



Automated Redaction Proof of Concept Report

Tom Clarke
Di Graski
Jannet Lewis
Barb Holmes

September 2017

This report was developed under grant number SJI-16-P131 from the State Justice Institute. The points of view expressed are those of the authors and do not necessarily reflect the official position or policies of the State Justice Institute.

Executive Summary

In July 2017, the National Center for State Courts published “Best Practices for Court Privacy Policy Formulation,” describing the State Justice Institute-funded project to update model public access and privacy guidelines published by the Conference of State Court Administrators in 2002. “[T]his report deliberately asserts the view that policies and redaction capabilities should be considered simultaneously.”¹

Quantifying the accuracy of automated redaction capabilities and publishing those results to the court community are the goals of the Proof of Concept. NCSC is grateful for the cooperation of private solution providers who currently offer automated data-identification tools in their commercial, off-the-shelf products, as well as state and local courts who shared court case records for the POC.

The purposes of this POC Report are to provide some background about the principles underlying the Revised Model Policy for Electronic Public Access to Court Case Records, a description of the categories of protected information included within the scope of the POC, the methodology NCSC and the participating vendors followed during the POC, and, most importantly, the impressive results achieved in automated redaction of the data targets.

¹ Clarke, Thomas M., Ph.D., Jannet Lewis, and Di Graski, [Best Practices for Court Privacy Policy Formulation](#), National Center for State Courts, Web. July 2017.

Background

Section 1.0 of the Revised Model Policy for Electronic Public Access to Court Case Records states the fundamental purposes of the policy:

- a. Maximize accessibility of court case records.
- b. Protect users of the court from harm.
- c. Make effective use of court resources.

The policy goes on to explain:

Accessibility is maximized for several reasons: to enhance public trust and confidence, to be accountable, to be transparent, to improve customer service, and to reveal common law. Protection from harm includes individuals, business organizations, government agencies, and the public at large. When balancing openness against potential harm, courts should make the rationales for their decisions explicit. Remote public access is part of a much larger strategy to provide court services online to improve access and convenience and to reduce cost. Cost and efficiency considerations refer to both user costs and court operational costs.²

Automated redaction capabilities hold the promise of furthering all three goals:

- a. Maximizing accessibility of court case records: with proven redaction capabilities, policy makers could confidently expand the case types and document types that are available to the public.
- b. Protecting users of the court from harm: especially for courts that rely upon filers to redact sensitive information from their documents, several high-profile audits have revealed that confidential information like Social Security Numbers are available on public access sites at alarming rates.
- c. Making effective use of court resources: especially for courts that rely upon their clerk staff to review and redact every case filing, leveraging technology could free precious court resources to tackle more complex and higher value tasks.

In pursuit of all three overarching goals, NCSC undertook an auto-redaction Proof of Concept in August 2017. Two vendors agreed to participate, CSI and Extract.

² *Id.* at p. 10.

Proof of Concept Process

Redaction Targets

NCSC selected a list of redaction targets based upon three factors:

- (1) How typical the redaction target is in state court rules defining confidential information that must be removed from public access.

As revealed in the Fall 2016 Council for Court Excellence survey of state courts' public access policies, there is little consistency in the types of information that are considered confidential. However, many courts require redaction of personal identifiers, juvenile and crime victim names, and financial information.

- (2) The potential harm that could be caused if a redaction target were exposed to the public.

For example, disclosure of a Social Security Number poses a significant risk of identity theft and financial harm.

- (3) NCSC's goal of exercising the data-identification software on both structured and unstructured data elements.

A Social Security Number is a classic structured data element because its content and syntax follow a clear pattern: three numerals, followed by a hyphen, two numerals, another hyphen, and four numerals. Financial account numbers like bank accounts, loan numbers, investment account numbers, and sales tax account numbers are unstructured, because financial institutions do not use a single pattern for their content and syntax.

Target Description	Structured or Unstructured Data?
Telephone Number of Party, Victim, or Minor	Structured
E-mail Address of Party, Victim, or Minor	Structured
Taxpayer Identification Number (Social Security Number or Employer Identification Number)	Structured
Credit Card Number	Structured
Street Address of Party, Victim, or Minor	Unstructured
Date of Birth	Unstructured
Name of Victim or Minor	Unstructured
Financial Account Number	Unstructured
Driver's License, State Identification Number, or Vehicle License Plate/Tag	Unstructured

Description of Court Case Records

In preparation for the POC, NCSC assembled court case records from a variety of jurisdictions in Colorado, Florida, Ohio, Pennsylvania, and Texas. The case types were diverse: small claims, civil, family, and adult criminal. Some of the case records were gathered from court document management systems; other case records were mocked-up by NCSC staff.

Both structured and unstructured documents were included. Structured documents are forms designed to elicit specific information from a filer by labeling check boxes, blocks, or blanks with descriptions of the data to be included. Typical examples include financial affidavits in dissolution of marriage cases or requests for court-appointed legal counsel or *in forma pauperis* waivers of court fees. Many courts have adopted standardized forms for case types with a high occurrence of self-represented parties, such as small claims, family, and landlord-tenant. Structured documents also include court-generated documents, typically in high-volume case types, that label data elements and present them in a predictable order. Typical examples include arrest warrants that clearly and predictably list an adult criminal defendant's personal identifiers, contact information, and physical descriptors.

Unstructured documents include court case records containing primarily narrative text, such as motions and briefs.

Several of the POC court case records were handwritten documents, typically structured court forms that were completed by a *pro se* party using an ink pen.

Learning Set

Once the court case records were assembled, NCSC analyzed each document to determine its type and whether it contained redaction targets. Next, NCSC assigned a majority of the records to the Learning Set, the group of documents that the participating vendors would use to teach their software to identify redaction targets accurately. The Learning Set comprised 213 documents and included several instances of each document type and each redaction target.

During the Learning phase of the POC, the participating vendors asked clarifying questions, such as, "Are legal (metes and bounds) descriptions of real property included in the scope of 'Party Address'?" For purposes of the POC, NCSC staff provided clear "Yes, No" answers to such questions but with the understanding that each jurisdiction will need to establish policies based upon their state and local law, which will surely vary.

Test Set

Following the Learning phase, NCSC released the Test Set of 26 documents to the participating vendors. The 26 test documents contained 160 redaction targets. The vendors ran their data-identification software on the Test Set and then submitted the redacted versions. NCSC staff reviewed and scored the redacted versions, and the compiled results are reported below.

Results

High-level Accuracy Rates

The overall accuracy rate of the data-identification software was 97.81%. The accuracy rate for typed documents, 98.33%, was slightly higher than the accuracy rate for handwritten documents, 96.25%. Notably, the occurrence of false positives – data redacted erroneously – was very rare.

Accuracy Rates by Redaction Target

For structured redaction targets, the performance of the data-identification software was flawless:

Taxpayer ID (SSN, EIN):	100%
Party, Victim, or Minor’s E-mail Address:	100%
Party, Victim, or Minor’s Telephone Number:	100%
Credit Card Number:	100%

Note that these results span both structured and unstructured documents, and both typed and handwritten documents.

The data-identification software also performed quite well on most categories of unstructured redaction targets:

Driver’s License or Vehicle License Plate/Tag:	100%
Financial Account Number:	100%
Name of Victim or Minor:	98.21%
Date of Birth:	95%
Street Address of Party, Victim, or Minor:	94.87%

Matrix of Accuracy by Data Types and Document Types

NCSC’s hypothesis was that the most challenging redaction targets for data-identification software would be unstructured data types in unstructured documents.

	Structured Data	Accuracy	Unstructured Data	Accuracy
Structured Documents (n=14)	Telephone Number (n=15)	100%	Street Address (n=29)	98.28%
	E-Mail Address (n=8)	100%	Date of Birth (n=15)	100%
	Taxpayer ID (n=6)	100%	Name (Victim, Minor) (n=22)	100%
	Credit Card Number (n=5)	100%	Financial Account Number (n=15)	100%
			Driver's License or Vehicle License (n=10)	100%
Unstructured Documents (n=11)	Telephone Number (n=3)	100%	Street Address (n=10)	85%
	E-Mail Address (n=1)	100%	Date of Birth (n=5)	80%
	Taxpayer ID (n=4)	100%	Name (Victim, Minor) (n=6)	91.67%
	Credit Card Number (n=1)	100%	Financial Account Number (n=3)	100%
			Driver's License or Vehicle License (n=2)	100%