2019 Policy Paper

Court Data: Open, With Care

COSCA
Conference of State Court Administrators

COSCA Policy and Liaison Committee

Arthur W. Pepin, *Chair*
Director, New Mexico Administrative Office of the Courts

Sandra A. Vujnovich, *Co-Chair*
Judicial Administrator, Supreme Court of Louisiana

Sigfrido Steidel-Figueroa
Administrative Director of the Courts, Puerto Rico Supreme Court

Sally Holewa
State Court Administrator, Supreme Court of North Dakota

Milton L. Mack, Jr.
State Court Administrator, Michigan State Court Administrative Office

Mary T. Noonan
State Court Administrator, Utah Administrative Office of the Courts

David W. Slayton
Administrative Director, Texas Office of Court Administration

Sara B. Thomas
Administrative Director of the Courts, Idaho Supreme Court

COSCA Policy and Liaison Committee Staff

Michael Buenger
National Center for State Courts

Phillip Bye
National Center for State Courts

Melinda Kantor
National Center for State Courts

2019 Policy Paper
Court Data: Open, With Care
Conference of State Court Administrators

State courts have the authority and responsibility to establish policies to control access to and manage release of court case data. In discharging that responsibility, state courts need to consider limits on access imposed by the need to also protect individual privacy, consider costs, concerns for data integrity, and the limits of technology.

It is now well-settled that, with limited exceptions, documents making up court records are public records. For many years members of the press, the public, and the bar have had and continue to have access to court records by getting copies of paper documents at the courthouse. Today many states provide online public access to individual electronic case level information and often to electronic court records. This paper addresses court case *data*, not just court *records* or individual case level information. The public, attorneys, members of the media, researchers, and for-profit data aggregators seek access beyond records to court case data in formats that allow the data to be analyzed as fits the interests of these parties. Courts already expend significant resources responding to requests for court case data and creating customized reports of such data.

The Conference of State Court Administrators (COSCA) endorses providing access to court case data with recognition of the need for resources that make it possible to respect important exceptions and with careful attention to the potential for misuse of court case data. The policy endorsed by COSCA is to make court case data open and accessible to the maximum practical degree when balanced with legal restrictions, protection of privacy interests, data security, and within resource constraints.

What is Court Case Data?

State courts create and maintain tremendous amounts of data. Some of this data is related to the administration of the courts, while much of the data is related to individual cases. For purposes of this policy paper, COSCA adopts the following definition of court case data:

> "Any information that is collected, received, or maintained by a court or clerk of court connected to a judicial proceeding that are generally maintained in a case management system. The data may contain both public and confidential information."

Examples of court case data include, but are not limited to, data fields such as case numbers, party demographics, attorney names, judges assigned to the case, filing dates, disposition dates, types of disposition, judgment or sentence orders, risk assessment information, and post-judgment proceedings.

Electronic access to court *case records* allows one to acquire and read documents online. Access to court *case data* includes allowing one to view, obtain, sort, or tally the information that the court or clerk maintains however the user chooses. For example, with electronic access to court case data one may be able to extract all civil cases with a judgment in excess of $50,000, or all criminal cases resulting in a hung jury, or how many orders for free process were filed and granted or denied in a given time period, or in which cases within a set time period did a specific expert witness testify – all searches that reference court case data available in many case management systems. With unrestricted access to court case data, one might query the data to determine how often a specific judge granted motions to suppress in comparison to other judges in the court, or which judge in the district

has the longest average time to disposition, or which judge grants custody to fathers most often when child custody is in dispute.

Examples of the types of datasets that are already being compiled by those outside the courts can be found in a project of Duke University School of Law which has collected 29 datasets on many topics. The "Data Sources: Courts" website links to datasets such as: "Judges of the United States Courts – a directory of biographical information on Federal Judges, including time on the bench, race, gender, education, and professional career;" and "Supreme Court Justices Database: A database on 263 characteristics of all Justices nominated for the U.S. Supreme Court."[1] Most, such as the Supreme Court Justices Database, are available in formats that can be manipulated, searched, and collated such as CSV, XLSX (Excel), DTA and XPT.[2]

The Duke website includes, in addition to datasets that may aim to meet curiosity about judicial personalities, datasets that are confined to case data from state and federal courts such as "CourtListener APIs & Bulk Data – Free" – which is reported to contain bulk data files containing millions of court opinions from federal and state courts gathered from numerous websites.[3] For those who can pay, the CourtListener website provides, "We have a powerful system for gathering bulk data from PACER. We have millions of opinions from hundreds of jurisdictions in our database. We have tens of thousands of oral argument audio files. We have information about nearly every reporter written. Much of this data is available via our APIs, but making sense of it can be time consuming. To fill these needs, we offer data services. Jobs are billed hourly or by project milestones" at hourly rates between $70 to $750 per hour.[4]

Whether courts are willing participants or not, significant amounts of court case data are already available through non-profit and for-profit data aggregators. Whether or not courts find such information has value, the information is found in public records, and there is a keen interest in such information among many outside of the courts. To the degree such datasets are produced by courts and not left to whatever rules and practices are adopted by outside organizations, courts can address challenges to providing access to such data that include how to protect private information that might be included in court case data, the cost of providing a platform in which such data can be accessed, and how to protect against abuses.

Court Case Data Is Public Data With Important Exceptions

COSCA and courts more broadly have long struggled with how to provide access to court case information. The first judicial electronic access system came online in June 1989, providing records from four Baltimore courts to 860 registered users in a system supported by five staff members. The users included "local and state bars, title companies, credit check firms, Washington D.C. area media, and the largest

[1] *Data Sources: Courts* at the Duke University School of Law website accessed at https://web.law.duke.edu/lib/facultyservices/empirical/links/courts/

[2] File formats such as CSV (Comma Separated Value), XLSX (Microsoft Excel's workbook files), DTA (Database Tuning Advisor that works with SQL Server) and XPT (a file format compatible with Mozilla Firefox and data formats crated by SAS applications) allow manipulation of the content of datasets. They have the advantage of compatibility with many data formats while each has advantages and disadvantages to the user. LifeWire, *What is a CSV file?,* accessed at https://www.lifewire.com/csv-file-2622708; LifeWire, *What is an XLXS file?,* accessed at https://www.lifewire.com/what-is-an-xlsx-file-2622540; ReviverSoft, *What is a DTA file?,* accessed at https://www.reviversoft.com/file-extensions/dta; ReviverSoft, *What Is .XPT file extension?,* accessed at https://www.reviversoft.com/file-extensions/xpt

[3] Id.

[4] 2019 Free Law Project website, *Data Services and Consulting* accessed at https://free.law/data-consulting/

users group, private investigators."[5] This capability has been steadily advancing and many courts have increased online access to court *records*.[6]

In the decade from 1995 to 2005, the Conference of Chief Justices (CCJ), COSCA, and the National Center for State Courts (NCSC) grappled with public access to court *records* and how the term "court records" was evolving during the emerging transition to an age of electronic records. In 1995 NCSC issued *Privacy and Public Access to Electronic Court Information: A Guide to Policy Decisions for State Courts* that recognizes "[t]he right of public access to court information has deep roots in our judicial history."[7]

The 1995 NCSC paper begins with the understanding that traditional paper documents were being surpassed by computer databases with "vast amounts of information about litigants, trial records, judicial orders, aggregate compilations of case data, images of filed documents, and judge statistics" and foreshadows a very real challenge courts face in 2019 by noting that "[s]ome requesters ask for magnetic tapes of 'raw' data so they can sort and compile it themselves."[8] The NCSC paper recognizes that, "While reasonable interpretations would not exclude all electronic data from the purview of public records, neither would they grant complete public access to all electronic information stored in computer records. A reasonable compromise might be to consider the intent of the traditional definition, and then attempt to fashion an imperfect but definable parallel between the paper records and the electronic system data."[9]

The struggle with establishing appropriate policies on access to electronic information only grew when in 2000 the COSCA Policy Committee prepared a paper intended to endorse a position on "access to court records" but the membership first amended the paper's recommendations and then adopted it only as a "concept paper."[10]

In 2002, CCJ and COSCA adopted *Guidelines for Public Access to Court Records*.[11] The objective of the *Guidelines* "is to provide maximum public accessibility to court records, consistent with constitutional or other provisions of law and taking into account public policy interests that are not always fully compatible with unrestricted access."[12] The *Guidelines* apply a presumption of access to a court record "that is available in any type of electronic form" including information that is textual or graphic, imaged, and "data in the fields or files of an electronic database."[13] A subsequent 2005 report suggested language for educating litigants and the public, development of internal policies for handling records, and access to family court records.[14]

---

[5] Kevin P. Kilpatrick, *Electronic Handshake: Public Access to Court Databases 2*, NCSC (1995), accessed at https://ncsc.contentdm.oclc.org/digital/collection/accessfair/id/224/

[6] According to the National Center for State Courts' State Court Organization (Table 5.6a – Remote Online Public Access to Case Information"), 18 states provide online public access to court documents while 64 of the nation's 152 largest courts that responded to the State Court Organization survey indicated that they provide online public access to court records.

[7] Susan M. Jennen, with excerpts by Jane Nelson and Debra Roberts, *Privacy and Public Access to Electronic Court Information: A Guide to Policy Decisions for State Courts,* NCSC (1995), NCSC Publication No. R-170, p.5, accessed at https://ncsc.contentdm.oclc.org/digital/collection/accessfair/id/222/

[8] Id. at page 2.

[9] Id. at pp.22-23.

[10] COSCA, *Concept Paper on Access to Court Records*, published by NCSC Government Relations Office (August 2000)

[11] Martha Wade Steketee and Alan Carlson, *Developing CCJ/COSCA Guidelines for Public Access to Court Records: A National Project to Assist State Courts,* NCSC and the Justice Management Institute (2002), accessed at https://ncsc.contentdm.oclc.org/digital/collection/accessfair/id/210/

[12] Id. at p.4.

[13] Id. at p.20.

[14] Martha Wade Steketee and Alan Carlson, *Public Access to Court Records: Implementing the*

In a 2008 policy paper endorsing publication of court performance measure results, COSCA recognized that, "How state courts use public resources must be visible to both funding bodies and the public. With sufficient accountability and transparency, state courts have the opportunity to earn the trust and confidence of the public and the other branches of government regarding the effectiveness of state court system operations and the efficient use of public tax dollars."[15]

In a 2017 proposal for a revised model policy to update the 2002 *Guidelines*, NCSC reemphasized that court *records* are presumptively open to public access and that a policy regarding court *data* should maximize accessibility "for several reasons: to enhance public trust and confidence, to be accountable, to be transparent, to improve customer service, and to reveal common law. . . . 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."[16]

In the nascent stages of the creation of "electronic public access systems" in the mid-1990s, courts faced challenges surrounding access to *records* that are echoed today in moving beyond records to court case *data*; challenges in protecting privacy interests, funding the costs of providing access through fees or appropriations, technological bottlenecks, and concerns that there would be an increased need for management oversight of public access systems.[17] Balanced with respect for the privacy interests of the people involved with court cases, the public and others have a right of access to the data harbored within the electronic recesses of our courts. Courts should facilitate that access.

COSCA Favors Open Access To Public Court Case Data In An Accessible, Machine-Readable Format Within Constraints Imposed By Issues Of Data Standards, Data Integrity, Funding, and Privacy

To achieve the broadest possible transparency, court case data must be available "in a modern, open, electronic format that is machine-readable."[18] A definition of "machine readable" stated in federal legislation is information or data "in a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost."[19] For example, a PDF document containing tables of data is human readable but not machine readable, but the same data can be in a format such as CSV, JSON, or XML that can be automatically read and processed by a computer.[20] An academic attempt to explain the concept for lay persons provides:

> [A]s commitments to open government and transparency increase, efforts to make information available must

CCJ/COSCA Guidelines Final Project Report, NCSC and the Justice Management Institute (2005).

[15] COSCA Policy Paper, *Promoting a Culture of Accountability and Transparency: Court System Performance Measures,* (December 2008), p.13, accessed at https://cosca.ncsc.org/~/media/Microsites/Files/COSCA/Policy%20Papers/2008WhitePaper-PerformanceMeasurement-Final-Dec5-08.ashx

[16] Thomas M. Clarke and Janet Lewis De Graski, *Best Practices for Court Privacy Policy Formulation,* National Center for State Courts, July 2017, Appendix D, accessed at https://ncsc.contentdm.oclc.org/digital/collection/tech/id/876/.

[17] *Electronic Handshake, supra* note 5 at pp. 39-44.

[18] 2018 Florida Senate bill CS/CS/SB 1392 at sections 900.05(4) (mandating publication under open data standards for criminal justice data in Florida courts beginning January 1, 2019), accessed at https://www.flsenate.gov/Session/Bill/2018/1392/BillText/er/PDF

[19] OPEN Government Data Act, HR H.R.4174 (P.L. 115-435), Section 202 (definitions) accessed at https://www.congress.gov/bill/115th-congress/house-bill/4174/text#toc-H8E449FBAEFA34E45A6F1F20EFB13ED95

[20] Glossary, *Machine Readable*, Open Data Handbook, accessed at http://opendatahandbook.org/glossary/en/terms/machine-readable/

> include machine readable versions of datasets and not only reports about this information in document form (such as PDF, HTML, and JPG). Consider a bar chart in a government report. You can read the report in PDF format and understand the analysis the chart provides. However, neither the chart itself nor its underlying data is available in a way that allows further processing of that information. Next generation efforts in opening government must ensure that users have access, for example, not just to a static bar chart image, but also to information about the source of that bar chart and the underlying data itself . . .
>
> In a practical sense, machine readable information helps government agencies to bridge the gap between "documents" (which are typically static and frozen in their format) and "data" (which may be dynamic and can be open to further processing). By adopting a machine readable perspective, these same agencies can meet their open government and open data objectives more completely, reliably, and responsibly.[21]

"Open data" provides unrestricted access in a machine readable, non-proprietary format that is searchable and sortable; available at reasonable reproduction cost, available for re-use and redistribution, and "interoperable" - meaning it is in a common format that permits diverse systems and organizations to intermix different datasets.[22] However, as NCSC recognized as long ago as 1995, no court would "grant complete public access to all electronic information stored in computer records."[23] It is essential to consider the challenges and concerns that constrain open access to court case records.

A. The Complex Challenges Of Data Standards And Data Integrity

Data is a surprisingly complex commodity. For example, it would appear easy enough to establish and compare how much time it takes to reach a disposition in murder cases in two jurisdictions over a stated time period. However the task can result in grossly misleading results if the data relied upon is not uniformly determined. To fairly compare data it is critical to know they measure the same thing in the same way. Do both jurisdictions have a common definition for "murder" (only capital and intentional homicide or including manslaughter and felony murder), how does each court count a conviction (only if there is a guilty finding on the original charge, or are lesser included offense convictions counted as well, and is there a conviction when the jury returns a verdict or when the court enters a judgment and sentence), and when does the time start (upon arrest or when arraigned)?

It is not surprising that the definitions and rules that produce court case data can vary from county to county and do vary from state to state. Court case data is accumulated in an extraordinary variety of forms among approximately 83 million state court cases annually.[24] Development of a national model for state court case data began with the Court Statistics Project in 1975 as a

[21] Jim Hendler and Theresa A Pardo, *A Primer on Machine Readability for Online Documents and Data* (September 24, 2012, published on the Data.Gov blog, cited at 2018 Florida Senate bill CS/CS/SB 1392 at sections 900.05(1) (legislative findings and intent) and 900.05(4) (data publicly available), accessed at https://www.flsenate.gov/Session/Bill/2018/1392/BillText/er/PDF

[22] Open Data Handbook, *What Is Open Data?,* accessed at https://opendatahandbook.org/guide/en/what-is-open-data/

[23] *Privacy and Public Access to Electronic Court Information*, *supra* note 7, p.23.

[24] Court Statistics Project 2017 overview accessed at http://www.courtstatistics.org/

partnership between COSCA and NCSC.[25] However, the Project has been limited in its utility due to the fact that only summary caseload statistics are provided rather than detailed case-level data.

Recognizing the persistent interest in court case data and its frequent use to compare different courts or states, NCSC has undertaken the herculean task of normalizing data standards for case types across state courts. The National Open Court Data Standards (NODS) project has developed a set of data elements for seven case types (criminal, juvenile, delinquency, dependency, family, civil, probate, and traffic) in order to "strengthen state courts' ability to solve internal business problems, reduce the burden that external data requests have placed on courts, and provide open, transparent, and reliable data to the state courts and to consumers of state court data."[26] The NODS project is identified as a priority topic by the Joint Technology Committee (JTC) of COSCA, the National Association for Court Management (NACM), and NCSC.[27] The project is led by the chairs of the COSCA Statistics and Joint Technology Committees with an Advisory Council that includes judges, COSCA and NACM members, representatives from county, state, and federal courts and agencies, academics, and data requestors from the non-profit and for-profit sectors.

The NODS group convened workgroups during 2019 to draft the data standards. NCSC opened a Public Comment section on the NODS website for the period October to December 2019 to solicit "the public's commentary to aid in the development of a national set of voluntary state court data standards."[28] Opening any of the tabs on the NODS Public Comment Folder reveals the extraordinary scope and effort dedicated to this undertaking.[29] The final set of standards are scheduled for release in early 2020. With the recent endorsement from COSCA and the National Association for Court Management and the expected endorsement from many other public and private entities, the anticipated result is gradual, voluntary adoption of the state court data standards and thus more consistent reporting of court case data by counties and states. The NODS project has the potential to provide a critical enhancement in the ability of courts to provide access to court case data that can be fairly compared and analyzed.

If an examination of the NODS project does not sufficiently establish the challenge of making comparisons of court case data fair and reliable, the not yet successful efforts of others to date may suffice. In Florida, Senate Bill 1392 became law on July 1, 2018, setting targets in 2019 and 2020 to create "a model of uniform criminal justice data collection." The bill requires the clerks of court, state attorneys, public defenders, county detention facility administrators, and the Department of Corrections to collect specified data on a biweekly basis and report it to the Florida Department of Law Enforcement (FDLE) monthly, requires FDLE to publish the data and make it searchable and accessible to the public, makes any clerk of the court or county detention facility that does not comply ineligible to receive any state funding for five years after the date of noncompliance, imposes annual requirements for reporting on pretrial release programs, digitizes the Criminal Punishment Code sentencing scoresheet, and funds a pilot project in the Sixth Judicial Circuit for the purpose of improving criminal justice data transparency.[30]

---

[25] NCSC website, Court Statistics accessed at https://www.ncsc.org/services-and-experts/areas-of-expertise/court-statistics.aspx

[26] National Center for State Courts, *National Open Court Data Standards (NODS)*website, at Public Comment section, accessed at https://www.ncsc.org/nods

[27] *Priority Topics*, Joint Technology committee, accessed at https://www.ncsc.org/~/media/Files/PDF/About%20Us/Committees/JTC/JTC%20Priorities%2018-19.ashx

[28] NODS website, supra note 41 at the Public Comment section.

[29] NODS website, supra note 41, at the NODS Public Comment Folder accessed at https://nationalcenterforstatecourts.app.box.com/s/j67gwg9evgbqg866rwoyn9uzpc75xocm

[30] 2018 Florida Senate bill Summary, CS/CS/SB 1392, accessed at

The Florida legislation also intends that FDLE will "publish datasets in its possession in a modern, open, electronic format that is machine readable and readily accessible by the public on the department's website. The published data must be searchable, at a minimum, by each data element, county, circuit, and unique identifier."[31] As of October 2019, the published data provides a wealth of statistical information in Excel format about criminal offenses on the FDLE website that is downloadable pursuant to the statute.[32] The data is without any individual identifying information.

The law is intended to require Florida's "jails, prosecutors, public defenders, courts and prisons to coordinate their data collection, enabling lawmakers and the public to track how someone moves through the entire criminal justice system, from arrest to release."[33] In practice, differences in how each entity defines data elements has hindered efforts to reach this goal. The legislature funded pilot projects in the courts in Pasco and Pinellas counties (Tampa area) "to figure out the best practices for collecting the new information. . . The goal of the pilot is to link police, court and other computer systems, standardize data collection, and create a blueprint for how all 67 counties will submit data to the state agency."[34] Implementation challenges resulted in the subsequent passage of HB 7125 (2019) delaying until 2020 numerous requirements of the original legislation.

Standardizing data across executive and judicial entities is an enormous challenge. As the NODS website cautions, "these standards are intended to cover data collected and maintained by state court systems for business purposes. The standards will not cover data collected by other entities such as jails, departments of correction, probation departments, or criminal history repositories, except to the extent that the court system already obtains and stores data from these sources for internal business purposes."[35]

The federal government adopted the Open, Public, Electronic, and Necessary (OPEN) Government Data Act signed into law on January 14, 2019, as part of H.R. 4174 (P.L. 115-435), the Foundations for Evidence-Based Policymaking Act. The OPEN Government Data Act requires that federal government data be "machine readable" meaning "in a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost" and provides for unrestricted access in a machine readable, open format meaning "an underlying open standard that is maintained by a standards organization."[36] The federal statute is the codification of an effort initially begun by executive order in 2013 to make "open and machine readable" data the default for all federal government data.[37]

---

https://www.flsenate.gov/Committees/BillSummaries/2018/html/1769

[31] 2018 Florida Senate bill CS/CS/SB 1392 at sections 900.05(1) (legislative findings and intent) and 900.05(4) (data publicly available), accessed at https://www.flsenate.gov/Session/Bill/2018/1392/BillText/er/PDF

[32] Statistical information through 2018 on crime indexed by jurisdiction, violent offense types, property crimes, and other categories can be found at https://www.fdle.state.fl.us/FSAC/Data-Statistics-(1).aspx

[33] Nicole Lewis, *Can Better Data Fix Florida's Prisons?,* The Marshall Project Justice Lab, April 14, 2014, accessed at https://www.themarshallproject.org/2019/04/14/can-better-data-fix-florida-s-prisons

[34] Id.

[35] NODS website, supra note 41, at Public Comment section.

[36] OPEN Government Data Act, HR H.R.4174 (P.L. 115-435), Section 202 (definitions) accessed at https://www.congress.gov/bill/115th-congress/house-bill/4174/text#toc-H8E449FBAEFA34E45A6F1F20EFB13ED95

[37] Executive Order, *Making Open and Machine Readable the New Default for Government Information* (May 9, 2013, President Barack Obama, accessed at https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-

While the federal law mandates open data, no uniform federal data standards exist and "the mandate is unfunded, meaning that many CDOs [Chief Data Officers] will have to convince their agency leadership that the publication or their data provides value to their agency and the American public."[38] The tasks for CDOs without funding confronts the perception of a "tendency within government or any other large institution to favor risk aversion and opacity."[39] Some federal government datasets can be found on the Data.Gov clearinghouse website in 14 categories often in machine readable formats (Excel, JSON, RDF, CSV, and XML), although as of late 2019 several have not been updated since 2015.[40] Other deficiencies in attempts to legislate open court data can be found, for example, in model legislation and even legislation that has been introduced which propose open data without addressing data standardization, funding, or how to protect privacy.[41] Given that court case data is produced and managed by courts, legislation is an inappropriate and demonstrably ineffective way to determine access and limits of access to court case data.

In addition to more uniform data standards as the NODS project portends, before court case data can be made more accessible the issue of data integrity must be addressed.[42] Information in the best designed case database must be reliable in order for the resulting data to have validity. All courts grapple to some degree with data entry training and uniformity issues. Ideally all clerks in a case management system shared within a county or an entire state enter the same event in similar cases in the same way so that valid comparisons of the cases are possible. In fact, it is not always possible to establish and enforce uniform ways of recording the same events and uniformity that is achieved can be eroded by staff turnover and limited training resources.

For example, in two very similar cases the different ways of recording a dismissal can have very different implications when data is analyzed. If a prosecutor brings a motion to dismiss a pending case, it may not matter in an individual case that one clerk enters the event correctly as "dismissed by prosecutor" while another incorrectly enters "dismissed by the court." Either way, the dismissal event was recorded, and the case will terminate. However when repeated hundreds or thousands of times these two different ways of recording the same event can lead to inaccurate conclusions about the prosecutor or judge. Those who manage courts are familiar with the Sisyphean challenge of having all clerks record all events correctly. Explaining to court employees why this is critical and ensuring they have the training to correctly enter the data remains a challenge not easily met.

Courts need to establish standards and provide effective training so that courts achieve a level of confidence in the reliability of data entered into the system from which datasets released to data requestors will be drawn. In combination with that confidence it may be appropriate to flag possible data integrity issues. Courts may also invite data requestors to seek review of their interpretation of data by the courts so that the courts may alert the users to possible misinterpretations or the possibility of errors in the data that might not be evident to the user.

Efforts by state courts to provide broader access to state court case data will be greatly advanced

---

[38] Jessie Bur, *What comes after legally mandated OPEN data,* Federal Times, February 7, 2019, accessed at https://www.federaltimes.com/it-networks/2019/02/07/what-comes-after-legally-mandated-open-data/

[39] Id., quoting Christian Troncoso, director of policy at The Software Alliance.

[40] Data.Gov website, accessed at https://www.data.gov/

[41] Criminal Data Transparency Model Act, American Legislative Exchange Council (2018), section 3, accessed at https://www.alec.org/model-policy/criminal-justice-data-transparency-model/

[42] The COSCA Court Statistics Committee adopted in December 2019 a model Data Governance Policy that addresses the organizational structure necessary for effective data governance, the life cycle of data over time, and best practices for improving data quality.

with the data standards that will result from the NODS project and from the work on a data governance policy by the COSCA Court Statistics Committee. In addition, courts must work to achieve an acceptable level of data integrity and/or provide clear warnings of data integrity issues so that data requestors are aware of the limitations that may affect use of or reliance on the data. Examples from a few court websites include:

- Kentucky: Requires an MOU for any bulk data report, or any report delivered in editable format, or any aggregate report that might contain confidential personal identifying information, and also provides, "Information received from the AOC is subject to change, reprogramming, modifications in format and availability, and may not reflect the true status of court cases due to ordinary limitations, delay or error in the system's operation. The AOC disclaims all warranty as to the validity of the information obtained. The AOC accepts no responsibility for the conclusions drawn by any individual who has received data from the AOC."[43]

- Oregon: Requires execution of a Terms of Use document that includes a limitation of liability (paragraph 17) and indemnities (paragraph 18) as well as a warranty disclaimer in paragraph 16 that includes, "You understand and agree that OCJIN and its website, services, information, and data provided are being provided 'as-is' without warranty of any kind, whether express or implied, and that they may be subject to delay, deletion, theft, errors or omissions."[44] For those who request access to bulk data, Oregon requires in addition to the Terms of Use document execution of a Bulk Data Transfer Agreement containing additional requirements and restrictions on use of the data.

- Colorado: "The data on this website is provided in live time from the Colorado Judicial Department electronic database for county and district courts; however, the Colorado Judicial Department or contractors declare that the information provided does not constitute the official record of the court. Any user of the information is hereby advised that it is being provided 'as is' with no warranties, express or implied, including the implied warranty of fitness for a particular purpose. The information on this website does not represent all the cases or case types filed with the Colorado courts."[45]

- Superior Court of California, Shasta County: "Although the data found using Superior Court of California, County of Shasta access systems have been produced and processed from sources believed to be reliable, no warranty expressed or implied is made regarding accuracy, adequacy, completeness, legality, reliability or usefulness of any information. By choosing a language from the Google Translate menu, the user acknowledges that the Court cannot guarantee the accuracy of the translations and is not liable for any such inaccuracies. This disclaimer applies to both isolated and aggregate uses of the information. Superior Court of California, County of Shasta provides this information on an 'as is' basis. All warranties of any kind, express or implied, including but not limited the implied warranties of merchantability,

[43] Kentucky Court of Justice website, Administrative Office of the Courts, Statistical Reports Research & Statistics Reporting Procedure, accessed at https://kycourts.gov/aoc/statisticalreports/Documents/reportingprocedure.pdf

[44] Oregon Judicial Case Information Network Terms of Use agreement accessed at https://www.courts.oregon.gov/forms/Documents/OJCINTermsofUse.pdf

[45] Colorado Judicial Branch website, Courts Records Search page disclaimer, accessed at https://www.courts.state.co.us/Administration/Program.cfm?Program=11

> fitness for particular purpose, freedom from contamination by computer viruses, and non-infringement of proprietary rights ARE DISCLAIMED. If you find any errors or omissions, we encourage you to report them to Superior Court of California, County of Shasta."[46]

Beyond the issues with the data itself and privacy, challenges regarding funding and privacy still must be addressed.

B. Funding

The most efficient, transparent, and effective court still needs funding to operate. Courts have made dynamic advances in implementing electronic case management systems that allow electronic filing of documents in a paper-on-demand environment. Whether purchased from a vendor or developed in-house, every electronic case management system that replaced paper files with data bytes had significant initial costs and continuing costs to operate. Those investments were not necessarily made with the additional goal of facilitating public access to court case data.

The technology that provides for management of thousands and even millions of cases is not necessarily designed to easily or inexpensively satisfy requests for court case data. Electronically tracking a specific case from beginning to end is a very different matter than, for example, producing a dataset that contains all cases over the past year that involved a defendant held in jail due to inability to post a money bond, or all cases in which the state agency responsible for risk management entered into a settlement of more than $5,000.

To meet the need for funding to implement access to court case data, some courts have chosen to initiate fee-based online access to data in formats that permit manipulation of the data. For example, through the CCAP (WCCA) website, Wisconsin provides free access to non-confidential court case records and also provides online access to court case data through a vendor, Court Data Technologies LLC, which fulfills subscriber requests for court case data in an open format at a cost of $500 per month or $5,000 per year.[47]

As an alternative, some courts have chosen to seek appropriations to provide access to court case data. In 2019 the New Mexico legislature appropriated up to $250,000 for projects that include, "developing data-sharing agreements and methods of data sharing among criminal justice agencies . . . to allow system-wide analysis of criminal justice operations within the judicial district and statewide."[48] Beginning in January 2020, the courts will publish to justice partners (prosecutors, public defenders, law enforcement agencies, etc.) seven datasets in machine readable format for use by justice partners. After a pilot with court data, justice partners are expected during 2020 to also upload datasets for use by other justice partners.[49] This sharing of court case data is limited to government justice partners by legislative direction and funding to create intergovernmental sharing of data across a common platform.

There are different ways to pay for technology that allows access to court case data. COSCA has previously endorsed the view that courts should be substantially funded with appropriated funds while allowing for reasonable fees to offset specific services or programs that extend beyond court functions or that provide a direct private

[46] Superior Court of California, Shasta County website disclaimer Information accessed at http://www.shastacourts.com/General-Info/Disclaimer.shtml

[47] Wisconsin Circuit Court Data Subscription Agreement Non-State of Wisconsin Government Subscribers accessed at https://wcca.wicourts.gov/download/RESTagreement.pdf

[48] House Bill 267 (2019), section 8(B)(2) accessed at https://www.nmlegis.gov/Sessions/19%20Regular/final/HB0267.pdf

[49] Report from Dave Wasson, New Mexico Administrative Office of the Courts Chief Information Officer, October 1, 2019.

benefit.[50] Wherever access to court case data falls between traditional court functions and additional, ancillary services, part of the consideration in providing access to court case documents must include funding to pay for it.

C. Litigants And Case Participants Have A Legitimate Interest In Protecting Some Court Case Data From Public Disclosure

Long before electronic court case data existed, there was a tension between public access to court records and the privacy interests of those involved in the cases. Arguments about how tenaciously courts should protect private information have only been amplified by the ease with which electronic data can be accessed today.

> Court records present a conundrum for privacy advocates. Public access to the courts has long been a fundamental tenant of American democracy, helping to ensure that our system of justice functions fairly and that citizens can access an astonishing amount of private and sensitive information, ranging from social security numbers to the names of sexual assault victims. Given that '[t]he courts are a stage where many of life's dramas are performed, where people may be shamed, vindicated, compensated, punished, judged, or exposed,' it should come as no surprise that court records, which serve as a chronicle of these dramas, are littered with private and sensitive information about the litigants, witnesses, jurors, and others who come voluntarily or involuntarily into contact with the court system.[51]

The issue of privacy is one that state court systems must address as each begins to consider how to appropriately provide access to court case data. Guidance on this task was provided in NCSC's 2017 statement that it is the best practice to start from the position that information should be "presumptively open to public access" and that courts should then consider what to restrict based upon a need to protect court users from harm.[52] As state courts consider providing open data, it is important to review the available court case data using this analysis to provide as much transparency as possible while reducing potential harm.

One path followed by courts has been to require that any protected information revealed to those who are given access to court case data must be protected by the requestor. As with the earlier discussion of disclaimers about data reliability accepted by the requestor, the requirements to protect private information are often included in a use agreement executed by the requestor and/or embedded into court use rules intended to bind the data requestor. Kentucky requires that a requestor of data in a format that can be manipulated after receipt, such as Excel instead of PDF, execute a memorandum of understanding agreeing to restrictions on the use or dissemination of such information.[53] The general Oregon Terms of Use requirements include the requestor's agreement to restrict use of any confidential information only as permitted by law, rule, court order and contract, and, "If You access confidential information through OJCIN,

---

[50] *Courts Are Not Revenue Centers,* COSCA Policy Paper (2011), Principle 1, page 7, "Courts should be substantially funded from general governmental revenue sources, enabling them to fulfill their constitutional mandates. Court users derive a private benefit from the courts and may be charged reasonable fees partially to offset the cost of the courts borne by the public at-large. Neither courts nor specific court functions should be expected to operate exclusively from proceeds produced by fees and miscellaneous charges." Accessed at https://cosca.ncsc.org/~/media/Microsites/Files/COSCA/Policy%20Papers/CourtsAreNotRevenueCenters-Final.ashx

[51] David S. Ardia, *Privacy and Court Records: Online Access and the Loss of Practical Obscurity* (August 4, 2017) (internal citation omitted). University of Illinois Law Review, Vol. 2017, No. 5, 2017; UNC Legal Studies Research Paper. Available at SSRN: https://ssrn.com/abstract=3013704

[52] *Best Practices for Court Privacy Policy Formulation*, *supra* note 1 at pages 2-3.

[53] Reporting Procedure, *supra,* note 47.

You agree to take all reasonable steps to maintain the confidentiality of the information. You must immediately report to OJD all suspected or actual instances of improper access to or disclosure of confidential information, including mistaken access and inadvertent disclosure."[54] Further, for requestors with access to protected information, Oregon requires an additional agreement that requires the requestor to specify the need for access and expressly lists the allowable uses of the protected data.

Regarding court case data, practical obscurity presents no practical protection of private information. Policies and required agreements with those who receive court case data can help to identify and advance the protection of protected information. Only when a court is confident that there are adequate protections against broad public dissemination of protected information can the court provide broader access to court case data.

Internal Benefits to Access to Court Case Data

There exists an expectation by the other branches of government and the general public that courts will use the data available to them to better manage the courts for which we have responsibility. This management may include overall policy or specific actions taken in certain cases. One of the best ways to accomplish this is to have access to case data that permits courts and state court systems to analyze how policies or decisions are working. However, many courts themselves do not have ready access to a machine readable database of court case data, instead having access only to summary level generalized court case data. This limitation hampers efforts to evaluate the effectiveness of various practices in the courts and hamstrings judges in making evidence-based decisions. This deficiency was aptly described during an episode of the podcast "Serial":

> "What works? The court doesn't gather statistics on sentencing, and that's true for most of the country by the way. No data that says defendants…do better after 6 months of probation than after 3 years of probation or, in terms of reoffending, 4 years in prison yields better results than 7 years in prison. We just don't know, which I find rather astounding when I realize that no one is tracking this. The court keeps extensive data regarding efficiency – how many cases are moving through whose docket and how quickly – which I'm not knocking efficiency – it's important, it's why people here are not generally waiting years and years for their cases to resolve and that's good – but there's no database, locally or nationally, that shows what works, so each judge in the building has to muddle it out for him or herself."[55]

All courts have the ability to track and record individual cases. Adding the ability to draw evidence-based comparisons from court case data adds the ability to measure success across case types, competing programs, and among the many policies and practices that compete for resources. The clamor from outside of courts for court case data should be an echo of the demand within courts for the ability to put the abundance of data now in court databases to work to support evidence-based decisions about how courts operate.

A Policy Of Open Access Does Not Resolve All The Issues

COSCA recognizes that this paper does not address all of the issues surrounding providing access to court case data. COSCA and other state

[54] Terms of Use, *supra* note 46 at paragraph 12.

[55] Koenig, Sarah, host. "You've Got Some Gauls." *Serial*. September 20, 2018. https://serialpodcast.org/.

court leaders need to continue studying and formulating policies to address those outstanding issues. For instance, this paper does not resolve:

- How to combat misinformation that is generated in the age of open data
- How to address misuse of court case data by authorized users; and
- How to successfully ensure that users of court case data are accessing the most current versions, including situations where expungement, scheduled destruction, or sealing of court records has changed the official document version and thus changed the underlying data.

While these and other issues are vitally important, COSCA believes that state courts cannot wait for their resolution before beginning the transition to open data. Waiting may allow others to capture a market needing data in ways that may present even more complex concerns for courts and harm to the public.

COSCA Endorses Open Access To Court Case Data Within Limitations Based on Data Standards, Funding, and Privacy Challenges

As a starting point courts need reliable data that shares common elements in order to make supportable comparisons of court case data. States will not be uniform with regard to whether access is government funded or is supported by fees nor will there be a uniform approach to the methods and degree to which courts attempt to safeguard privacy interests in court case data. However, COSCA recognizes these issues must be addressed as courts act on the developing consensus that access to court case *data* means far more than access to electronic copies of *records*. COSCA supports efforts to standardize court case data, to achieve better reliability of the data, and to provide access to court case data in a machine-readable form the protects privacy interests as defined by each state consistent with state polices and laws as well as the ability to fund such access. Court case data transparency will promote court accountability and will promote the trust and confidence in courts that is essential to the rule of law. To this end, COSCA urges state court leaders to:

1. Establish a Data Governance Policy;

2. Adopt court case data access policies that provide open access to court case data while protecting court users from harm;

3. Consider adoption of the National Open Data Standards (NODS) data definitions and data technical standards;

4. Establish data standards and provide effective training to improve the reliability of data entered;

5. Develop funding models and seek funding to permit state courts to provide open court case data; and

6. Develop the capability for individual judges and other court leaders to utilize court case data to improve practices and policies within the state court system.