

# AOPC DATA GOVERNANCE POLICY

## Explanatory Report

June 2020

# I. Overview

*"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."*

National Center for State Courts Data Governance Template  
January 2020

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 governance is the framework by which the AOPC can reach and communicate organizational decisions around data, ensure that business activities and data management practices are synchronized, and develop and document strategies around the collection, use, and disposal of data. Data governance encompasses the people and procedures that ensure data are fit for important processes such as analysis, decision making, planning and budgeting. Without strong data governance, AOPC risks basing decisions on poor quality or misunderstood data, releasing inaccurate data to the public, redundant data collection efforts, wasting time correcting bad information, and making decisions repetitively and sometimes inconsistently.

The Data Governance Committee was established in September of 2018 to develop a policy and recommend how a policy might be implemented within AOPC. In addition

to establishing policies and procedures, there is also a need for regular communication among departments and individuals across the agency that produce and use AOPC data. The committee will continue to meet quarterly to communicate and collaborate around current and new data topics.

## WHY DO WE NEED DATA GOVERNANCE?

Data must be accessible, accurate, and standardized to be useful. Having a data governance strategy and policies make this possible. As the judiciary collects and shares more data, and as many of those data are publically available by law, accessibility, accuracy and standardization must be clearly defined and consistent across departments. This is the role of a data governance policy. Attorneys, the public, other branches of government, researchers, and media organizations have a reasonable expectation that data and terminology are consistent across AOPC. Getting different answers from different departments erodes public trust and confidence in the reliability and

integrity of court data and by extension, the judicial branch. Standard policies mitigate the risk of proliferating invalid, contradictory, incomplete or unverified data.

A data governance policy also standardizes the expectations of data requesters who need to know what types of requests are feasible, what time frame can be anticipated for the fulfillment of requests, and that all requestors are treated equally in accordance with standard business practices.

Consider the following hypothetical scenarios:

- A reporter covering Protection from Abuse (PFA) topics in the Commonwealth regularly requests data from the Research Department to write articles about PFA. For another story on domestic violence, the reporter makes a Public Access request for data from the automated criminal case management system for the same period, but is told there is a 10-year limit on criminal data requests. The Research Department has not been imposing limits on this same requestor and the requestor is suspicious as to why AOPC limits access to only select data.
- A legislative staffer wants to know why the number of drug court graduates provided by the research office is different from the number of drug court graduates provided by the drug court program.
- In order to consider a new procedural rule, the Supreme Court asks the Problem Solving Court Unit to examine Driving Under the Influence cases over the past 10 years. An analyst requests and receives an enormous dataset from the Common Pleas Case Management System and spends weeks sorting through hundreds of thousands of cases to prepare the data for analysis. She later discovers that the Research Department maintains an annual DUI dataset that includes the same data she was interested in, and had also spent weeks preparing the data. Both departments would have saved the agency and the court time and resources had they known of each other's redundant efforts.
- The Human Resources department circulates a report to managers about employee leave statistics, intended for monitoring and evaluation of potential abuses. In a casual conversation at a committee meeting, a district court administrator (DCA) expresses interest in seeing the report to help her evaluate her own staff patterns. Even though the report is clearly marked "For Internal Use Only", it is shared with the DCA whom the AOPC manager considers to be an "Internal" entity, when in fact the "Internal" labeling of the report was intended to restrict circulation outside the agency.

These scenarios can occur when data functions are separate and distinct between

departments, as opposed to standardized and uniform across the agency. The latter are codified in a Data Governance Policy that treats data as an asset of the agency, not a product of a department.



## STATEMENT OF GENERAL POLICY

The Data Governance Policy aligns data practices among departments; ensures that the rules informing the collection, use, storage and disposal of data are consistent; and, sets uniform quality standards for agency data.

External requests for data from the automated case management systems that are managed by IT are governed by the Public Access Policies of the Unified Judicial System and are excluded from the Data Governance Policy.



## SCOPE: DATA INCLUDED IN THE GOVERNANCE POLICY

The AOPC Data Governance Policy covers all data collected, compiled and maintained by the departments and personnel of the Administrative Office of the Pennsylvania Courts for both internal and external purposes. It includes data:

1. collected from the judicial districts;
2. provided to AOPC from judicial district case management systems;
3. requested from other local and state agencies;
4. collected via survey or data collection form; or
5. otherwise compiled by AOPC personnel for business purposes.

Examples of datasets included in the policy are:

- Caseload statistics
  - Criminal, civil, dependency, delinquency, PFA, custody, divorce, and other case types appearing in the annual statistical report and online dashboards
- Problem solving court statistics
- Collections and disbursements
- Judicial Needs Assessment
- Finance and human resource data
- Data provided to Communications and Legislative Affairs for public release

### **Exclusion of public access requests**

Public Access requests processed by the Data Exchange Unit are expressly excluded from the policy. However, it is important to clarify that the Data Exchange Unit provides "raw" data extracts to AOPC departments that are transformed in significant ways by AOPC personnel such that they no longer resemble the bulk data extracts. These transformations include combining several

datasets together, performing computations such as percentages, or correcting systematic errors. After these transformations, the final datasets are considered AOPC work products and become subject to the Data Governance Policy. In this manner, bulk data that are available through the Public Access Policies may take forms that are non-public once processed by AOPC into an internal project deliverable or work product.

### **Exclusion of AOPC work products**

The Data Governance policy requires that agency data products are designated either public or non-public based on 1) assessment of compliance with the UJS Public Access Policies; 2) data quality in accordance with the Data Governance Policy Data Quality Standards; and, 3) whether the data qualify as a work product. These determinations are made by the department director and the Data Steward, and the designation is included in the Data Catalog.

For example:

A report on how often each Magisterial District Judge confines low-level offenders is produced for a Supreme Court Rules Committee to evaluate the impact of a new procedural rule. The report uses confinement data from the automated Magisterial District Justice System, all of which are available via request through the UJS Public Access Policies. However, for the final report to the Committee, a data analyst has calculated the rates of confinement for each judge, and ranked judges according to those rates and caseload size. This report is now designated as Non-Public because the underlying data have

been configured into a work product by an AOPC employee. While similar source data items can be requested from the Data Exchange Unit, the report itself and data analyses cannot. The report is categorized as non-public because it is intended for use by employees of AOPC and the committee members only, and is not made available via the AOPC website or distributed beyond its internal business purpose.

Where clarification of the public/non-public designation is needed, the Public Access Committee and/or the Court Administrator would make the determination.

Ad hoc queries fulfilled on behalf of the Court, the Court Administrator, or in service of any short term directive are not considered agency data and are excluded from the policy.

## II. Data Governance Policy

*"Data Management is about culture change. It can no longer be "someone else's responsibility." It must be the responsibility of every individual within a firm to ensure that the data they provide into the operational workings of a firm are accurate, complete and timely."*

*John Bottega, Chairman,  
Enterprise Data Management Council*

*a global organization, with over 200 member organizations from the US, Canada, UK, Europe, South Africa, Japan, Asia, Singapore and Australia, and over 7000 data management professionals as members*

### DATA GOVERNANCE COMMITTEE

The Data Governance Committee's purpose is to optimize AOPC's data resources to advance decision making, support evaluation and research, inform the public, and facilitate communication and transparency around data.

Ongoing objectives include:

- improving the overall quality of agency data;
- streamlining processes for providing data to internal and external requesters;
- maintaining, improving and expanding data products in accordance with professional and agency standards, including public dashboards, annual reports, topical reports and budget documents; and,

- facilitating the availability and understanding of agency data and related products to internal and external users.

These objectives will be met through six main activities:

1. establishing and maintaining an internal website that lists available agency datasets and how to access them;
2. documenting the steps involved in producing the various datasets;
3. publishing schedules of when datasets are refreshed or updated;
4. creating agency wide standards for validity, reliability, accuracy and uniform definitions;
5. identifying existing tools, software licenses and personnel skill sets that optimize the use of data and facilitate sharing among departments; and
6. reviewing and monitoring requests for new data and recommending processes that

are in line with available resources and competing priorities, including agency studies, surveys and ad hoc data collections impacting judicial districts.

The Data Governance Committee does not authorize datasets, research projects, technology purchases or any other data related initiatives, but can review and make recommendations upon request.



## DATA CATALOG AND DATA STEWARDS

Documenting how each dataset is created, who in the agency is responsible for its maintenance, and how often the data are updated is a critical task that supports the availability and proper use of the agency's datasets by AOPC personnel. The Data Governance Committee will maintain an electronic Data Catalog as a "user's guide" to existing AOPC datasets and graphics. The goals of the Data Catalog are to streamline and simplify access to datasets; ensure the most recent version of the dataset is being shared; identify the most knowledgeable AOPC point person who can assist with the proper use and interpretation of the data; tighten departmental control over the flow of non-public or sensitive data; and provide a full view of the data, software, and stock graphics available for use by AOPC personnel at large.

The most important piece of information in the catalog is the identity of the Data Steward. Each included dataset will be assigned a Data Steward by the department's director. The Data Steward is the sole

contact for accessing the dataset(s) assigned. A rule of thumb for assigning a Data Steward is to choose the last person that manipulates or finalizes the dataset, even if it flows through multiple hands. For example:

At the end of each year, IT program analysts extract the annual criminal caseload statistics using data definitions provided by the Research Department. The data are given to the Research Department, which in turn shares them with each District Court Administrator and Clerk of Courts for corrections. The final numbers are combined with other datasets, calculated into different case management measures, and edited for publication. Among all of the contributors: the county court personnel, the IT Department and the Research Department. Research is the last to "touch" the data and publish final numbers. Accordingly, in this example someone in the Research Department would be assigned as the Data Steward despite multiple contributors along the way.

Maintenance of the assigned dataset is assumed to be part of the assigned Data Steward's existing duties. The Data Steward will be required to document information about the dataset in the Data Catalog, and thereafter update the information when needed. Examples of information detailed in the Data Catalog by the Data Steward include how often the data are updated, whether the data set is public or non-public, what the sources of the data are, and what reports are generated from the data. This information appears in a standardized Data Catalog template.

Information from the AOPC case management systems are routinely requested by Finance, Research, Communications and the public. It is important that those results match what the program or county sees in its case management reports and in annual reports. The Data Steward insures this by using the same source of information for all requests and having subject matter experts involved in the process of disseminating it.

## DATA QUALITY STANDARDS

A common misperception about court data is that they are a by-product of case management systems, and produced simply via a system download. In fact, "raw" data are aptly described as disorganized and unfriendly<sup>1</sup>, characterized by omissions, human errors, system errors, and outliers. They are a series of 1s and 0s that humans would not be able to read (and nor would they want to). Worse, they may appear as an endless list of free-form text entries that offer no readily discernable patterns.

Accordingly, data have to be made *fit for use*, a conventional phrase used by data specialists to describe data that have been corrected, supplemented, verified, or otherwise processed so that they are sufficient for calculation, analysis and interpretation.

Data that are not *fit for use* present considerable risk. The U.S. Government

Accounting Office recommends that a Data Steward consider whether the data:

- may be used to inform legislation, policy, or a program that could have a substantial effect;
- may be used to inform important decisions by individuals or organizations with an interest in the subject;
- will be the basis for numbers that are likely to be widely quoted;
- are relevant to a sensitive or controversial subject; or
- have been evaluated for their quality by experts or external stakeholders.<sup>2</sup>

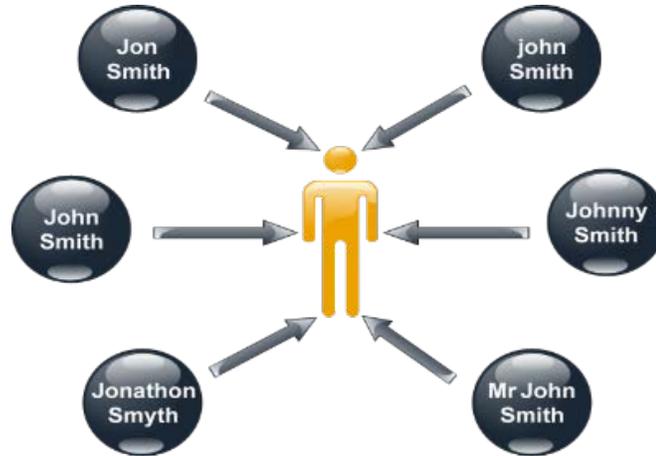
The AOPC policy recommends five categories of assessment to determine whether or not data are *fit for use*:

1. **Completeness.** If more than 10% of the data are missing, computations will be inaccurate resulting in misleading conclusions.
2. **Reliability.** Not all data elements are entered the same way by system users. For example, consider a system field where a clerk is to enter a defendant's name. Even the slightest variations in the format of this field can result in one individual being identified as several different people, or the same person with multiple records. See figure 1.
3. **Comparability.** The judiciary and its systems are constantly evolving. From the

<sup>1</sup> <https://www.dqglobal.com/2014/05/27/what-is-the-difference-between-data-and-information/>

<sup>2</sup> U.S. Government Accountability Office, *Assessing Data Reliability*, GAO-20-283G (Washington, D.C.) December 2019.

Figure 1. Data are not *reliable* if the same item is entered in different ways. Data are made *fit for use* after a data specialist identifies and corrects these discrepancies.



moment we capture information, our data begin to erode as the processes, systems, statutes and rules they were created to capture change. The more time that passes, the more disparities accumulate, rendering the data less coherent year over year. This is commonly referred to as *data decay*.

For this reason, among others, both the Public Access request procedures and the Data Governance policies assert that data requests are limited to a 10-year period.

4. **Timeliness.** Related to *Comparability*, Timeliness refers to the age of the data and how long it remains useful. Depending on how the data are being used, yesterday's data eventually become irrelevant to today's decision-making. A rule of thumb is "getting the right data, to the right people, **at the right time.**"

5. **Accuracy.** Data collected across multiple systems include typos, incomplete information, duplicate records, outliers and other inconsistencies.

Together, these five domains constitute measures of data quality, but they are not

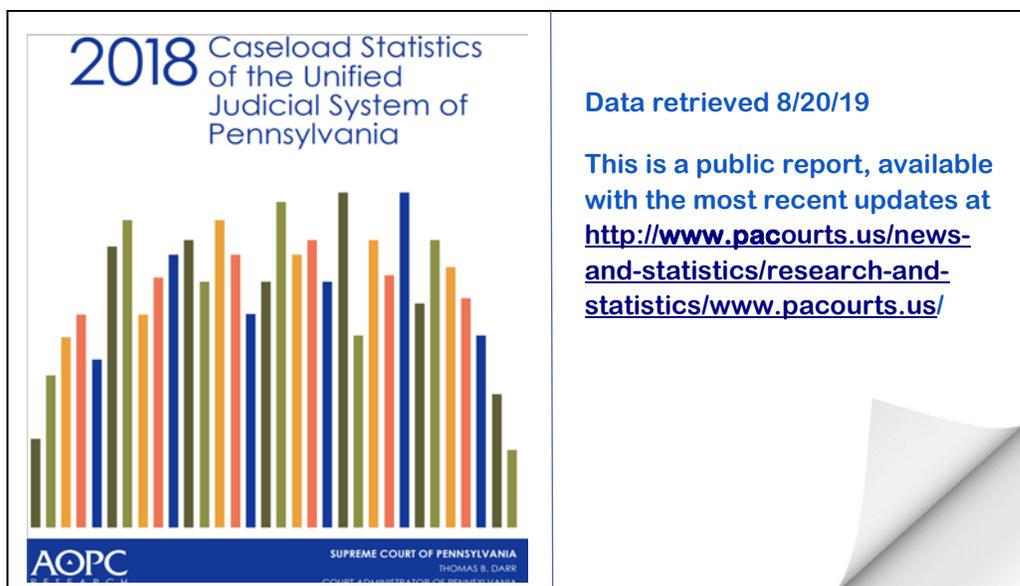
exhaustive; unfortunately there are many more ways data can go wrong! The Data Steward will be familiar with the strengths and weaknesses of a dataset and guide data users through appropriate data applications.

## REPORT STANDARDS AND FORMATTING REQUIREMENTS

Once data are disseminated, they travel distances, change hands and persist long after they're relevant. A standard system of citation is intended to brand official AOPC statistics as they are used by both outside users and AOPC users.

### Citations

When datasets are refreshed annually, like the caseload statistics and the collections and disbursement data, there is always the risk of the re-emergence of outdated versions simply because they are poorly labeled or not labeled at all. Citations serve



as a time-stamp on figures and specify what period they reflect. The Data Governance Policy requires standard language for the citation of figures in reports, graphics and dashboards that appear in print or online. The policy also specifies a location for the citation so checking references is made a consistent and user friendly practice. Moreover, a specific citation to accompany each dataset or graphic is provided for explicitly in the Data Catalog under "Suggested Citation".

In other instances, data are the product of a one-time study or data collection. For example, a workgroup conducted a statewide survey of domestic relations practices to inform policy recommendations. The data were the result of an intensive work effort and of broad interest to a range of government agencies. As the dataset is the sole source of detailed information, never collected before and never since, it's often cited and repeated by speakers, journalists and agencies. But as it ages, the "data decay" render the facts and figures

an outdated reflection of current practices. These AOPC data should travel with a standard citation. While not a hard and fast expiration date, the citation specifies the data as a point-in-time measurement.

### Public and Non-public reports

The terms public and non-public mean different things to different employees. For some, "the public" refers to ordinary people in the community that read newspapers, watch television and otherwise might encounter our data in the media. For others, "public" means not confidential, a report or a graph that can be passed around among departments or shared with a colleague at another government agency. Or, the word "public" can be interpreted as associated with the UJS Public Access Policy which defines whether the data are confidential and, if not, is required to be made available.

The definitions listed in the Data Governance Policy specifies that **public** means data that are available on the website or incorporated in a report or other

externally disseminated document, and comport with the UJS Public Access Policy.

**Non-public** means data that are intended for internal use only, and are not openly available on the website or incorporated in a report or other externally disseminated document. The non-public designation is made by the Data Steward based on the data quality, whether the underlying information comports with the UJS Public Access Policy, and whether the data format qualifies as an AOPC work product.



## REQUESTS FOR DATA

There is considerable time and effort involved in fulfilling an external or internal data request. The Data Steward will serve as the conduit for any data requests and in some respects, there will be variability in how data requests are fulfilled. However, the policy does standardize several data request practices.

### **Time for responding to requests**

The policy sets a 10-day response deadline for all requests. It may not be feasible to fulfill the request in the 10-day period, but the requestor should be contacted by the Data Steward to discuss the specifics of the request within that timeframe. This is consistent with the practice of the Data Exchange when fulfilling external requests.

### **10-year time frame limit**

The Data Quality Standards section describes the repercussions of data decay as data age. In addition, data fields and processing rules change over time. Older data are more labor intensive to produce and may need to be specifically updated or mapped to allow its inclusion in present day data requests. The 10-year limit would mitigate these burdens and is consistent with the limits placed on requests for electronic case records and bulk data requests. Exceptions to the limit will only be made at the direction of the Supreme Court or Court Administrator.

### **Facilitating external data collection**

Often academics, advocacy groups, outside agencies or other external agencies will seek assistance from AOPC to gain access to judges, parties, court personnel or court proceedings in order to conduct interviews, distribute surveys, observe operations or other research activity. The Data Governance Policy prohibits AOPC personnel from facilitating those requests by independent entities.

Pa.R.J.A. 505(a) sets forth data collection as a general function among the duties of the AOPC, and Pa.R.J.A. 506(a) requires cooperation with AOPC requests for information and statistical data. In order to preserve the mandatory cooperation among judicial districts, row offices and court personnel with Supreme Court and AOPC information initiatives, it is important not to co-mingle independent and non-required requests for information with AOPC directives.

There have been instances where independent entities such as the National Center for State Courts, the Pennsylvania or American Bar Association, the Sentencing Commission, or the Pennsylvania Conference of State Trial Judges have suggested surveys or information-gathering initiatives that would be of benefit to AOPC or the Supreme Court. In those instances, requests are to be forwarded to the Data Governance Committee to make recommendation to the Court Administrator for authorization.

## DATA ARCHIVING AND PURGING

The policy does not set specific retention requirements for all datasets. Rather it requires that all data sets publish a schedule of archiving and purging as determined by the department director and the Data Steward. The retention schedule is published in the Data Catalog alongside the data update cycle. When considering an archiving and purging protocol the following considerations are helpful in determining an appropriate schedule:

- are there rules or statutes that require the destruction of the paper or electronic records which are represented in the data?
- are the data sources deleted electronically, hard copies are destroyed, or both?
- what are the specific data decay concerns that might drive the archiving of aging data?

- how must the data be maintained, updated and preserved to respond to requests within the 10-year request timeframe?

## TECHNOLOGY RESOURCE SHARING

It is in the agency's collective interest that all of its work products, regardless of the originating department, are of the highest quality. Each department licenses technology and software to its employees for daily tasks and activities. For example, the Research Department licenses survey software that provides many time-saving functions and permits more responses than the free survey software typically used for inquiries; several departments have Tableau trained personnel and licenses to create dashboards and powerful data visualizations; others have licenses for Adobe Pro that can modify PDF documents.

Some of these licenses are transferrable and can be temporarily shared for a specific project task. Others are either not transferrable or require training to use. In these instances, departments and employees can collaborate to use the technology to upgrade the quality of a report, chart, data task or PowerPoint. The policy directs each Director to list the software or technology resources that the agency provides to his or her department in a specified location of the Data Governance SharePoint site. This accessible location will facilitate the sharing of these resources across the agency and make plain the range of products available to AOPC personnel.

## II. Conclusion

*“To improve is to change; to be perfect is to change often.” — Winston S. Churchill*

Because the people, processes, and systems associated with AOPC data are in constant flux, as are the subjects of the data themselves, perfect data production and delivery can never be achieved and should not reasonably be expected. A Data Governance Policy sets forth policies and procedures to meet current needs, and its Committee serves as an ongoing mechanism to monitor and evaluate the inevitable changes that will require a more elastic response.

Data Governance is above all a change in culture, which requires departments to begin to view their data as an agency asset, redefine problems with data as collaborative challenges, and solutions to those challenges an agency-wide responsibility.