <u>Output</u>: Publish the redacted documents as a TIFF, <u>Portable Document File</u> (PDF), secure <u>Content Sealed Format</u> (CSF) or other technological application.

g. A redaction software system that meets these guidelines should be run at least annually following the completion of the "back-file" redaction of SSNs from all land records from January 1, 1935. This will serve as a back-up to the existing provisions of the *Code of Virginia* that require SSNs be removed by the attorney or party preparing a document for recordation.

V. Recommendations

Recommendation 1. The Workgroup recommends Circuit Court Clerks perform redaction of SSNs from land records available via secure remote access on a:

a. "day-forward" basis from electronic **copies** of land records at least once every 12 calendar months through the use of a software redaction system; and

b. "back-file" basis from electronic **copies** of previously recorded land records through the use of a software redaction system.

*Recommendation 2.*The Workgroup recommends Circuit Court Clerks require in the vendor contract that their vendors adhere to the Guidelines for Functionality of Software Redaction Systems and the Redaction Principles set out in this Report.

Recommendation 3. The Workgroup recommends Circuit Court Clerks request that their vendors estimate the number of images in their land records database for the period January 1, 1935, to June 30, 2007, and provide that information so that it can be included as a supplement to this Report.

 Recommendation 4. The Workgroup recommends Circuit Court Clerks plan for an estimated average redaction cost of $0.04 per image for "back-file" redaction and a similar cost for "day-forward" redaction. The estimated average redaction cost is based on information VITA received from three vendors regarding their pricing structure for the software redaction system and the employee responsibilities attendant to running the software redaction system.

Appendix A: Sample Form for Recording Redaction Decisions

Sample form for recording redaction decisions

Redaction Date	Document Reference (Book & Page No.)	Reviewer's comments	Any other comment

Example of form to record decisions and pass on information to document editors carrying out redaction

Editing requirements: Document Name: _____
                      Document Section: _____

Book No.	Page No.	Paragraph No. (If available)	From (inclusive)	To (inclusive)

Appendix B: References

1. Broward County, Florida, *Automated Document Redaction Agreement* (RLI #20050124-0-CR-01), April 2006, Fort Lauderdale, FL, Contact: PKETZ@broward.org

2. Carlson, Alan, Wade Steketee, Martha, *Developing CCJ/COSCA Guidelines for Public Access to Court Records: A National Project to Assist State Courts,* October 18, 2002, National Center for State Courts and the Justice Management Institute http://www.courtaccess.org/modelpolicy/18Oct2002FinalReport.pdf

3. *General exemptions from inspection or copying of public records*(Title X, Chapter 119, § 119.071), Florida Statutes, 2006 Session, http://www.leg.state.fl.us/Statutes/index.cfm?App_mode=Display_Statute&Search_String=&URL=Ch0119/SEC071.HTM&Title=-%3e2006-%3eCh0119-%3eSection%20071#0119.071

4. *How to Minimize Metadata in Office Documents*. Microsoft.Com, January 28, 2005, http://support.microsoft.com/default.aspx?scid=kb;EN-US;Q223396

5. *Identity Theft Protection Act*, North Carolina General Assembly, 2005 Session, http://www.ncga.state.nc.us/Sessions/2005/Bills/Senate/HTML/S1048v6.html

6. Jackson, Joab, *Documents can reveal supposedly deleted data, but fortunately there are ways to make sure that secrets stay secret.* Government Computer News February 2o, 2006, http://www.gcn.com/print/25_4/38253-1.html

7. Kaplan, Ari, *A New Generation of Redacting Tools*, The National Law Journal, November 14, 2002, http://www.law.com/jsp/article.jsp?id=1036630382605

8. *Microsoft Word 2003 Redaction Add-in Tool*. Microsoft Corp., http://office.microsoft.com/en-us/assistance/HA012193991033.aspx

9. Mingis, Ken, Rosencrance, Linda and Weiss, Todd R. *Update: Fla. residents' data exposure a statewide issue*, (2006, April 11) Computerworld http://www.computerworld.com/securitytopics/security/privacy/story/0,10801,110389,00.html

10. NewsFactor, *State Sites Removing Personal Data* (2006) http://www.newsfactor.com/news/State-Sites-Removing-Personal-Data/story.xhtml?story_id=12000B101PIO)

11. Office 2003/XP Redaction Add-in (free download) http://www.microsoft.com/downloads/details.aspx?FamilyId=144E54ED-D43E-42CA-BC7B-5446D34E5360&displaylang=en

12. Orange County Comptroller, *Request for Proposal for Redaction pf Confidential Information* (RFP #2005-01-OR), (2004) Orlando, FL, Contact: carol.foglesong@occompt.com

13. P*rivacy and Public Land Records: Making Practical Policy*. (2006, January 4). A White Paper by the Property Records Industry Association (PRIA). http://www.pria.us/Papers/PRIAWhitePaperFinal010406.pdf

14. Rainbow, Justin. *Redaction Toolkit: Guidelines for the Editing of Exempt Information from Documents Prior to Release,* (2004, November) The National Achieves. The National Archives, Kew, Richmond, Surrey, TW9 4DU, UK http://www.nationalarchives.gov.uk/recordsmanagement/advice/pdf/redactiontoolkit.pdf

15. *Redaction Instructions, Form & Video*, Clerk of the Circuit Court, Pinellas County, FL, http://www.pinellasclerk.org/aspInclude2/ASPInclude.asp?pageName=redaction.htm

16. *Redacting of Confidential Information in Electronic Documents*, (2006) Adobe Systems Incorporated San Jose CA http://www.adobe.com

17. *Redacting with Confidence: How to Safely Publish Sanitized Reports Converted From Word to PDF*. National Security Agency. (2005, December 13) http://www.fas.org/sgp/othergov/dod/nsa-redact.pdf

18. *States Removing Personal Data from Official Web Sites*. (2006, April 24) http://www.newsfactor.com/news/State-Sites-Removing-Personal-Data/story.xhtml?story_id=12000B101PIO

19. U.S. District Court web page on Redaction: https://ecf.cand.uscourts.gov/cand/faq/tips/redacting.htm

20. U.S. Navy, Chief of Naval Operations, *FIOA Policy Re: Redaction Techniques* (Ser DNS-36/5U826221), May 5, 2005, www.foia.navy.mil

21. Wait, Patience, *NSA urges use of better redaction methods*, Government Computer News February 20, 2006, http://www.gcn.com/print/25_4/38286-1.html

22. Wait, Patience, *NSA urges use of better redaction methods*, Government Computer News February 20, 2006, http://www.gcn.com/online/vol1_no1/38079-1.html

Contact for further information regarding this report:

Eric B. Perkins
Virginia Information Technologies Agency
IT Investment & Enterprise Solutions Directorate

411 E. Franklin Street, Suite 500
Richmond, VA 23219
Telephone: 804.786.0516
Fax: 804.371.2277

Email: eric.perkins@vita.virginia.gov

Appendix C: Glossary

Authentication – The term "authentication" refers to the process of verifying the identity of a user.

Authorization – The term "authorization" refers to the process of establishing and enforcing a user's rights and privileges to access specified resources.

Back-file – The term "back-file" redaction means the redaction of social security numbers from land records that exist in the database of a circuit court clerk prior to the date of completion of redaction.

Best Practice - The term "best practice" means a guideline or specification that is advisory in nature and whose compliance is strongly recommended; however, it is not binding on Agencies.

Confidence Level - A measure of how reliable a result is, expressed as a percentage that indicates the probability of the result being correct.

Data – The term "data" includes, but is not limited to, data in a database, information about an OS, operational policies and procedures, system design, organization policies and procedures, system status, and personnel schedules.

Day-forward – The term "day-forward" redaction means the redaction of social security numbers from land records that exist in the database of a circuit court clerk since the last redaction. As provided in the "Guidelines", "day-forward" redaction should occur at least once every 12 calendar months, measured from the most recent redaction.

Encryption – The term encryption refers to the process of converting computer data and messages to something incomprehensible by means of a key, so that the data can be reconverted only by an authorized recipient holding the matching key.

Information – The term "information" means any communication or representation of knowledge such as facts, data or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative or audiovisual forms.

Information Resources – The term "information resources" includes government information, information technology and associated personnel.

Information Systems - The term "information systems" means a discrete set of information resources organized for the collection, processing, maintenance, transmission and dissemination of information in accordance with defined procedures.

Information Technology –The term "information technology" means hardware and software operated by an organization to support the flow or processing of information in support of business activities, regardless of the technology involved, whether computers, telecommunications, or other. In the Commonwealth of Virginia, Information Technology means telecommunications, automated data processing, databases, the Internet, management information systems, and related information, equipment, goods, and services.

Quality Control – The term "quality control" means the assessment of product compliance with stated requirements. Quality control should be independent from production.

Redaction- The term "redaction" refers to the software redaction system that redacts social security numbers from land records maintained by a circuit court clerk.

TIFF – (Tag Image File Format) is a common format for exchanging raster graphics (bitmap) images between application programs, including those used for scanner images.

User – The term "user" means an individual or group who has access to an information system or its data.

Appendix D: Model Redaction Workflow

Model Redaction Workflow

Input

Software Redaction System Functionality

Output

Input:
- TIFF File Scanned Copy
- Paper Document
- Microsoft® Word
- Microsoft® Excel
- Adobe® PDF
- MPG, JPG, etc.

Software Redaction System Functionality:

To Field Identification

- Field Identification
- Validation
  - Manual or
  - Automated
- Quality Control
- Redaction
- Release

Fields Found - To Validation

To Quality Control

Passed - Redaction Needed

Failed - Return to Validation

To Release

No Fields Found - Redaction NOT Needed

Output:

To Output

- TIFF FIle
- Database
- Secure Remote Access System for Circuit Court Clerk's Electronic Records