STATEWIDE INFORMATION TECHNOLOGY WHITE PAPER

White paper: **Redaction Software**

Effective Date: July 15, 2006

Replaces and Supercedes: Not applicable.

I. Purpose

The purpose of this white paper is to provide a brief overview of the need for and the use of redaction software in instances where the state and local government makes records available to the general public.

This white paper is provided for informational purposes only. Comments, questions and suggestion are welcome and may be directed to the contact information provided below.

II. Definition(s)

Refer to the Statewide IT Policies and Standards Glossary for a complete list of definitions.

III. Closing

For questions or comments on this white paper, e-mail ITpolicy@mt.gov, or, contact the Information Technology Services Division at:

Chief Information Officer
PO Box 200113
Helena, MT 59620-0113
(406) 444-2700
FAX: (406) 444-2701

The technical contact for this white paper is:

Policy and Planning Services Bureau
910 Helena Avenue
Helena, Montana 59620-0116
(406) 444-5476
FAX: (406) 444-4644

IV. Cross-Reference Guide

- http://www.osha.gov/as/opa/foia/foi-act.html
- http://www.usdoj.gov/04foia/foiastat.htm
- http://www.usdoj.gov/oip/foia_updates/Vol_XVII_4/page2.htm
- http://www.whitehouse.gov/news/releases/2005/12/20051214-4.html
- http://www.computerworld.com/securitytopics/security/privacy/story/0,10801,110486,00.html
- http://www.hartic.com/pr_view.php?prid=35
- http://practice.findlaw.com/cyberlaw-041806.html
- https://ecf.cand.uscourts.gov/cand/docs/go53.htm
- http://www.gcn.com/print/16_31/32148-1.html
- http://www.gcn.com/print/16_34/31925-1.html
- http://news.thomasnet.com/fullstory/484039
- http://www.rapidredact.com/
- http://www.govtech.net/magazine/story.php?id=98593

V. Administrative Use

History Log	
Approved Date:	Not applicable
Effective Date:	July 15, 2006
Change and Review Contact:	ITPolicy@mt.gov
Review:	Event Review: Any event affecting this white paper may initiate a review. Such events may include a change in statute, key staff changes or a request for review or change.
Scheduled Review Date:	Not applicable
Last Review/Revision:	
Changes:	

OFFICE OF THE CHIEF INFORMATION OFFICER

WHITE PAPER

Redaction Software

July 15, 2006

STATE OF MONTANA
Department of Administration
INFORMATION TECHNOLOGY SERVICES DIVISION

Brian Schweitzer
Governor

CREDITS

Primary Authors:

Jamee Fields, Policy and Planning Services Bureau

Warren Dupuis, Policy and Planning Services Bureau

Other Contributors:

TABLE OF CONTENTS

Executive Summary 1
Technical Description 1
Business Effect 2
Relevant Federal And State Laws And Proposed Legislation 3
Major Programs, Projects, Pilots Within Montana And Across The Country 3
Agency Plans Or Links To State IT Strategic Plan 4
Near Term effect On The State 4
Recommendations On Follow-Up Research 4

Redaction Software

EXECUTIVE SUMMARY

The purpose of this paper is to provide a brief overview of the need for and the use of redaction software in instances where the state and local governments make records available to the general public. The personal data of millions of U.S. residents could be exposed by the public posting of official documents, and state and local governments (State and County) are increasingly looking for ways to automate the process of cleaning up data being put online.

The issue of removing sensitive information including Social Security Numbers, bank account information, driver's license data and personally identifying details from public documents is gaining attention in light of concerns from privacy advocates. They have argued that the number of public documents being posted online with sensitive data included could open the door for a new wave of identity theft and fraud. To meet that concern, government officials across the nation are turning increasingly to software to remove that data.[1]

There are instances where law suits have been filed for injuries resulting from their failure to remove sensitive information prior to making it available to the general public.[2] In March of 2006, an Ohio man filed a class action suit against the Ohio Secretary of State for posting his and other residents' Social Security Numbers in records on a publicly searchable web site.[3] Montana law does not currently address the issue of redaction, though it is being discussed by the Economic Affairs Interim Committee.

TECHNICAL DESCRIPTION

Redaction is the careful editing of a document to remove confidential information. The document can be in paper or electronic format (Acrobat PDF, .doc, etc.). Sensitive government documents, appraisals, legal documents, criticism, and insurance contracts are often redacted before being made available to the public. In the context of United States government agency documents, redaction refers more specifically to the process of removing classified information from a document during declassification, prior to its publication. [4]

Redaction can be done manually or automatically using redaction software. Some software automatically indexes and redacts images using algorithms that look for targeted numbers or words or seeking out related words in context -- adjacent words like “account number” or “Social Security number.” Once keywords are found, the software automatically redacts the information. The software can also remove personal information by indicating a certain area on a scanned form for automatic redaction, as long as the forms have a standard layout with information in fixed locations.

BUSINESS EFFECT

The business effect on the state is dependent on which situation you are addressing: the redaction of documents prior to being filed electronically; or the redaction of documents that have already been filed and are currently available to the public, or could be requested under the Freedom of Information Act.

The first effect is a requirement to establish the policies and procedures for posting electronic information available to the public, to include the provisions for redacting personal and sensitive information. The oversight responsibilities must be addressed and the manpower requirements identified.

For documents that have yet to be filed electronically, there are a variety of solutions available. These include free, but labor intensive, cut and paste solutions, Microsoft Word plug-ins, as well as several relatively inexpensive (depending on the number required) software solutions. It is important to note that there appears to be no single solution for Word, WordPerfect, and Adobe PDF documents.

For documents that have already been filed electronically, the solution is more expensive and time consuming. The software looks for candidates for redaction from among millions of document images. Several thousand pages are culled and analyzed individually by a person who can verify that the information should be redacted. As the pool of documents is reviewed, the software automatically adjusts to redact the remaining records based on the choices made manually. The typical review process can take two to three months and software costs typically range from $200,000 to $300,000 depending on the size of the project.

For example: the Orange County, Florida Comptroller contracted with Austin, Texas-based Hart InterCivic to provide redaction services to the County via the company’s new Anthem™ Redaction Services program. The contract is valued at $587,500. Orange County with its county seat of Orlando has a population of over one million people.[5] The number of documents containing sensitive information may be lower than people assume. Orange County, for example, is in the midst of inspecting about 30 million pages dating back to 1970 for Social

Security Numbers, bank account numbers, credit card numbers and debit card numbers. So far, seven million pages covering 2.2 million documents recorded between June 1, 2002, and April 30, 2005, have been inspected: Out of those pages, 119,000, or 1.63%, have information that needed to be redacted.[6]

RELEVANT FEDERAL AND STATE LAWS AND PROPOSED LEGISLATION

At the federal level, the Freedom of Information Act (FOIA) mandates public access to many documents as long as sensitive content is not disclosed. According to US Court General Order 53, pursuant to electronic court filings, parties must refrain from including, or must redact where inclusion is necessary, all Social Security Numbers, names of minor children, dates of birth and financial account numbers.

The Electronic Freedom of Information Act of 1996 (as amended by public law), requires that agencies of the federal government put their frequently requested documents into an electronic reading room after removing any personal or sensitive information as outlined in subsection (b) of the act.

The Egovernment Act of 2002 codifies the process by which data will be made available between governmental agencies and to the general public. The act contains provisions for the redaction of certain categories of information in order to protect privacy and security concerns.

The State of Montana Electronic Government Services Act does not specify any specific provisions for protecting privacy in regards to egovernment.

The Montana Code Annotated addresses privacy rights as they pertain to medical and insurance records but it does not specifically address the issue of other records that may be made available.

The Administrative Rules of Montana do not specificity address the redaction of personal information.

MAJOR PROGRAMS, PROJECTS, PILOTS WITHIN MONTANA AND ACROSS THE COUNTRY

The Legislative Economic Affairs Committee has established a working group on identity theft. The group requested data from state agencies in 2006 regarding Social Security Numbers used in state and local government pursuant to the Privacy Act of 1974. The committee has also discussed legislation to provide

penalty for state, and local governments that do not comply with Privacy Act of 1974. A member of that working group is reporting to the Information Technology Board.

The Montana Supreme Court is currently looking into IntelliDact redaction software from Computing System Innovations (CSI) to handle the 9,000 (est.) cases that are currently filed at the Supreme Court. Each case contains 80 to 100 pages in various formats. CSI says that it will cost two-to-four cents per page to do the redaction. That puts the cost of redacting current Supreme Court documents at $18,000 to $36,000.

Montana Department of Corrections is currently looking into Microsoft add-on redaction software to redact Word documents prior to posting. The Department of Corrections is also interested in seeing what the Montana Supreme Court decides to use for image documents.

In Florida, counties are required by statute to have all online public records redacted for sensitive personal information by January 1, 2007. Other states are looking at similar record-keeping issues. The SJR 38 white paper on Social Security Numbers outlines some specific actions that have been taken by other states.

AGENCY PLANS OR LINKS TO STATE IT STRATEGIC PLAN

Many agencies IT plans reference egovernment and e-filing in their goals and/or initiatives. Agencies such as the Office of Public Instruction and the Office of the Commissioner of Higher Education have initiatives for data warehouses and the expansion of document imaging. These all could have potential redaction issues.

NEAR TERM EFFECT ON THE STATE

The near term effect is that this may become a topic of discussion during the upcoming legislative session where the need for such legislation will be discussed. It is anticipated that this discussion will include an examination of the State's current laws and policies regarding the protection of personal information.

RECOMMENDATIONS ON FOLLOW-UP RESEARCH

It is recommended that any interim policy regarding redaction be reviewed by Department of Administration legal counsel to ensure that it complies with the intent of the Privacy Act.

It is recommended that further research be done to determine the enterprise approach that is being taken by other states.

---

[1] Todd Weiss, "Data exposure: Using software to redact personal data from public documents," *Computerworld,* April 13, 2006
[2] Ari Kaplan, "A New Generation of Redacting Tools," *The National Law Journal,* November 14, 2002
[3] Anita Ramasastry, "Can States Legally put Residents' Social Security Numbers and Other Identifying Data Online?" *Modern Practice,* April 18, 2006
[4] Amit Agarwal, "Redact Adobe PDF & Microsoft Word documents," *http://Labno.blogspot.com,* posted December 25, 2005
[5] Hart InterCivic, "Hart InterCivic to Implement Groundbreaking Redaction Project in Orange County, FL to Protect Data in Public Records," *Press Release*, May 27,2005
[6] Jaikumar Vijayan, "Data exposure: Counties across the U.S. posting sensitive info online," *Compuworld,* April 12, 2006