CSP
Court Statistics Project

Data Governance Policy Guide

organizational structure

data quality life cycle

2019

National Center for State Courts

Version 1.0 Adopted December 2019

Data Governance Policy Guide

A joint project of the Conference of State Court Administrators' Court Statistics Committee
and the National Center for State Courts' Court Statistics Project

Diane Robinson
Sarah Gibson

National Center for State Courts
300 Newport Avenue
Williamsburg, Virginia 23185
1.800.616.6109
www.ncsc.org or www.courtstatistics.org

PROJECT STAFF
Nicole L. Waters, Director
Diane L. Robinson, Senior Court Research Associate
Kathryn Genthon, Senior Court Research Analyst
Robert C. LaFountain, Senior Court Research Analyst
Sarah Gibson, Court Research Analyst
Alice Allred, Program Specialist

Acknowledgments

This project originated from a meeting of the Conference of State Court Administrators (COSCA) to which data specialists were invited in December 2018. The members of the COSCA Court Statistics Committee gave generously of their time, talent, and experience, and their participation was invaluable to project staff.

Pamela Q. Harris, Maryland, Chair
Laurie Dudgeon, Kentucky, Vice Chair
Patrick Carroll, III, Connecticut
Rodney Maile, Hawaii
Lily Sharpe, Wyoming
Corey Steel, Nebraska
Robin Sweet, Nevada
Jonathan Williams, Massachusetts
David Slayton, Texas, Liaison from COSCA/NACM Joint Technology Committee
Paul F. DeLosh, Liaison from NACM
Kim Nieves, Pennsylvania, data specialist liaison
Angela Garcia, Texas, data specialist liaison

Many dedicated court administrators and data specialists in courts around the country also shared their knowledge, provided guidance, and generously shared materials in the development of this work. In particular, we would like to thank the following for their work on this project:

Nancy Cozine, Oregon Judicial Department
Deb Dailey, Minnesota Judicial Branch
Craig Hagensick, Minnesota Judicial Branch
Laura Hutzel, Michigan State Court Administrative Office
Hans Jessup, Nevada Administrative Office of the Courts
Amanda Johnson, Virginia Office of the Executive Secretary
Lee Kavanaugh, Massachusetts Executive Office of the Trial Court
Nicole Le, Superior Court of Orange County
Greg Montgomery, Oregon Judicial Department
Marcy Podcopaz, Minnesota Fourth Judicial District
Jessica Roeser, Oregon Judicial Department
Leah Rose-Goodwin, Judicial Council of California
Danielle Rosete, Superior Court of Guam
Daniel Sturtevant, Kentucky Administrative Office of the Courts
Conor Wall, Oregon Judicial Department
Dave Williams, Indiana Office of Judicial Administration

Table of Contents

Executive Summary

As courts rely more heavily on data for case management, strategic planning, budgeting, and improving court performance, we recognize that data are more than by-products of case processing.

The public and justice partners increasingly depend on ready access to data, and accurate and timely data are essential for public trust and confidence in the judiciary. Data are now strategic assets of the courts and courts need strong data governance policies and practices.

Data governance is the framework by which courts reach and communicate organizational decisions around data, ensure that business activities and data management are synchronized, and develop and document long- and short-term strategies around the collection, use, and disposal of data. Data governance encompasses the people, court processes, and procedures that ensure that data are fit for managing cases, planning, and budgeting. Governance is about creating a culture around data creation and use, including how data rules are created and enforced and how disputes are resolved. Without strong data governance, courts risk wasting time and energy searching for missing information, collecting unnecessary information, correcting bad information, entering data redundantly, and making decisions repetitively and sometimes inconsistently.

This resource guide is intended to help courts create, evaluate, revise, and maintain good data governance policies and practices within the context of their own laws, rules, and regulations. Recommendations are included for staffing a data governance committee and assigning data governance roles within the organization. Concerns around the life cycle of data are addressed, from collection to use to deletion or archiving. Best practices around data quality are provided, along with a glossary and resources that have been useful for other courts.

Writing, revising, and maintaining data governance policies and procedures is an ongoing process that requires significant time and attention. Policies and procedures need to be robust enough to address the current needs and climate of the court but elastic enough to adapt to changes that will inevitably occur. It is our hope that this resource guide serves to help courts develop and improve effective data governance policies by providing both a solid foundation for courts who are at the beginning stages of development while also providing insights and resources to assist in the continuing development and improvement of policies for courts who are further along in the process of establishing data governance. The development of data governance policies and procedures is an ongoing, iterative process, and as such, this document too will evolve and adapt. We invite court leaders, data specialists, data users, and others who rely upon court data to share their experiences with data and impart best practices to the court.

Data Governance for Courts

WHY DATA GOVERNANCE?

Data are strategic assets for courts, increasingly necessary not only for the operation of the court and management of cases, but also for strategic planning, developing policies and budgets, and improving court performance. This is a significant shift from the view of data existing primarily as by-products of case processing or court management.

Planning, developing policies and budgets, and improving court performance "require small data to be well organized, consistent in quality, provided timely, and presented in a manner that is easily understandable" (Schauffler 2014). Having a data governance strategy and data governance policies make this possible. As courts collect more data, and as data become more readily accessible to both court users and the public, citizens and court leaders alike have increasing expectations of court data. This makes having a data governance policy essential.

Court governance[1] is "the framework by which courts reach and communicate organizational decisions, establish business activities, and develop long- and short-term strategies" (Tobias and Billotte 2019). Similarly, **data governance** is the framework by which courts reach and communicate organizational decisions around data, ensure that business activities and data management are synchronized, and develop and document long- and short-term strategies around the collection, use, storage, and disposal of data. Data governance encompasses the people, court processes, and procedures that ensure that data are fit to serve as strategic assets. Governance is about creating a culture around data creation and use, including how data rules are created and enforced and how disputes are resolved.

In this document, the term "court" will be used broadly to indicate all courts that share a common data system. In some areas, this may mean an entire state, in others it will be a single court.

How Can Data Governance Help You?

The days of inaccurate or inconsistent data being hidden in the courthouse basement are over. With increased transparency and expectations for access, accurate data are essential to public trust and confidence. Consider the following scenarios:

- The "pro" side of a political campaign for a ballot issue requests attorney contact information from one division of the State Court Administrator's office and receives it, while the same request from the "con" side is refused by a different division. Both decisions took hours of discussion/debate to reach.
- The legislature receives a report with the number of graduates of a problem-solving court from the research office, but the communications office publicly releases a report with a very different number of graduates.
- An individual with a misdemeanor conviction in one county has her case automatically sealed after five years, but in the neighboring county, an individual with an identical conviction five years ago still shows up on the publicly available database.

[1] All items in bold are defined in the glossary.

- Three counties define the number of criminal cases as follows:
    - County A: all the charges involved in a single incident for a single defendant;[2]
    - County B: the number of individual charges;
    - County C: all the charges involved in a single incident regardless of the number of defendants.

Funding for prosecutors and public defenders depends upon the number of cases.

- One county closes cases at disposition and reopens them if a new petition is filed. A neighboring county keeps domestic cases open in anticipation of continued litigation. Funding for the court is based on the number of new cases opened plus the number of cases reopened.

Data governance increases efficiency and improves communication by aligning data practices in different courts within the same state or territory. This alignment of data definitions and rules helps ensure consistency. Attorneys, the public, other branches of government, researchers, and media organizations have a reasonable expectation that the use of a term or method of counting is consistent across county lines. Without common definitions, the public and court staff are likely to misunderstand or misinterpret data. Getting different answers from different offices or courts erodes public trust and confidence in the reliability and integrity of court data and by extension, the judicial branch.

"The goal of promoting judicial branch excellence and innovation requires continuing our examination of court operations at all levels to identify what we can do differently or better to achieve greater efficiencies and outcomes. This effort requires accurate data and the tools necessary for leadership to understand and use data to inform decision making." (Supreme Court, State of Arizona 2019)

Supreme Court, State of Arizona 2019

What Data Do We Mean?

Courts maintain and provide different kinds of data, all of which may need to be considered in your data governance policy. If there is no existing and accessible inventory of datasets managed by the court, it may lead to inefficiency and redundancy in data collection, storage, and use.

Case management data are data used in the processing of court cases and typically housed in a case management system. Typical elements include filed date, disposed date, case type, manner of disposition, parties to a case, motions made, and orders issued.

Bulk data are data extracted from a case management system without modification or compilation.

Compiled data are data extracted from one or more case management, auxiliary, or administrative data systems and reformulated for a particular need or purpose.

Administrative data are data used to manage the court system. This may include contact information for judges, attorneys, and court staff; bar exam results; and continuing education/training schedules and materials.

[2] This method is consistent with the Court Statistics Project and the National Open Data Standards (NODS).

Aggregate data are most often summary reports, traditionally presented in annual reports. An example would be the number of civil cases filed or disposed in a particular county. Aggregate data may also be used for court performance measures, including clearance rates, time to disposition, and age of active pending caseload. Case management data, auxiliary systems data, and administrative data may all be shared or presented as aggregate data.

"Data governance emerges organically out of serial transgressions. We are bombarded with requests for data. It becomes chaos, so standards and policy are needed to mitigate risk."

– *Court Research Analyst*

Within a typical database, there are at least four kinds of data courts use:
• *Transactional data*
• *Reference data*
• *Relationship data*
• *Metadata*

Supplemental data are data not stored in the case management system but used in the management of court programs. Supplemental systems may include data used for problem solving courts, interpreter scheduling and tracking, and accounting and collections systems data.

Work product are notes, e-mails, data, presentations, data collection instruments, source code, and other elements used internally for decision-making or pending issues. Work product may be exempt from laws or rules addressing public access.

Within a typical database, there are at least four kinds of data courts use:

Transactional data are static data that represent an activity or event at a point in time. Court examples include hearing dates and filings of petitions or orders.

Reference data are data that uniquely identify a person or business. Court examples include personal identifying information about judges, attorneys, and litigants.

Relationship data describe the relationship between entities. Most electronic case management systems use relational databases. The relationship data (known as keys) link the many tables that make up the database.

Metadata describe individual data elements. These are typically elements that the average user does not see, but which describe the relationships between data fields, technical information, who can access the data, and who created it.

Data are stored in different ways. Most courts store data in a **case management system.** Some courts or states that do not share case management systems or wish to transform their case management data for analysis, decision-making, and reporting may choose to use **data warehouses.** Auxiliary and administrative data may be stored in separate databases. Regardless of where and how the data are maintained, the data governance policy should address any data over which the court exercises control.

Data Governance Basics

A few principles will help keep the focus on data governance.

- Treat court data as a strategic asset, not simply as a by-product of managing cases.
- Establish and keep **data quality** as part of the strategic plan and day-to-day practice of the courts.
- Identify key personnel and who has responsibility for data governance and data quality. This should be written into job descriptions and performance evaluations.
- Have practical data standards in place. Data standards are the rules by which data are described. They should be consistent with court policy and practices and make sense to court users.
- Have a plan and consistent strategy to identify and solve data problems.
- Make innovation and learning key parts of the court's culture.
- Knowing that disputes about data will arise, establish a mechanism to resolve conflicts among stakeholders.

How to Succeed at Data Governance

For data governance to succeed, the court must value and put the necessary resources into the effort. The court must be focused, not only on the day's docket, but on the larger issues of case management. Data governance efforts must be seen as an integral part of the day-to-day work of the courts.

Path to failed data governance efforts	Path to successful data governance
"It's IT's job."	"What is my role in data governance?"
"All I need to care about is the person standing in front of me."	"I also need to worry about those who AREN'T standing in front of me. What's happening with their cases?
"I don't have time to worry about data governance."	"I don't have time to revisit the same controversies about data again and again."

MAKING THE CASE FOR DATA GOVERNANCE

The first step in building a data governance program is to enlist the support of senior court leadership. If this is a statewide program, it may mean the State Court Administrator, the Chief Justice, or the judicial council support the effort. If it is a local court, the Chief Judge, Chief Clerk of the court, and/or the Court Administrator support the effort. Without the support of senior leadership, no data governance process will be effective or sustainable.

If senior leadership are not already aware of data governance issues, it may be necessary to build a business case first. How does the mission of the court depend upon data? What problems are created by data quality issues? Where do needs of various users conflict? What real-world problems would a data governance program help address? What are the costs to the court of low-quality data? Costs can include:

- Time spent hunting missing information;
- Redundant data entry;

- Decreased productivity;
- Data clean up;
- Difficulty in making decisions due to incomplete or inaccurate data;
- Inability to conduct meaningful evaluations;
- Staff morale; and
- Time spent making repetitive decisions because policy is not clear

If a log of common data problems does not already exist, it may be necessary to begin one to document the challenges found. A survey or brainstorming session with users of court data is also likely to expose points of friction or frustration.

GETTING STARTED

While no one data governance policy or model will work for every court system, and while each court system is likely at a different stage of developing data governance, this document is intended to provide a useful resource for state and local courts. Each section below will require bringing together, evaluating, and revising old policies and, likely, creating new policies. An entire data governance policy need not (and cannot) be developed all at once but is instead a process of continuous improvement and innovation.

Common vs Separate Case Management Systems

This resource is intended to be helpful for individual courts and for court systems. For court systems sharing a case management system, common case governance is essential across the courts. Without clear goals, definitions, and business practices, even courts sharing a case management system may be using different definitions or making different assumptions about data. For court systems in which different courts use different case management systems, composite reports and planning information are only possible if common definitions are in use.

Why common definitions? To cite one example, the definition of a criminal case in NCSC's *State Court Guide to Statistical Reporting* is a "defendant and all charges involved in a single incident" (Court Statistics Project 2019). If a court defines a case as a single charge, the number of filings will not be comparable to the courts that count a case as it is defined in the *Guide*. If a court defines a case as all defendants named in the case, it will also not be comparable. In many locations, charging decisions are solely those of the prosecuting attorney. Courts may be able to engage with the executive branch to establish standards for cases, such as one defendant, one case. If courts do not use the same definitions, either because of differences in practices of justice partners or because of different court practices, mapping the data to reporting guidelines is essential.

One important note for states without a shared case management system: a statewide agreement on data governance is just as important as it is for states with shared case management, although the agreement may be more challenging to create. The public expect equivalent access to court information, regardless of which county they happen to be in. The need for consistent information for purposes of background checks, vital records, and public safety is present regardless of the structure of the judiciary.

Creating a **data warehouse** is one approach taken by states with multiple data systems. A data warehouse is a database or set of databases created to compile, store, analyze, and report on data. A state may use a data warehouse to collect case level data or summary information from trial courts using a variety of case management systems. Having a data governance policy for a data warehouse is essential. A data warehouse can also be used as an additional level of data verification at the state level: that is, if the case information in the data warehouse differs from aggregate or summary reports, the difference is an indication of inconsistency in measurement.

Who Should Be Involved?

Creating a strong data governance policy will involve virtually every level of court staff. In determining where your court stands when it comes to management of data, and where your court wants to be, ask the following questions:

- Who are the key people needed to move in that direction? Who best knows the current state of court data?
- Who knows what data courts need to operate efficiently, both in the courtroom and behind the counter?
- What technology initiatives may be necessary to improve the data?
- What human resources changes may be necessary?
- Who has the information and authority to make the decisions necessary?

Many courts create a committee or workgroup to begin the data governance process. In some courts, this becomes the ongoing **data governance committee.** When establishing this committee or workgroup, consider including individuals filling the following roles:

- Court research & statistics;
- Information Technology (IT);
- Trial clerks and others doing direct data entry;
- Appellate clerk(s);
- Those who use/review/consume court data, including judges and/or court staff;
- Regional field positions;
- Public relations or public information office;
- General counsel; and
- Legislative liaison.

Use this information to form a data governance workgroup. The first task of this workgroup will be to identify the business use of the data: getting the right data to the right people, at the right time, in the right amount, and in the right format. This also means identifying what areas of the court rely upon others' entry of data elements. For example, what information does the court need prior to scheduling a matter or prior to holding a hearing? What information does the clerk need to be able to open or close a case? Do the different groups who touch the same data have the same definition? If the same person is involved in multiple court cases, does the court know? Is it relevant?

Data Ownership

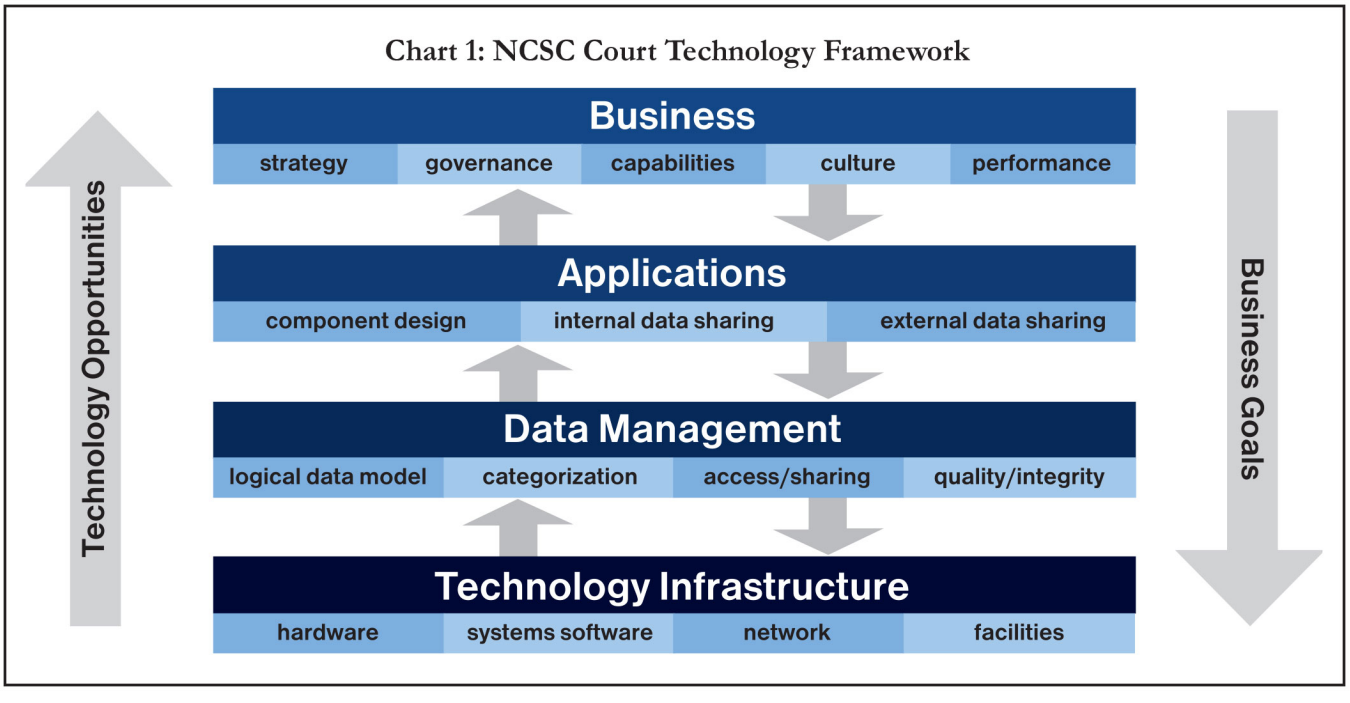
Any data governance policy needs to address who owns the data and who can release the data. Generally speaking, the organization that created the data is considered the owner of the data. If, for example, the court receives data from another agency, the data would belong to the other agency. Consider these situations:

- The court receives sex, race, and date of birth (DOB) from the State Drivers' License Agency. If a data request was made for case level criminal information, could the court release that information? If the court discovers that the DOB for a criminal defendant is incorrect, who can change it?
- In a non-unified state, the Office of the State Court Administrator (SCA) collects case level data from local courts in a data warehouse. A data request is made for statewide data – can the state office release case level data?

Any Memorandum of Understanding (MOU) or data exchange agreement should explicitly address what rights the receiving organization has in terms of data sharing, data archiving, and data updates.

DETERMINING ORGANIZATIONAL STRUCTURE AND STAFFING

Courts must determine where data governance fits in the organization. There is considerable overlap between IT functions and data governance functions. In looking at the NCSC Court Technology Framework (Chart 1), a distinction should be drawn between IT governance and data governance. The technology infrastructure is clearly an IT function. Questions of access/sharing and quality/integrity are data governance functions. The applications that enable data sharing are IT functions, while the policies of what should be shared and with whom are data governance functions.



Data Governance Committee

The data governance committee has ongoing responsibility for the data governance of the organization. Because data governance is a process of continuous improvement, the role of the committee is to evaluate and refine the data governance policy, both on a regular basis and in response to problems that arise. The data governance committee may serve in an advisory or governing role. The data governance committee for a state should work closely with the Chief Data Officer and may report to the judicial council, the SCA, or the Chief Justice. For a trial court, a data governance committee may report to the court administrator or chief judge. The committee should be representative of the data users and may be the same as the initial data governance workgroup (see previous section).

Chief Data Officer

One way to ensure that data governance efforts are institutionalized is to have a designated Chief Data Officer (CDO). A CDO's tasks include:

- overseeing the data governance process, including leading the data governance committee;
- ensuring that adopted policies and procedures are followed;
- defining data management strategy, including identifying and managing **master data;**
- defining processes for introducing and revising data fields or values;
- establishing processes for identifying, reporting, and resolving data quality issues;
- ensuring that sensitive data are identified, only collected/stored when necessary, only available to court personnel with a business need to access the information, and are not released or made public;
- improving the ability of courts and the public to access data within the scope of the law and court rules;
- establishing and facilitating training or certification for data stewards and super users of the data;
- improving the data analysis skills of court staff;
- ensuring that data are presented in ways that are meaningful, useful, and representative of the work of the courts including effective data visualization;
- critically analyzing data;
- ensuring that data destruction occurs consistent with state laws and court rules; and
- managing change.

Having a CDO is not required to address these issues but having data governance as a central or primary portion of the job function of a senior staff member is essential to developing, implementing, maintaining, and improving a data governance policy.

The **Chief Information Officer** (CIO) and CDO have different roles and responsibilities in relation to the data governance. In rare instances, the positions may be shared; however, in those situations clear expectations for the CDO responsibilities should be outlined. An example of the different functions of a CIO and a CDO is shown in Table 1.

Table 1: Roles of CIO and CDO

Role of CIO	Role of CDO
Providing the capability to capture, measure, and track court data	Providing the capability to find meaning in the data
Selecting and implementing technology to meet business requirements	Defining data requirements to meet business needs
Focusing on technology governance	Focusing on data governance
Supervising staff overseeing the hardware and software	Supervising or working closely with data stewards and data analysts

Public Access Manager

As courts face increasing demands for data, having a person whose job responsibility includes public data access is essential. In smaller jurisdictions, the CDO may have this responsibility, but in other areas the job functions may rest with someone who reports to the CDO and also works closely with communications. This position would include the following tasks:

- Receiving and evaluating ad hoc data requests; and
- Ensuring that data on a public-facing website is updated, complete, and appropriate.

Data Quality Analysts

In some jurisdictions, individuals are specifically tasked with monitoring and addressing data quality issues. Such a position would include some of the following tasks:

- Serving as a business process analyst, verifying that established business processes are followed with fidelity;
- Running and monitoring data quality reports;
- Responding to reports of data quality problems;
- Educating other staff on the importance of data quality;
- Recognizing excellence in data quality; and
- Conducting training on business processes and data quality.

In some jurisdictions, a person in this position may also be tasked with auditing revenues and disbursements to ensure that they are consistent with applicable statutes and rules. These tasks may also be part of the job description of employees who are not devoted full-time to analyzing data quality.

Data Stewards

A data steward is responsible for maintaining the integrity of the data, including data definitions and business rules. A data steward is typically assigned in each group or for each set (or subset) of data. As an example, a court may designate a data steward for data from juvenile court. This person would be a specialist in data generated in and used by the juvenile court, and could have the following tasks:

- Making recommendations to the data governance committee on matters pertaining to juvenile data;
- Approving release of aggregate juvenile data;
- Participating in decisions regarding exchange of juvenile data (with, for example, the child welfare agency);
- Monitoring data quality of juvenile data;
- Validating the accuracy of juvenile data; and
- Verifying that data are suitable for release.

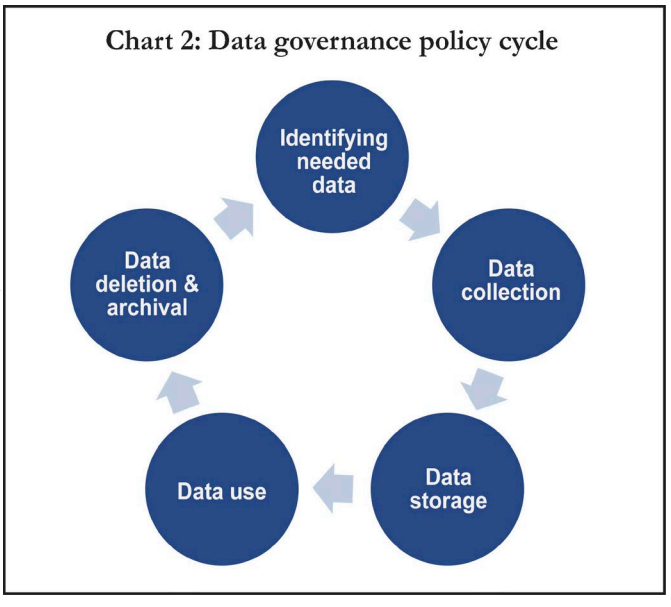
Data stewards are likely to wear more than one hat. The individual may be a clerk who is particularly knowledgeable about the subject matter, an analyst responsible for a subset of the data, or a court business analyst. Regardless of the person's job title, performing as a data steward should be a formal part of the individual's job description and be part of the individual's performance evaluation.

LIFE CYCLE OF DATA

A data governance policy should consider the entire life cycle of data, which encompasses all stages of data use from acquisition through permanent storage/archive or destruction. At a minimum, the policy should address the areas shown in chart 2.[ii]

Data Stewards in Pennsylvania are responsible for cataloging data on a site shared by all agency personnel. Cataloging includes the source of the data, interpretation details, reports and publications that rely upon those data, and how the data may be accessed.

Chart 2: Data governance policy cycle



Historically, courts were concerned about the paper court files and physical factors that affect paper files including labeling, storage, retrieval, ground water, humidity, and fire. Laws, court rules and policies regarding court records may be based on the outdated assumptions that court files are paper files. As part of data governance, each relevant law, rule, or policy should be identified, and current compliance documented. In some locations, courts may find it difficult to comply with regulations written for paper files in the age of electronic case records. When this occurs, courts will need to work to amend court rules and policies or work with their legislature to update relevant statutes.

Identifying Data Needed

For what purposes are court data needed? Typical purposes include the following:

- Maintain accurate court records;
- Manage cases;
- Inform judicial decision-making (e.g., pre-trial release decisions, risk assessments, and sentencing in criminal cases; adoptions in family cases; and risk assessments in delinquency cases);
- Inform policy;
- Plan and monitor budgets;
- Request resources from funding organizations or grantors;
- Measure performance;
- Increase efficiency and/or reduce costs;
- Comply with laws or rules.

Within these needs, basic questions that should always be answered include:

- Are these data actionable?
- What will the court do with these data?
- What will change if the court has these data?
- What will happen if the court does not collect these data?
- Are the courts the right place to collect these data?

Thinking that it would be "nice to know" is not sufficient reason to gather data, especially since tradeoffs often exist between the quantity and quality of data. As clerks are asked to enter more information, there may be an increase in incomplete data entry.

In some jurisdictions, natural language processing technology is being used to do much of the routine data entry. In these cases, the quantity of data sought becomes less important, though the clerks' role in ensuring data quality is undiminished.

SENSITIVE DATA

If data are sensitive, special consideration should be made about the necessity of collecting and storing the data. There should be a higher bar for collection of sensitive material, as the best protection against a data breach is to not have highly sensitive information in the case management system in the first place. Examples of data that merit special attention include physical and mental health evaluations, financial account numbers, citizenship status, race, and addresses. In criminal cases, witness information may be sensitive. In some jurisdictions, non-adjudicated criminal cases are considered confidential. Judicial officer contact information may be sensitive, both for reasons of security and to prevent ex parte communication. Key questions to ask include:

- What is the business need for these data?
- Who needs access to these data?
- How can access to these data be limited to those with a legitimate business need for it?
- If the court collects these potentially sensitive data, do laws or court rules require release response to a public records request? Under what circumstances?
- Who would be harmed if there were a data breach?
- What measures are in place to protect these data in the event of a data breach (encryption, separation of key elements)?
- What is stored in data fields and what is available only within a document?
- If documents are required to be redacted, is that the filing party's responsibility or the clerk's responsibility? For information on automated redaction, see the "Automated Redaction Proof of Concept Report" in the resources section.
- If a field or document is configured to be sealed or confidential, what is the process to override the default?

Several of these questions require communication and cooperation with the IT department, who should be represented on the data governance committee. The role of the data governance committee is to make decisions on what data are considered sensitive; the role of IT is determining the best way to protect sensitive data.

NATIONAL OPEN COURT DATA STANDARDS (NODS)

For a broad (but not inclusive) list of data elements collected by courts, see the NODS database, available at www.ncsc.org/nods. NODS provides guidance on business and technical court data standards to support the creation, sharing, and integration of court data. NODS is optional, aspirational, and separable: courts are not required to collect data elements included in NODS and some data elements included in NODS may not be appropriate for release. However, using common standards will improve understanding of what court data represent. The goals of NODS are:

- To make case-level state court data available to researchers, policymakers, the media, and the public to provide transparency in court operations and improve public policy;
- To make data available for public and court system use in a consistent manner that reduces the possibility of error and misinterpretation; and

- To reduce the burden on court system staff in responding to data requests.

Data Collection

Once the court has decided what information to collect, the next decisions are around how to collect the data. Court data come from a variety of sources. Much of it is entered by the staff members of clerks' offices, though increasingly, data enter the court system through electronic filing systems and through natural language processing of court documents. Clear and easily accessible training and reference materials are essential for anyone entering data, whether that is an employee in a clerk's office or an attorney filing a case.

TYPES OF DATA FIELDS

The data should be structured in a way that encourages accurate data entry. Limiting data entry to accepted options (as with a list of pre-defined values) can be helpful. Selections from pre-defined values make for more consistent and usable data, but if there is an "other" option, it is likely to be used disproportionately often if it is not closely monitored. Text fields offer great flexibility but make it very difficult to use data to monitor or improve court performance.

UNIQUE IDENTIFIERS

Having a unique identification number to identify each individual within a court database is a significant challenge. Ideally, an electronic case management system will make it easy for anyone entering information to identify and use an existing ID for an individual. Unfortunately, in nearly every case management system, the process to generate a random ID is quicker and easier than to find an existing one. The problem is further complicated by misspelled names, missing or incorrect birth dates, and lack of other demographic information, as well as by the fact that some individuals do not wish to be accurately identified by the court!

The data governance policy should address the problem of identifying information. Common questions to consider include:

- What is the unique identifier? What supplemental ID numbers are also used?
- Are there begin and end dates for identifiers?
- Is there an automated system to match multiple IDs that seem to belong to the same person? If so, is there a degree of certainty that would allow a match to be automatically made, or is human review required? How many data elements must match?
- Is the preference to err on the side of incorrectly combining IDs, or missing some duplicates?
- If duplicate IDs are entered by different jurisdictions sharing a case management system, how are these to be resolved?
- If IDs are merged incorrectly, is there a way to "undo" the merge?
- Are the standards different in different subject matters? A court may have greater interest in accurate IDs for criminal defendants than for civil litigants, for example.

Court cases also need unique identifiers. For courts that use a common case management system, cases are typically assigned unique numbers. In systems where clerks enter the case number, duplication of case numbers can occur. If multiple courts are contributing case level information to a data warehouse, a means of assigning each case to the appropriate court becomes critical. One common solution is an alphanumeric prefix or suffix added to the case number to quickly and accurately identify the court.

IMPROVING DATA QUALITY

Hennepin County (Minnesota) has implemented an e-check-in system, similar to that used by airlines. Defendants in criminal cases are given the opportunity to confirm their address, email, and phone numbers. They are also given an opportunity to self-identify race and ethnicity. The system validates the addresses. The incentive to participate is that the defendant can register for reminders of court events.

IMPROVING COVER SHEETS

When Arkansas took on the project of creating a statistical guide for the state courts, the state also revised case types to be more easily mappable to the Court Statistics Project, revised cover sheets, and implemented disposition sheets. Attorneys, judges, AOC staff, and clerks were all involved in reviewing, testing, and approving cover and disposition sheets for use statewide. Information about the forms was shared with attorneys at bar meetings statewide.

COVER & DISPOSITION SHEETS

If clerks are entering data from court filings, cover and disposition sheets make it easier for clerks to identify what should be entered. The job of the clerk's staff does not include reading and interpreting legal documents, such as the source of the criminal case (grand jury or indictment), or whether an order closes a case or is interlocutory. Staff should be able to easily locate information that needs to be entered into the case management system to initiate a case and to dispose of a case. Cover and disposition sheets must be consistent with the data dictionary and, ideally, organized to capture the most essential information in a format consistent with the case management system used by the office. They should be thoroughly tested before being put into use.

Outreach is essential to those responsible for filling out the cover and disposition sheets to convey the purpose of the forms and how accurate data affect courts and attorneys. If attorneys and court staff do not understand the role they play in collecting accurate information about court cases, cover and disposition sheets may not improve data quality. Since many attorneys and court staff rely upon ready access to court information, it is helpful to highlight the connection between accurate data and accurate and timely information.

ELECTRONIC FILING

Electronic filing (or e-filing) can be a way to improve data collection. Having data entry as close as possible to the source can improve the quality of the data, but electronic filing is not a panacea. Clerks have vital roles in verifying the information entered by attorneys and their office staff. Common problems include the overuse of "other" as a case type and as a document type. Unless attorneys enter reliable information for parties, it will be incumbent upon clerks to match any parties to parties already in the case management system.

When cases or documents are electronically filed, clear rules are necessary. Is a case considered filed when it is entered or when the clerk accepts it? In some jurisdictions, filings must be made available to the public as soon as they are filed, but this does not negate the need for clerk review.

One best practice is to run data quality reports on the rejection rate of electronic filings. If it is extremely low, it may be an indicator that clerks are not adequately reviewing filings. If it is extremely high, there may be unnecessary barriers to filing court cases. If employee identifiers and rejection codes are included, these reports become excellent tools for training.

DATA FROM EXCHANGES

Some data come from information exchanges, such as the State Drivers' License Agency, the Statewide Automated Child Welfare Information System (SACWIS), a vendor processing payments, or corrections. As part of the data governance program, the court needs a global policy on data exchanges with other organizations as well as an agreement or an MOU for each data exchange. Each MOU should include, at a minimum:

- What data will be exchanged;
- Format of data to be exchanged;
- Frequency of data exchange (e.g., real time, hourly, daily, monthly);
- How to handle inaccurate data;
- How each agency will be notified if the data sent are changed in format or content;
- Who the primary contact persons are at each agency;
- Expiration date of the agreement.

Additionally, the court must determine what data are automatically accepted into the case management system and what data require review. To learn more about data exchanges conducted by other states, see Section 5.1: External Information Exchanges in the State Court Organization survey found at www.ncsc.org/sco.

Data Storage

The data governance policy should address what data items are stored, whether history of entry should be stored, and how long data should be stored. Some data fields may be overwritten when they are updated, but for others, each change should be maintained and include a date stamp. One example would be criminal charge history. If a defendant is found guilty of different charges than were initially filed, it is important to be able to see both filing charges and disposition charges.

Data storage and backup are primarily IT functions, though the data governance policy should reference the Continuity of Operations (COOP) plan, and the COOP plan should identify what data are needed immediately for continued operations. A COOP Planning guide is available on the NCSC website at www.ncsc.org.

If data are to be stored indefinitely, the data governance plan should account for the risk of digital records not being useable in the future. Are data records transferred to a standards-based digital preservation system? An excellent resource is the JTC Resource Bulletin: Developing an Electronic Records Preservation and Disposition Plan found at www.ncsc.org.

Data Use

Strategic use of data involves getting the right data to the right person, at the right time, in the right amount, and in the right format. Communicating data effectively and efficiently is a key component of data governance.

DATA DISTRIBUTION MODELS

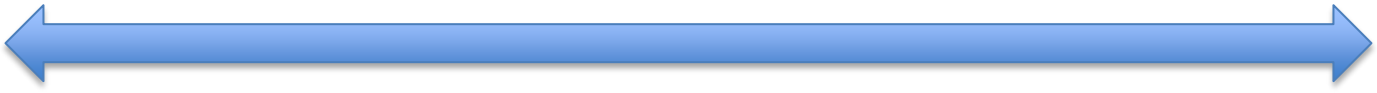
Decisions about data distribution models for the court should consider the advantages and limitations of a range of data models, from a centralized/restricted access approach to a completely open approach. Questions to consider when developing a data distribution model include:

- How reliable are the data?
- Do some users have more access to data than others?
- Who can create reports?

- What access is given to the public?

The approach to data distribution can be thought of as a continuum, from a highly centralized model in which few users have deep and wide access to a dataset, to a completely open one.

> Courts may use different distribution models along this continuum for different databases depending on the nature and reliability of the data.



Centralized Data Model	**Partially Accessible Data Model**	**Widely Accessible Data Model**	**Open Data Model**
Deep, broad access limited to core data team	Access based on roles and permissions	Many internal (and possible external) users have broad access to data	Data completely available to the public
Reports created by data team	Designated users can create reports after receiving training and/or certification	Many users create reports	Anyone can create reports
	Data team creates certified data sources for use by larger group of users		

When courts relied upon paper files, data distribution was centralized by default. With electronic case management systems and other databases, courts have more choices. Decisions must be made regarding the access to data both by internal and external users. At the extremes, a centralized model limits data access to a small group of specialists who create reports by request. An open data model allows anyone to download data and create reports.

The courts may also choose an approach of a widely or partially accessible data model. In a widely accessible data model, many users have access to the data and can create their own reports. In a partially accessible data model, designated users can access data and create reports, possibly requiring training and/or certification to do so.

Advantages and Disadvantages of Data Distribution Models

In a centralized model, the court can maintain tight control of data access, interpretation, and publication. This is appropriate for situations in which the complexity of the data outpaces the data literacy of the audience, either because the data are especially complicated or because the audience does not possess the knowledge to accurately use and interpret the data. However, content generation tends to be slow and limited, typically falling far short of

the court's need for data insights. A centralized model is unlikely to keep pace with the decision-making needs of court and agency leadership.

An open model allows the broadest access to data, allowing for internal and external users of data to download data in machine-readable formats to use without restriction (OMB 2009). This greatly reduces the burden on court staff to create reports, but creates greater risks for the misuse or misinterpretation of data.

Many courts choose approaches between these extremes. A widely accessible model allows for less specialized users to create content as needed, greatly increasing the pace at which insights can be gained from data. Interactive data dashboards or access to curated data sources[11] allow data savvy court users to answer their own questions by interacting directly with the data. For many courts, a widely accessible model of data distribution for internal users is likely to have greater institutional advantages over time as data become more and more central to the work of the courts. On the other hand, more users and more content mean more opportunities for mistakes, incorrect or misleading publications, and conflicting reports, as well as a higher risk of unintended access to sensitive data.

To mitigate these risks, many courts rely upon a partially accessible data model, in which access is based on roles and permissions. In this type of model, data stewards or other skilled users create their own reports. The court may use a centralized data team to create the court's most important content and also create certified data subsets for use by other groups of users. Users who create their own reports may be required to have training or certification.

Subsets of data created for use in a partially-accessible, widely accessible, or open data model are most useful when they:

- use flat data,
- use de-identified data,
- use simple labels,
- have standardized documentation and data dictionaries, and
- are organized in a consistent manner across all certified datasets.

Unreliable Data

Courts often struggle with what to do when there is a request for data known to be unreliable. In some jurisdictions, policies block release of such data. If state law or court rules require the release of data even if they are known to be of poor quality, the policy must also address how to handle the release, including any annotations or transmittal language. In a centralized model, this may be a standard disclaimer that goes out with unverified or unreliable data. In a partially accessible model, these data may be unavailable, or they may be available to users with explicit information about the limitations of the data. The court may make only the most reliable data available in an open model to account for these challenges (and still use a disclaimer on public-facing websites acknowledging limitations). The data governance policy should designate the person with the authority to make decisions about publication if release is not either completely prohibited or mandated. There should also be a clear mechanism for both internal and external users of data to notify the court or agency about inaccurate data.

Regardless of the distribution model, sensitive cases, including those involving minors or other vulnerable individuals, must have adequate security protocols in place to ensure that only individuals with a legitimate business need can access the information, whether they are internal or external users. Balancing protection of confidential case information while providing public access to court information remains a challenge, requiring

agility as technology, public expectations, and the law evolve (Clark, Lewis and Graski 2017, Joint Technology Committee 2018).

INTERNAL DATA USERS

Internal users of court data include judicial officers, court staff, and clerks. Key questions to consider include:

- What data are needed for each position to complete required tasks?
- Are security settings (roles & permissions) appropriate to ensure that sensitive data are only available to those who need access to complete their job functions?
- How long does it take a user to access the information they need?
- Would access to additional data improve decision-making or court processes?

Internal users of data may have competing priorities for data; the data governance policy must balance the needs of the internal users.

Another use of internal data is to monitor court performance and data quality issues. Judicial officers and court administrators should not be surprised by the content presented in reports to the legislature, another government agency, or to the public, and courts should have an opportunity to review and comment on the provided reports. Data can serve as an excellent starting point for a conversation about court practices, by providing insight into what is happening, but it cannot typically explain why.

USING DATA IN COURT PERFORMANCE

Oregon's court leadership used videoconferencing to present dashboards to trial court administrators. They explained the purpose, the path of the data from collection to the report, and how the reports were actionable.

EXTERNAL DATA USERS

State law and court rules typically address what data are publicly available online, available only at the courthouse, available through a public records request, or not available. The data governance policy on the data distribution model must be consistent with state laws and court rules.

Important questions for a court to consider regarding public access to data include:

- What case level data can the public access without a request (e.g. online)?
- How long is case information made available online?
- What aggregate or performance data can the public access without a request? Can they drill down by jurisdiction or by judicial officer? By case type?

Questions regarding data requests include:

- What data must be released when requested?
- Who is the decision maker on release of data? Does it vary by data type?
- Who in the agency or court is notified of the data request?
- How are data requests tracked?
- Does the court charge for data reports? How much? Who determines it? Who is exempt from payment?
- Who prepares the report?
- What is included in the cover memo?
- What is the policy on data believed to be unreliable? Must it be released? Is it released with a tag, such as unverified or incomplete?

Typically, if a data request is targeted at a particular judicial officer, court, or county, it is best practice to notify that judicial officer, court, or county of the request. The best practice in any release of data is to include the source(s) of the data, how the report was prepared, the date range of the records provided, the geographic area included, the known exclusions (such as expungements), and the limitations of the data.

DATA EXCHANGE PARTNERSHIPS

The San Antonio (TX) school district receives frequent data requests. To streamline requests, they established a "trusted partner" program, in which frequent requestors complete an MOU describing the exchanges, and then submit a simpler form for each data request. The MOU is approved by the board, and then the data requests can be handled at the staff level.

Public Access Policy

Most courts receive many requests for data. As part of the data governance policy, there must be standard policies for requesting data. The policies must be in compliance with state law and court rules. Questions to address in data governance planning and policy include:

- Are laws and court rules consistent with the infrastructure of case management systems?
- Who is responsible for releasing data?
- If more than one office has authority for releasing data (public information office and research office, for example), how is consistency ensured?
- How are decisions made regarding the release of non-routine or sensitive data?
- Who needs to be notified of a data request? If a data request targets a judge or court, how are they notified?
- What information is included in a cover letter or memo, if required?
- For what time frame are data released? If older, less-reliable data are requested, are they released? Is any additional information or warning about reliability included?
- In what format are data released?
    - Raw case level data: straight out of case management system, without modification or compilation. Only **personally identifiable information** (PII) (in some cases) or sealed cases are removed.
    - Curated case level data: data after some cleansing or transformation. This is typical of data in a data warehouse, datamart, or reporting software.
    - Aggregate data: summarized version of either curated or raw case-level data. This is often formatted as high-level summaries by court, case type, year etc.
- If raw case level data are released, what is the mechanism to alert the user if criminal cases are later sealed or expunged? How is removal of the record verified or enforced?
- Is there a minimum cell size for aggregate data to be released? If, for example, there are only three juvenile dependency cases in a county, it may violate confidentiality to release aggregate information for that county.
- Do data release policies cover case information, administrative databases, or both?
- Are datasets created or gathered for internal purposes releasable?
- Are the data collection instruments releasable? Are they considered work product or data?

Data should always be released in a similar format, with clear explanation of how the data were retrieved and basic information regarding how data are collected. Providing a link to the data catalog provides useful reference and context for data users.

Release of Data

Data should always be released with information providing context for the release. This information can include:

- link to the data catalog;
- clear explanation of how the data were retrieved;
- basic information regarding how the data are collected;
- the date the data were extracted;
- notice that the extract reflects the data as of that date.

It is also good practice to have a tracking number assigned to all released data so that courts can alert users to the fact that data are no longer accurate.

Means of Access

Organizations may regularly receive data from the court or have a role in direct data entry into the case management system. Examples may include external data vendors, such as the vendor of the case management system, payment system, or ODR system. Others may be justice partners, such as the State Drivers' License Agency or the Department of Corrections.

In cases where there is ongoing data exchange, critical questions to answer about the data vendor, partner, or customer include:

- Is the organization accessing the live case management system or receiving data extracts?
- If the organization is accessing the live database, what restrictions are necessary to protect court operations and accuracy of data? Some restrictions may include time of day, required data verification, or other measures.
- For what purposes can the organization use the data?
- If the court stores information in the cloud, who owns it?
- What is the organization's duty to remove or destroy data?
- Is a MOU or contract a more appropriate tool?

Data Deletion or Archival

Data governance must address the entire life cycle of data, including eventual archival or deletion of data. The data governance policy should include retention requirements and deletion policies for case management data, administrative data, working data, and data in auxiliary systems.

CASE INFORMATION

Data deletion is an area where laws and rules may be consistent with paper court files but do not address digital records. Where courts use relational databases, completely deleting a record may be difficult, particularly if there are financial records tied to the record. On the other hand, it is possible to implement an automated destruction schedule for digital records.

Deletion of a record implies destruction of a record, and this may be required under applicable laws and/or rules. This is common in juvenile cases in many jurisdictions. Policies should also consider "soft deletion" of a record and sealing records in addition to destruction. A soft deletion removes or masks PII but ensures that other information is available for statistical or research purposes. This is an important option to consider if the court wants to be able to study issues like recidivism over time.

Some individuals have their cases ordered sealed or destroyed as a result of pardons or exonerations. The data governance policy should delineate how those cases are handled.

Questions to consider in this area are:

- Is deletion of records mandatory or permissive for local courts?
- Does deletion mean a complete purge (hard deletion) or does it mean removing or masking PII (soft deletion)?
- For what period are records available for public access?
- Do records refer to data elements, documents, or both?

A useful guide is the JTC Resource Bulletin: Developing an Electronic Records Preservation and Disposition Plan available at ncsc.org.

DATA FIELDS

Data elements should also be periodically reviewed to ensure that they are still relevant and necessary. In some cases, a data field was added in reaction to a specific situation or event that no longer exists. This sometimes happens in response to a funding requirement, a legislative request, or in response to a particular concern. The effort to collect the data may long outlive the utility.

Questions to consider in this area are:

- What was the original purpose/reason for creating this data element?
- Does anyone currently use this data element? What is it used for?
- Are the data collected in another way?
- Should the information previously collected be retained?

When a data element is ended, be careful about reusing the field or the name, as it can cause confusion and data quality problems.

Data Breach Management

Unfortunately, the question of data breaches has become more "when" than "if" and courts must be prepared. IT security is properly the responsibility of IT, but the response to a data breach falls under the data governance umbrella. The data governance committee and IT should collaborate to establish a post-incident response plan.

Federal grants may require grant recipients to have written procedures in place to respond to an actual breach of data or to an imminent breach. The federal Office of Management and Budget states that potential harms that could result from the compromised PII which includes "the effect of a breach of confidentiality or fiduciary responsibility, the potential for blackmail, the disclosure of private facts, mental pain and emotional distress, financial harm, the disclosure of contact information for victims of abuse, the potential for secondary uses of the information which could result in fear or uncertainty, or the unwarranted exposure leading to humiliation or loss of self-esteem" (Office of Management and Budget 2017, 20). In assessing the risk to individuals, the OMB states that the agency must consider:

- "Nature and sensitivity of the PII potentially compromised by the breach, including the potential harms that an individual could experience for the compromise of that type of PII;
- Likelihood of access and use of PII, including whether the PII was properly encrypted or rendered partially or completely inaccessible by other means; and
- Type of breach, including the circumstances of the breach, as well as the actors involved and their intent" (Office of Management and Budget 2017, 21).

All 50 states now have data breach notification requirements, though definitions of personal information and data breach vary (Joint Technology Committee 2018)

USEFUL RESOURCES FOR DATA BREACH POLICIES

The following resources are helpful:

- From the Joint Technology Committee:
  - JTC Resource Bulletin: Responding to a Cyberattack
  - JTC Resource Bulletin: GDPR for US Courts
- Data Breach Notification in the United States and Territories from the Privacy Rights Clearinghouse
- Data Breaches & Victim Service providers: Considerations for developing effective policies from the National Network to End Domestic Violence
- Frequently Asked Questions of the VAWA confidentiality provision from the U.S.

ENSURING DATA QUALITY

Ensuring data quality must be a shared responsibility between judges, clerks, court administrators, other staff who work with data including IT and research, and data partners. With the move to electronic case management systems, data quality is fundamentally about the integrity and accuracy of the court record. Data quality encompasses **validity, accuracy,** and **timeliness** of data entry.

Many states and larger courts have research or statistics units, whose responsibility it is to research, plan, implement, and maintain a system of trial court statistics in order to provide accurate reports to both internal and external users of data. One of the continuing challenges in this task is data quality.

Court rules or state statutes may address the collection and compilation of consistent statistical information concerning the filing and resolution of cases in courts. Statutes may require a uniform system for collecting and compiling statistics and other data regarding the operation of the state court system. The court of last resort in a state may require similar statistics but may also require reports on other matters of interest. Court orders may go further and require specific information, address technology, and require audits of systems and data collection.

Why Is Data Quality Important?

Data quality is essential for measuring and meeting performance goals and for the integrity of the court. Many courts now make data accessible to anyone with an Internet connection. All users of court data, internal and external, have an expectation that data are accurate. Without a solid foundation of high-quality data, the court's ability to function effectively and serve the public is undermined. Time and resources are squandered, and confidence in the judiciary may be undermined.

For data to be accurate it must be correct and unambiguous in form and content (Harris 2014). Correct means that the value in the data field is the one intended. Unambiguous means that it is understandable by any user of the data. For example, a date of 10/11/2019 means October 11 to an American, but November 10 to a European.

Data accuracy also requires consistency: if a child can be entered with a party type of child, juvenile, minor, or ward, it will be difficult to aggregate data efficiently and correctly. City and county names must also be entered consistently: using St. Louis, Saint Louis, St Louis, and STL interchangeably will create problems.

Many **judicial officers** now have the expectation that they will be able to access accurate data about their cases and about court performance measures from wherever they happen to be: their offices, the bench, or elsewhere. Accurate data are essential for both the work on the bench in individual cases and for the larger-picture issues of monitoring court performance and budgeting. Reliable and accurate case files are "fundamental to the effectiveness of day-today court operations and fairness of judicial decisions" (National Center for State Courts 2005).

For **citizens,** accurate and timely data about court actions is essential. Consider the following examples:

- Citizen A doesn't get a job because a case that was dismissed after completion of drug court was never sealed.
- Citizen B is arrested and her children are placed into foster care because an arrest warrant was never cancelled after she resolved a failure to appear issue.
- Citizen C is hired as a nursing home aide where he assaults a patient after his previous conviction for assault is not transmitted to the criminal background check system because of a missing data field.
- Citizen D is unable to access life insurance benefits because the disposing order in her deceased husband's divorce from his previous wife was never filed.

Partner agencies rely on accurate court data, as well, for specific cases, for planning, and for budgeting and grant writing. While data sharing varies from state to state and jurisdiction to jurisdiction, some of the common case level data sharing agreements include:

- Crime Information Centers rely on the courts for outcomes of court cases, whether convicted, acquitted, or dismissed.
- State Drivers' License Agencies rely on the courts for disposition of driving offenses and license suspension information.
- Vital records rely on the courts for information regarding divorces and annulments.
- Legislative auditors rely on court data for funding and other purposes.
- Justice partners (prosecuting attorneys, public defenders, and others) rely on timely and accurate information.

Court clerks and **administrators** rely on accurate data to create dockets, to collect fees and fines, to apply for needed funding, and to manage the workflow of the court. In some cases, data collection is mandated by law for purposes such as monitoring the performance of problem-solving courts.

Policy makers rely on accurate data for decision-making, program evaluation, and budgeting.

External users may acquire data case-by-case or in bulk form. They use court data for criminal background checks, research, or other purposes. Media organizations use court data on a regular basis.

Best Practices In Data Quality

Best practices in data quality involve the people, processes, technology, and the data itself. Data quality must be integral to the business rules/standard operating procedures of the court. Preventing data quality problems is always preferable to finding them later, but strategies must be in place to ensure data quality throughout the life cycle of the data.

Frequent and active use of court data helps to ensure data quality. Regular review of information can expose data quality issues and help determine if those issues are systemic or the result of a simple data entry error. Identifying and correcting data quality issues is a process of continuous learning and improvement on the part of the court.

PEOPLE

Each person who touches court data should understand how it is used, why it is important, and their own role in ensuring accurate court records. When an issue arises with data entry, learn whether it was a simple error (such as a mistyped date) or whether the error occurs across multiple people or offices. Further investigation can help determine an appropriate response by identifying whether it is a one-time error, a common error made by a single user (which requires re-training of that individual) or a widespread error made by many users (which may require a change in the system or in training materials).

Training

Consistent, on-demand training is essential for all data entry staff. High turnover is the norm in many clerks' offices and passing on "how we do things" informally will lead to data quality problems over time. Training must be consistent with written guides for entry of case information and with day-to-day practice in the office.

The training program should also address how new rules or procedures are communicated and monitored. Any new process should receive additional data quality scrutiny in the early phases of rollout. All individuals who perform data entry should be included in training. For courts with e-filing, where attorneys or their staff members are directly entering case information, attorneys should be offered formal training, ideally with continuing legal education (CLE) credit. If non-clerk employees of a court enter data into the system, they should also be included in training.

When consistent errors are revealed, supervisors should take the time to re-train employees and have them correct the errors. When the supervisor simply corrects the error herself, an opportunity for learning is lost.

In response to continuing training needs, Massachusetts created on-demand video training of correct data entry. Both Massachusetts and Ohio have implemented training on the importance of data.

Recognition

Recognizing work well done is an important component of any data quality program. Consider data quality awards for employees who meet performance benchmarks and for courts, units, divisions, or counties that consistently achieve quality goals or show improvement in data quality. "Catch 'em being good" is an essential part of ensuring data quality.

PROCESSES

Clear and documented processes are essential for good data quality. Processes include those for case management and other data entry tasks as well as for reporting on and remediating data quality problems.

Published Court Administration Processes

The court should have easily accessible and clearly written business processes. In addition to definitions, the following should be included:

- How cases are initiated in the case management system;
- What needs to be entered;

- Which case participants need to be entered and how they are identified;
- How court actions are recorded; and
- How cases are closed in the case management system.

Developing standardized business processes must include subject-matter experts from the courts to ensure that they are workable. If multiple courts are involved, they may be tested or piloted by a few courts. Once they are implemented, they should be evaluated to look for inconsistencies or common problems with adherence. They can then be amended or revised, if needed, or training may need to be adjusted.

Written business processes must be reviewed on a regular basis. New court rules or statutes may mandate changes or court practice may change. Published court administration processes undergo a continuous cycle of upgrade and improvement.

Reporting Data Quality Issues

The data governance policy should make it clear what data users are to do when they find data quality problems. These users may include judges, clerks, and other court staff.

- How/to whom should users of court data report a data quality problem?
- What data errors can be corrected by any court staff member? What errors must be corrected by a staff member in the clerk's office?
- What changes or corrections need to be approved by a supervisor?
- How are systemic issues to be reported?
- Where can suggestions for changes in policy or practice be made and who will consider the suggestions?
- How are issues brought to the data governance committee?

A review of security levels of cases by type revealed an inappropriately low security level for a sensitive case type. Further review revealed that the problem was in a single county and, further, by a single user. This user was a deputized clerk who, it was discovered, had not been included in ongoing training efforts and was routinely changing the default security level, which the user believed to be in error. The remediation plan included training the user on this issue as well as making sure she was included in future training.

Remediation

When data entry errors are found, a remediation plan to correct the errors should be put into place. If an error was simply a missed digit or letter, an extensive plan is not necessary (though if any one user is making many of those errors it will need to be addressed). If an error is a result of a lack of understanding of a business process or an undeveloped or underdeveloped business process, the mediation plan may address the process and training on the process. Whether the error comes from a few users or is widespread will affect the remediation plan.

The remediation plan will also include data cleanup. The plan should address:

- What was the source of the error?
- Who has authority to change data or to approve changes?
- Can the errors be corrected through an automated process or does each case need to be corrected by hand?
- Who will verify that corrections have been made and how will the data be validated afterwards?
- Will past errors be corrected, or is it sufficient to correct the process going forward? If including past errors, how far back will errors be corrected?
- Should any reports or stored historical data be refreshed to reflect the corrected data?

- How will data users be affected by the remediation?
- Do any recipients of past data extracts need to be notified of the remediation? If so, which ones? How is notification made?

Any change made to data should be clearly documented. Rather than deleting data, best practice is for a correcting entry to be made. This allows for tracking of errors, reduces the chances for manipulation of court data, and allows for "continuous auditing" of court data (McMillan 2018).

Data quality reviews should also examine whether remediation plans have been sufficient to correct data quality problems. This includes, at a minimum, examining the error rate before and after the remediation plan was put into place.

TECHNOLOGY

Everywhere possible, technology should make it more difficult to enter poor-quality data and easier to enter good-quality data. Text boxes and choices of "other" should be used sparingly, perhaps requiring an extra step or justification. Technology should also make it easier to identify data quality problems. Common strategies include data validation, running exception reports, using electronic filing, and using data visualization to reveal problems.

Case management systems may be able to use data hierarchy to improve data quality. For example, selection of case types can be organized so that more likely predominant issues (divorce) are listed above likely secondary issues (child support). "Other" should never be the first option in a case type list!

Data Validation

Data validation should exist within the case management system to prevent logical errors whenever possible. Examples include the following:

- Dates entered into the system are within allowable limits: no hearing should be scheduled for 100 years in the future or the past, a disposition should not occur prior to case filing, and a criminal defendant should not have the current year or a future year as his/her birth year.
- Case types should be verified for logic and be consistent with subject areas. The system should not allow, for example, a contract dispute to be filed as a juvenile case.
- A trial type should be consistent with the case type. Data validation should prevent entering a jury trial on a juvenile status offense case if that is not permissible under state law.
- A judge assignment must be consistent with the case type and/or geographic location if that is the practice of the court.

Unusual circumstances arise in case processing. Some validation may prevent entry of data that are correct, but rare, so there must be a mechanism to override validation errors. The data governance policy should spell out who has the authority to override validation errors.

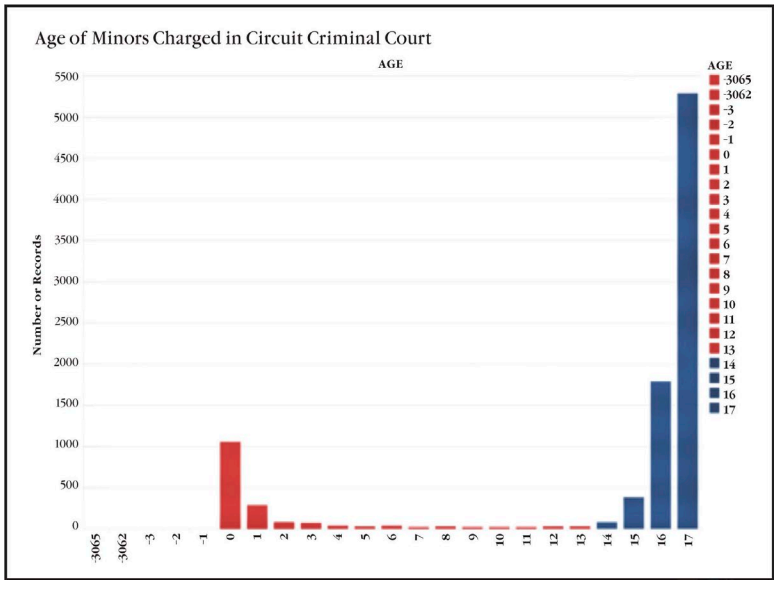
If the case management system used by the court does not allow for validation at data entry, the need for exception reports becomes even more pronounced.

Exception Reports

Every court should run exception reports on a regular basis. An exception report shows data that do not conform to expectations, such as criminal defendants with a date of birth in the same year as the alleged offense. Specific exception reports will depend on the business rules, data restrictions, and priorities of the courts, but typical categories of exception reports include:

- Logical errors, such as an implausible date or sequence of events;
- Inappropriate security levels for documents, cases, events, or parties;

- Case initiation errors, particularly those lacking essential data such as citation dates, fingerprint tracking numbers, or valid violations in criminal matters; or a marriage or separation date in a dissolution.
- Case processing errors, such as when a warrant is still active in a closed case.
- Case disposition errors, including cases in which a final order is filed but a case is not closed or the use of a civil disposition on a criminal case (or vice versa).



A judge in a rural county was inaccurately assigned to a case across the state from her district. The correct judge was never notified of filings and the case languished as a result.

Examples of more specific exception reports would include the following:

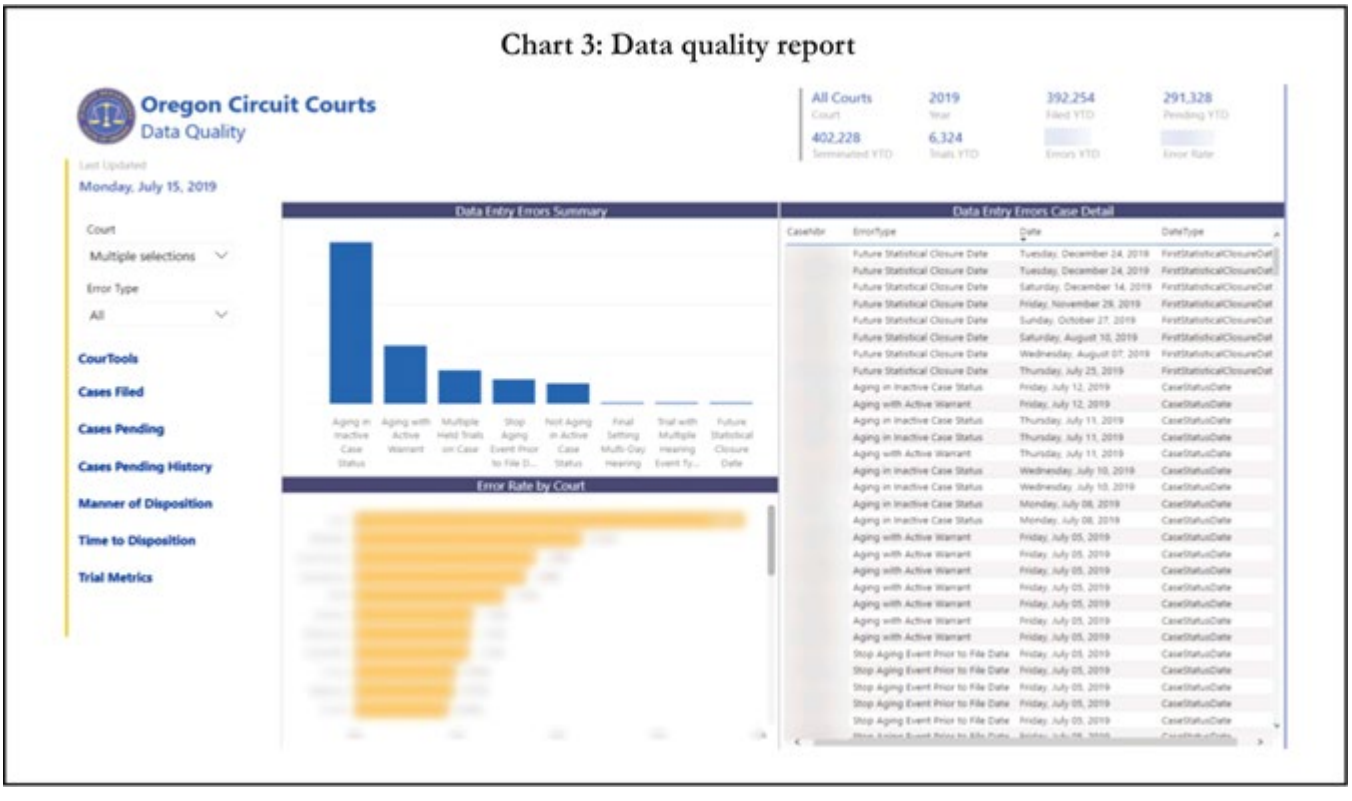
- Cases in which a trial or evidentiary hearing has been held but no order has been recorded;
- Cases that have had no activity for a designated period (the time will depend on law and court rules);
- A judicial officer assigned to a case not in the correct jurisdiction or of a case type the judicial officer does not hear.

Exception reports are only helpful if courts can see the cases or data elements that make up the number provided. In the example of cases with no activity for a designated period of time, knowing that ten cases meet that description is not actionable information. Knowing that these ten specific cases meet that description is actionable.

One example of a statewide dashboard of data quality exception reports is Oregon (chart 3). They allow courts to see areas of concern and how they are performing compared to other courts. The names of courts are deliberately blurred in this sample.

Electronic Filing

A well-designed electronic filing system has the potential to improve data quality by moving the data entry to the source of the data: the filer knows more about the case than a clerk receiving it. However, electronic filing systems may also cause or exacerbate data quality problems depending upon the design, policies and procedures, training, and use. An electronic filing system typically expands the number of users requiring training and updates, including many users who have less accountability to the courts as they are not employees. Robust clerk review is essential for ensuring data quality. There is sometimes a tendency to approve all filings without checking for accuracy or completeness. Even in jurisdictions with a requirement to make cases immediately available upon filing, clerk review is essential, as are policies and procedures for addressing errors found in the review.



Data Visualization

Data visualization can play an important role in data quality by making it easier to identify anomalies, to create positive peer pressure, and to highlight best practices. Some courts use highly visible symbols, such as stoplight colors of red/yellow/green to indicate problems, but consideration should be given to users who may be color blind or who may be printing in black and white when designing data visualizations.

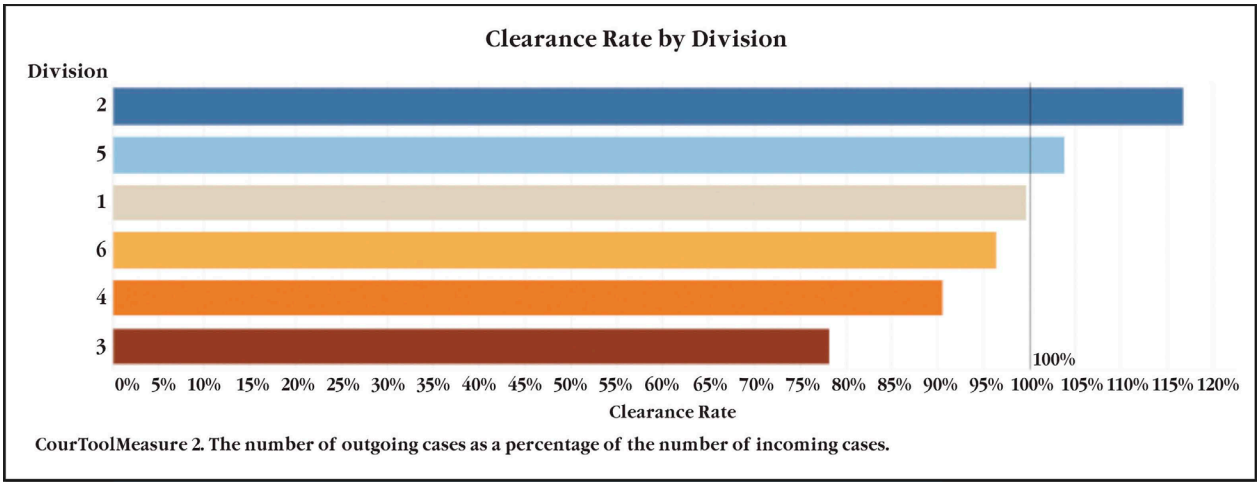
A judge's first reaction may be: "this report is wrong." Explaining where the data come from and how the data were pulled for the visualization is helpful. Always double check to make sure the methodology was sound: it can be very difficult to re-build credibility after errors. If the methodology was good, then it is an opportunity to discuss why the data are showing something unexpected.

Test all data visualizations, with users before publishing or sharing them widely. The following questions may be helpful for testers, who should be from the group targeted:

- What is the main message of this visualization?
- What would you do with this information?

Data visualizations are only helpful if they are monitored and understood by individuals in a position to act. For example, a judicial officer may find a report showing CourTool Measure 4: Age of Active Pending Caseload (National Center for State Courts 2013) helpful if sorted by case types. A chief judge may find a data visualization that includes comparisons of courts to be helpful. The example below in chart 4 shows clearance rate by judge at the halfway point of the year. A chief judge can see that the judges in divisions 2 and 5 have good clearance rates, but that the judges in divisions 3 and 4 do not. The judges can use this information to learn what practices in divisions 2 and 5 are leading to good clearance rates. They can also determine the root of the problems in divisions 3 and 4: have these divisions been assigned too many cases? Do they need more support? Is there a process problem, such as allowing many continuances or lack of follow-through on final orders? Are cases not being closed properly in the case management system in those divisions?

Chart 4: Sample clearance rate chart



A clerk may find a report showing invalid data helpful, particularly if the clerk can drill down to see the cases behind the visualization. In chart 5 below, defendants under the age of 18 charged in criminal court are shown. The blue bars indicate plausible data (in this jurisdiction minors aged 14-17 can be charged as adults for certain crimes) and the red bars indicate data representing circumstances that are not possible – a newborn cannot be charged with a crime.

DATA

Some data quality issues are a result of particular characteristics of the data, including poorly defined data and poor quality source data.

Data Definitions

Data dictionaries take two forms. The first is a technical document, used primarily by IT staff, that describes the attributes of the data (the metadata). Typical information includes the name of the data field, the format and length of the field, and who has access to the field. Possible format can be a list of potential values, a true/false or yes/no designation, a date, a number field, or a text field.

The second form of a dictionary is oriented to court processes and intended for users, sometimes known as a logical data dictionary. It explains what is meant by the various terms and by the values available in any given data field. If court terms are not clearly defined, data quality efforts will not be effective. The National Center for State Courts publishes the State Court Guide to Statistical Reporting. This guide provides a high-level overview of case

types measured at the national level, definitions of cases, case statuses, case characteristics, and manners of disposition. Many states also publish their own guides to statistical reporting, which can be more detailed.

The Conference of State Court Administrators and the National Center for State Courts have developed business and technical court data standards to support the creation, sharing, and integration of court data through the National Open Court Data Standards (NODS) project. NODS is a useful resource for courts in establishing their own data standards. More information can be found at www.ncsc.org/nods.

Poor Quality Source Data

If the court exchanges data with another agency, poor quality data entering from the other agency may be an issue. Any MOU should spell out how data quality problems are to be addressed. Poor quality data may be rejected. Possible errors may be flagged. There should be a process for addressing conflicting data, such as two agencies having different spellings of a person's name or two different dates of birth. If an agency flags poor quality source data, the originating agency should be notified.

Conclusion

High quality data are essential for effective and efficient court operation and serve to bolster public trust and confidence in the judiciary. In order to fully achieve these purposes, data must be accessible, accurate, and standardized. This requires carefully considered data governance policies.

Writing, revising, and maintaining data governance policies and procedures is an ongoing process that requires significant time and attention. Policies and procedures need to be robust enough to address the current needs and climate of the court but elastic enough to adapt to changes that will inevitably occur. It is our hope that this resource guide serves to help courts develop and improve effective data governance policies by providing both a solid foundation for courts who are at the beginning stages of development while also providing insights and resources to assist in the continuing development and improvement of policies for courts who are further along in the process of establishing data governance. The development of data governance policies and procedures is an ongoing, iterative process, and as such, this document too will evolve and adapt. We invite court leaders, data specialists, data users, and others who rely upon court data to share their experiences with data and impart best practices to the court community.

Glossary

Accuracy	Data value is correct and unambiguous.
Aggregate data	Summary data or data resulting from compilation of records.
Case management data	Data used in the processing of court cases. Typical elements include filing date, disposal date, manner of disposition, parties to a case, motions made, orders issued.
Case management system	Database containing essential data necessary for day-to-day processing of court cases, designed for fast and efficient storage and retrieval of data.
Chief Data Officer	The person in an organization who oversees the data governance process, defines data management strategy, quantifies the impacts of data quality issues, improves the ability of courts and the public to access data, critically analyzes data, and focuses on the data analysis skills of court staff.
Chief Information Officer	The person in an organization who oversees information technology strategy, implementation, and enterprise solutions to meet the needs of the court system.
Compiled data	Data extracted from one or more case management, auxiliary, or administrative data systems and reformulated for a particular need or purpose.
Court governance	The framework by which courts reach and communicate organizational decisions, establish business activities, and develop long and short-term strategies.
Data governance	The framework by which courts reach and communicate organizational decisions around data, ensure that business activities and data management are synchronized, and develop long- and short-term strategies around the collection, use, and disposal of data.
Data governance committee	The group in an organization with ongoing responsibility for the evaluating and refining data governance policy, both on a regular basis and in response to problems that arise.
Data quality	Data are reflective of actual court performance and are fit for court operations, decision-making, and planning.
Data steward	A person responsible for maintaining the integrity of the data, including data definitions and business rules.

Data warehouse	Database used for analysis and decision-making, separate from the case management system(s). A data warehouse sacrifices efficiency in order to optimize analysis and reporting using a process known as denormalization.
De-identified data	A record in which identifying information, including name, street address, day and month of birth, social security number, and phone numbers has been removed.
Deletion, hard	Complete removal of information from a case management system.
Deletion, soft	Removal or masking of personally identifiable information while ensuring that other information is available for statistical or research purposes.
Flat database	Data are stored in a plain text file in which each line contains a single record with fields most often indicated by tabs or commas
Master data	Data identified as critical data for court operations.
Metadata	Data that describe individual data elements. These are typically elements that the average user does not see, but which describe the relationships between data fields, technical information, who can access the data, and who created it. It can also provide chain-of-custody information for electronic records.
Personally Identifiable Information (PII)	Information that can be used to distinguish or trace an individual's identity, either alone or when combined with other personal or identifying information that is linked or linkable to a specific individual. (OMB 2010)
Relationship data	Data that describe the relationship between entities. Most electronic case management systems use relational databases. The relationship data (known as keys) link the many tables that make up the database.
Reference data	Data that uniquely identify a person or business. Court examples include information about judges, attorneys, and litigants.
Supplemental data	Data not stored in the case management system but used in the management of court programs. Auxiliary systems may include data used for problem solving courts, interpreter scheduling and tracking, and accounting and collections systems data.
Transactional data	Static data that represent an activity or event at a point in time. Court examples include hearing dates and filings of petitions or orders.
Validity	A data value is feasible, but may or may not be accurate.
Work product	Writing, notes, or other data elements used internally for decision-making or pending issues.

References

Clark, Thomas M., Jannet Lewis, and Di Graski. 2017. Best practices for court privacy policy formulation. Williamsburg, VA: National Center for State Courts. Accessed December 26, 2019. https://www.ncsc.org/~/media/Files/PDF/About%20Us/Committees/JTC/Best%20Practices%20Privacy%20-%20July%202017.ashx.

Court Statistics Project. 2019. State Court Guide to Statistical Reporting. Williamsburg, VA: National Center for State Courts.

Harris, Jim. 2014. The two characteristics of data quality. January 28. http://www.ocdqblog.com/home/the-two-characteristics-of-data-accuracy.html.

Joint Technology Committee. 2018. "GDPR for US courts." JTC Resource Bulletin. Accessed December 26, 2019. https://www.ncsc.org/~/media/Files/PDF/About%20Us/Committees/JTC/JTC%20Resource%20Bulletins/2018-09-19%20GDPR%20for%20US%20Courts-FINAL.ashx.

McMillan, James E. 2018. "Deleting Court Data." Court Technology Bulletin. June 16. Accessed October 31, 2019. https://courttechbulletin.blogspot.com/2018/06/deleting-court-data.html.

National Center for State Courts. 2013. "Age of Active Pending Caseload." Trial Court Performance Measures. Accessed October 30, 2019. http://www.courtools.org/Trial-Court-Performance-Measures.aspx.

—. 2005. "Reliability and integrity of court files." Courtools. http://www.courtools.org/~/media/Microsites/Files/CourTools/courtools_Trial_measure6_Reliability_And_Integrity_Of_Case_Files.ashx.

Office of Management and Budget. 2017. Preparing for and responding to a breach of personally identifiable information. Washington, DC: OMB. Accessed October 23, 2019. https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2017/m-17-12_0.pdf.

OMB. 2010. Guidance for agency use of third-party websites and applications. OMB Memorandum M-10-23, OMB. https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2010/m10-23.pdf.

OMB. 2009. M-13-13 -- Memorandum for the heads of executive departments and agencies. Memorandu, Washington, DC: OMB. Accessed January 20, 2020. https://project-open-data.cio.gov/policy-memo/.

Schauffler, Richard. 2014. Big Data: What state courts should know. Joint Technology Committee. Accessed December 26, 2019. https://www.ncsc.org/~/media/Files/PDF/About%20Us/Committees/JTC/JTC%20Resource%20Bulletins/Big%20Data%201%200%201-23-2015%20FINAL.ashx.

Supreme Court, State of Arizona. 2019. "Justice for the Future 2019-2024." 10.

Tobias, Patti, and Raymond Billotte. 2019. "NACM Core Competency - Court Governance." National Association for Court Managers Annual Conference. Las Vegas, NV.