



DATA GOVERNANCE POLICY

Introduction

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 AOPC data.

Scope

This policy covers all data collected, compiled and maintained by the departments and personnel of the Administrative Office of Pennsylvania Courts for both internal and external business purposes, including 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.

External requests for data from CPCMS, MDJS or PACMS are to be referred to the Data Exchange Unit using the request forms available at pacourts.us.

Public and non-public datasets are included. Non-public datasets include work products, such as the results of data analyses conducted by AOPC personnel and databases compiled into work products by AOPC personnel from otherwise public, unverified source data.

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

Definitions

Archiving means moving data no longer used to a separate storage device for long-term retention. Archived data consists of older data that, although no longer used in analysis, remains important to the organization and/or needs to be kept for future reference.

CPCMS means the Common Pleas Criminal Court Case Management System.



Data means

1. Unverified data: information extracted from an automated system or otherwise collected by AOPC from another agency or judicial district that has not been checked for errors, verified for accuracy, or coded for consistency, or otherwise processed for business purposes. Because it has not been cleaned of errors, inconsistencies and text fields, it is not suitable for analysis.
2. Verified data: information extracted from an automated system or collected by AOPC from another agency or judicial district that has been checked for errors, verified for accuracy and prepared by AOPC personnel. Verified data is stored in a spreadsheet and because it has been cleaned of errors, inconsistencies and text fields, is suitable for analysis.
3. Aggregate data: are summary statistics used for reporting, such as statewide or county totals. Percentages, averages and time calculations, such as the age of pending cases, are all aggregate data examples.

Data Steward means the AOPC designee assigned to maintain, update and make available a particular dataset(s).

External Requests refers to requests from any agencies or personnel who are not employed in one of the departments of the AOPC. External requests include those from:

1. judicial district officials and employees, including employees of the office of the clerk of courts, prothonotary, and any other office performing similar functions;
2. the Courts of Common Pleas and Magisterial District Judges;
3. individuals or entities, private or governmental, who assist the Unified Judicial System or related offices in providing court services; and
4. any federal, state, or local government agency or an employee or official of such an agency when acting in his/her official capacity.

Internal Requests are requests from employees of the AOPC departments.

IT means the Administrative Office of Pennsylvania Courts Information Technology Department.

MDJS means the Magisterial District Judge Automated System.

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 Public Access Policies, and whether the data format qualifies as an AOPC work product.

PACMS means the Pennsylvania Appellate Court Case Management System.

PAJCIS means the Problem Solving and Juvenile Courts Information System.

Public means data that are available on the website or incorporated in a report or other externally disseminated document, and comport with the Public Access Policies.

Public Access Policies refers to the Electronic Case Record Public Access Policy and the Case Records Public Access Policy of the Unified Judicial System of Pennsylvania.

Purging means permanently erasing and removing data from a storage space.

UJS means the Unified Judicial System of Pennsylvania.

Data Governance Committee

The Data Governance Committee will function as a standing work group, meeting quarterly and as needed, to address new and changing data governance topics, technologies and data demands.

The composition of the standing committee will include:

1. Co-chairs: the Director of Research and Statistics or designee; and the Director of IT or designee;
2. the Director of Judicial District Operations or designee;
3. the Problem Solving Courts Coordinator;
4. the Data Exchange manager;
5. a representative from the Communications Department;
6. an IT designee representing the delinquency and dependency case management systems;



Revised as of 7/06/20

7. an IT designee representing automated caseload statistics;
8. a designee of the office of the Court Administrator;
9. subject matter experts brought on for topic specific initiatives as needed.

The ongoing responsibilities of the committee shall be to:

1. establish and maintain an internal website that lists available agency datasets and how to access them;
2. document the steps involved in producing the various datasets;
3. publish schedules of when datasets are refreshed or updated;
4. create agency wide standards for validity, reliability, accuracy and uniform definitions;
5. identify existing tools, software licenses and personnel skill sets that optimize the use of data and facilitate sharing among departments; and,
6. review and monitor requests for new data and recommend 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 initiative, but can review and make recommendations upon request.

Data Catalog

- Each dataset governed by the Data Governance Policy is documented in an internal, online catalog that describes the contents and properties of the datasets.
- The Data Catalog entry is authored by the Data Steward assigned to that dataset. Standardized catalog content will be determined by the Data Governance Committee, and is subject to revision.
- The Data Catalog is for internal use only. Viewing access will be available to all AOPC employees; Data Stewards will have administrative access to make entries.



Data Stewards

- For each dataset listed in the Data Catalog, a Data Steward is assigned by a department director.
- A complete list of the individuals designated as Data Stewards and secondary contacts will be maintained in the Data Governance internal website, which will be accessible to all AOPC employees.
- The role and responsibilities of a Data Steward include:
 1. serving as the sole contact for accessing the assigned dataset(s);
 2. defining, managing, controlling and preserving the quality of the dataset(s);
 3. monitoring that the most current data are circulated, and that contradictory or inaccurate data are not released;
 4. issuing timely notifications of any revisions, corrections or updates; and,
 5. facilitating the proper interpretation of data in reports and data products, and advising users of the limitations of the data.

Data Quality Standards

- To determine whether a dataset is fit for an intended use, the following criteria should be considered:
 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 recorded the same way or have the same meaning across the judicial districts. Reliability refers to the data elements having the same meaning across different contexts.
 3. **Comparability.** Data formats and definitions are not held constant over time. Changes to counting methods, programming or definitions may render successive datasets incomparable, or unable to be combined or analyzed together.



- 4. Timeliness. Data more than 10 years old can be out of step with more recent changes to data collection methods, changes in statutes, rules or policies that eliminate or significantly redefine the meaning of a particular variable.
- 5. Accuracy. Data collected across multiple systems include typos, incomplete information, duplicate records, outliers and other inconsistencies. If no systematic data cleaning has been performed, the accuracy is unverified.

Report Standards and Formatting Requirements

- The following standard citation is to be used when incorporating data into internally and externally circulated reports, both hardcopy and electronic:

Administrative Office of Pennsylvania Courts, data retrieved MM/DD/YYYY. See _____ for latest updates and revisions.

The citation should appear on the inside, front cover.

- The following standard citation is to be used on all data visualizations and graphical displays, both hardcopy and electronic:

Last updated MM/DD/YYYY

The citation should appear directly below the display.

- All written reports should be classified as public or non-public. Public reports can be shared with external parties and agencies; non-public reports cannot be shared with external parties or agencies without express permission from the Court Administrator. The public or non-public classification should appear on the inside, front cover, alongside the suggested citation.
- Any reports that include non-public data must include in the header section the following notation: **NON-PUBLIC: NOT FOR CITATION OR CIRCULATION.** Internal reports may be considered NON-PUBLIC work products though they may contain information allowable under the Public Access Policy. In such instances, the data is to be separately requested using the Public Access request forms available at pacourts.us.



Revised as of 7/06/20

- Data visualizations intended for posting to the AOPC website are to incorporate the color palette and font conventions outlined in the AOPC Style Guide, available on the Data Governance internal website.
- Graphics to be posted on the UJS website will be reviewed by Communications for adherence to style guide specifications.

Requests for Data

- External requests for data from CPCMS, MDJS or PACMS are to be referred to the Data Exchange Unit using the request forms available at pacourts.us. Internal requests for system data are to be directed to the Data Exchange Manager.
- Internal and external requests for all other data are to be made to the Data Steward.
- Requests for data shall be responded to by the Data Steward within 10 business days of receipt of a data request, in one of the following ways:
 1. fulfill the request;
 2. notify the requestor in writing or via email that the information cannot be provided; or
 3. notify the requestor in writing or via email that the request has been received and the expected date that the information will be available. If the information will not be available within 30 business days, the Data Steward shall notify the department director.

Time to fulfillment of requests will vary based on the complexity of the request, and will allow sufficient time to review the final data for quality before it is released.

- The Data Governance Policy prohibits AOPC personnel from assisting research by independent entities, other than by providing requested data. Examples include academics, advocacy groups, outside agencies or other external agencies seeking assistance from AOPC employees to gain access to judges, parties, court personnel or court proceedings in order to conduct interviews, distribute surveys, observe operations or other research activity. In special circumstances, requests to facilitate independent entity research are to be referred to the Data Governance Committee in order to make recommendation to the Court Administrator for authorization.
- Datasets are to be maintained in a shareable and consistent format for a 10-year



period. Requests exceeding the 10-year timeframe are to be authorized by the Court Administrator.

Data Archiving and Purging

- Each dataset included in the Data Catalog will have a data life cycle designation, including:
 1. the time frame of available data if less than the 10-year allowance;
 2. an inventory of archived datasets; and,
 3. a purging schedule, if applicable.

Specification of these phases is to be kept current in the Data Catalog.

Technology Resource Sharing

- Where practical and reasonable, departments are encouraged to share technology and software resources across departments and projects.
- The department director will approve which technology and software resources are shareable and included in the Data Catalog Resource Sharing section.
- Approved department technology and software, software descriptions and license holders are to be recorded by the Data Governance Committee on the Data Governance internal website.
- Where possible, software licenses are to be purchased as transferrable.
- All requests for technology sharing from another department, whether a request for another department to collaborate with a service or to borrow or transfer a software license, must be made through the department director.