Florida Criminal Justice Data Transparency

Standard Operating Procedure

Recommended Procedures and Practices for Agencies Complying with the Data Reporting Requirements of Section 900.05, Florida Statutes

MEASURES FOR JUSTICE

This project was supported by a grant awarded by Arnold Ventures to Measures for Justice. The opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the Laura and John Arnold Foundation.

Table of Contents

Table of Figures

List of Tables

Executive Summary

The Florida legislature passed a Criminal Justice Data Transparency (CJDT) bill in 2018, enabling Florida to become a national leader in criminal justice transparency. The bill standardized and centralized case-level criminal justice data from agencies in all 67 counties, requiring criminal justice agencies to submit data to the Florida Department of Law Enforcement (FDLE). Along with other responsibilities, FDLE was required to create a complex data portal available to the public. A "pilot initiative" was established in the Sixth Judicial Circuit, encompassing Pasco and Pinellas Counties. Measures for Justice (MFJ), with the support of Arnold Ventures, was selected to facilitate the project with its data expertise and employment of embedded data fellows. This Standard Operating Procedure (SOP) documents the lessons learned from the pilot initiative and offers recommendations on the steps all agencies required to report to FDLE can take to implement the mandate.

Chapter one explains the business process used to identify and collect data elements. Some data elements were being collected prior to the CJDT initiative while others needed a collection and recording process to be created. Chapter two discusses data element definitions and the schemas needed for data collection. Several meetings with the agencies were held to determine existing data elements, define additional data elements needed, and characterize data schema descriptions with a strong emphasis on standardization. Chapter three clarifies the structure necessary for data extraction. The variety of agencies involved, and their numerous computer systems, made it essential to have a united, automated, and standardized data structuring goal. Chapter four demonstrates the data extraction translation, validation, and transmission process. Detailed translation instructions, validation methodology, business rules, and data formatting are contained within this chapter. Chapter five emphasizes the importance of embracing a data culture within each organization, allowing facts to direct decision-making. Chapter six outlines the value of data visualization and the importance of data publication. Communicating the story to others is easier and more memorable with data visualization.

Measures for Justice believes when there is no data, there is no change. True and accurate data can restore trust in the community and allow government to make informed decisions.[1]

Acknowledgements

This document synthesizes information compiled throughout the three years Measures for Justice offered CJDT implementation support to the criminal justice agencies in Pasco and Pinellas. The work this SOP is based on would not have been possible without the generous support of Arnold Ventures. Furthermore, MFJ is grateful for the input and feedback provided by local and state criminal justice agencies in Florida throughout the course of this grant. Without their expertise and support we would not have been able to create this document.

- Florida Clerk Technology Group
- Florida Court Clerks and Comptrollers Association
- Florida Criminal and Juvenile Justice Information Systems Council
- Florida Department of Corrections, Region III
- Florida Department of Law Enforcement
- Hernando County Clerk of Court
- Hillsborough County Clerk of Court and Comptroller
- Lee County Clerk and Comptroller
- Manatee County Clerk of the Circuit Court and Comptroller
- Marion County Clerk of Court and Comptroller
- Pasco County Clerk and Comptroller
- Pasco Sheriff's Office
- Pinellas County Business Technology Services
- Pinellas County Clerk of the Circuit Court and Comptroller
- Pinellas County Justice Coordination
- Pinellas County Sheriff's Office
- Public Defender, Sixth Judicial Circuit
- Seminole County Clerk of the Circuit Court and Comptroller
- Sixth Judicial Circuit of Florida
- State Attorney, Sixth Judicial Circuit
- Walton County Clerk of Courts and Comptroller

Finally, MFJ would like to acknowledge the hard work and contributions to this project made by our Data Fellows:

- Ariel Fonticiella
- DeSean Lewis
- Paula O'Neil
- Sam Harden

Introduction

During Florida's 2018 legislative session, the legislature passed Senate Bill 1392, making Florida a national leader in criminal justice transparency.[2] The legislation was ambitious in its scope, proposing to standardize and centralize case-level criminal justice data from agencies in 67 different counties. Clerks of the Circuit Court (Clerks), State Attorneys, Public Defenders, Regional Conflict, Jail Administrators, and the Florida Department of Corrections (DOC) were required to submit specific sets of data about each case and defendant. The Florida Department of Law Enforcement (FDLE), Florida's statewide executive branch law enforcement agency, was ordered to collect the data and publish it in "a modern, open, electronic format that [would be] machine readable and readily accessible by the public on the department's website."[3]

The legislation created a *Pilot Project* in the Sixth Judicial Circuit, encompassing Pinellas and Pasco Counties. Both counties received assistance from Measures for Justice (MFJ), a non-partisan, non-profit organization with a mission to bring data transparency to the criminal justice system, at no net cost to the counties. The work product of the Pilot Project was intended to inform and serve as a template for other counties in the state.

Part of MFJ's work with the Pilot Project involved creating a Standard Operating Procedure (SOP) document with guidelines for collecting, standardizing, and transmitting data to the FDLE. The goal was to provide agencies across the state with the tools to comply with the CJDT requirements, including establishing and nurturing a data culture and demonstrating various methods of data visualization. Some chapters were previously released to help the remaining Florida agencies prepare their offices for CJDT compliance.

Chapter 1: Business Process for Data Elements

The first step in developing a comprehensive business process to comply with CJDT involved an agency self-assessment to identify the status of each of the data elements required by the law. During that stage, it is helpful to refer to the following question framework. These questions are not intended to be exhaustive; rather, they serve as the starting point for agency self-assessment.

1.1. Is the data element being collected?

Many of the required data elements were already included in existing data streams, such as the Offender Based Transaction System (OBTS) and the Comprehensive Case Information System (CCIS). The CJDT legislation, however, included some elements and complexity not contemplated by preexisting systems.

1.2. What process could be created to collect data elements?

To develop a process to record data elements not being collected, agencies must determine the data element definitions, the level at which the data elements must be collected, whether they can be collected directly or inferred from other pieces of information, and whether they are being collected in the case management systems (CMS) or if new fields would need to be added.

1.2.1. *Is there a definition for the data element?* Many of the data elements were defined in Section 900.05, F.S., within the Data Dictionaries published by the FDLE, or in both. Although most data elements were defined, edge cases created uncertainty. For example, the FDLE Data Dictionary for the Clerks defined “trial type” as “the type of trial that was held for a case,” and allowed for three value options: None, Jury Trial, or Non-Jury. However, it was unclear how to classify a case where *voir dire* (jury selection) took place, and a jury was empaneled and sworn in, but the defendant pleaded guilty before opening arguments or in the middle of the case.

1.2.2. *What is the level at which the data element must be collected? Is it at the defendant, case, charge, or agency-level?* This question is important in considering where to source each data element. By virtue of their different roles in the criminal justice process, agencies would have either direct or indirect access to information about the defendants, the cases, the charges, and agency-level workload elements. For example, the county

detention facility has direct interaction with many defendants through the arrest and booking process while other agencies typically do not have any direct interaction with defendants. That creates unique challenges for reporting and originating defendant-centric data elements, such as the defendant's primary language.

1.2.2.1. *Data about defendants*: Does the agency interact with the defendant during the case process? If there is direct interaction, what is the protocol when the defendant refuses to divulge information?

1.2.2.2. *Data about cases and charges*: The stage of the charging or court process should be considered for these data elements, as that would affect the collection process. For example, the domestic violence flag element can only be evaluated once charges are filed, while the sentencing elements can only be evaluated in cases where sentencing has occurred.

1.2.2.3. *Workload elements*: The reporting requirements for agencies like the State Attorney's office included many workload data elements, such as the number of full-time prosecutors or the annual felony caseload. Some workload elements required unique extraction processes and some calculation to report accurately.

1.2.3. *Does the agency collect the data element directly, or does it need to be inferred from indirect indicators?* As stated above, some agencies do not have a regular, direct interaction with the original sources of some data elements. For those data elements, it is important to consider whether the element could be inferred from other data sources with a high degree of confidence. As an example, the Clerks must report the primary language of each defendant; however, the Clerks do not typically interact with each defendant to gather that information. Many Clerks record when the defendant requires an in-court language interpreter, which could be considered a reliable indicator that the defendant's primary language is not English.

1.2.4. *Is there a place to store the data elements in the case management systems (CMS)?* Many agencies' systems have capacity for the storage of additional data elements that were not already collected by those agencies. However, each agency has to consider the workload in *both*

inputting *and* extracting the information for reporting. While a free-text note field might be a convenient place to input a data element, it would increase staff workload for extraction and automation.

1.2.4.1. *If there is no place to store the data element in the CMS, should the agency contract its vendor to modify the CMS?* All vendor upgrades or modifications to agency systems can be expensive and time-consuming. Each agency needs to weigh the pros and cons of vendor vs. in-house solutions. In 2019, the Florida Legislature authorized grant funding, managed by FDLE, to help reporting agencies offset or reduce upgrade costs. In cases where vendors represent multiple counties, agencies could consider coordinating efforts across counties.

1.2.4.2. *Can the agency find an in-house solution to extract, translate, store, and transmit the data elements while ensuring consistent data collection?* An in-house solution could include the development of a custom-coded tool that performs all or part of the following processes: extracting data from the agency's database, translating the data into the format required by FDLE, and storing it, if necessary, until the time for transmission. The viability of in-house solutions would depend on multiple factors, including, but not limited to, staff time, funding, the agency's technical expertise, and the complexity of the agency's technical infrastructure.

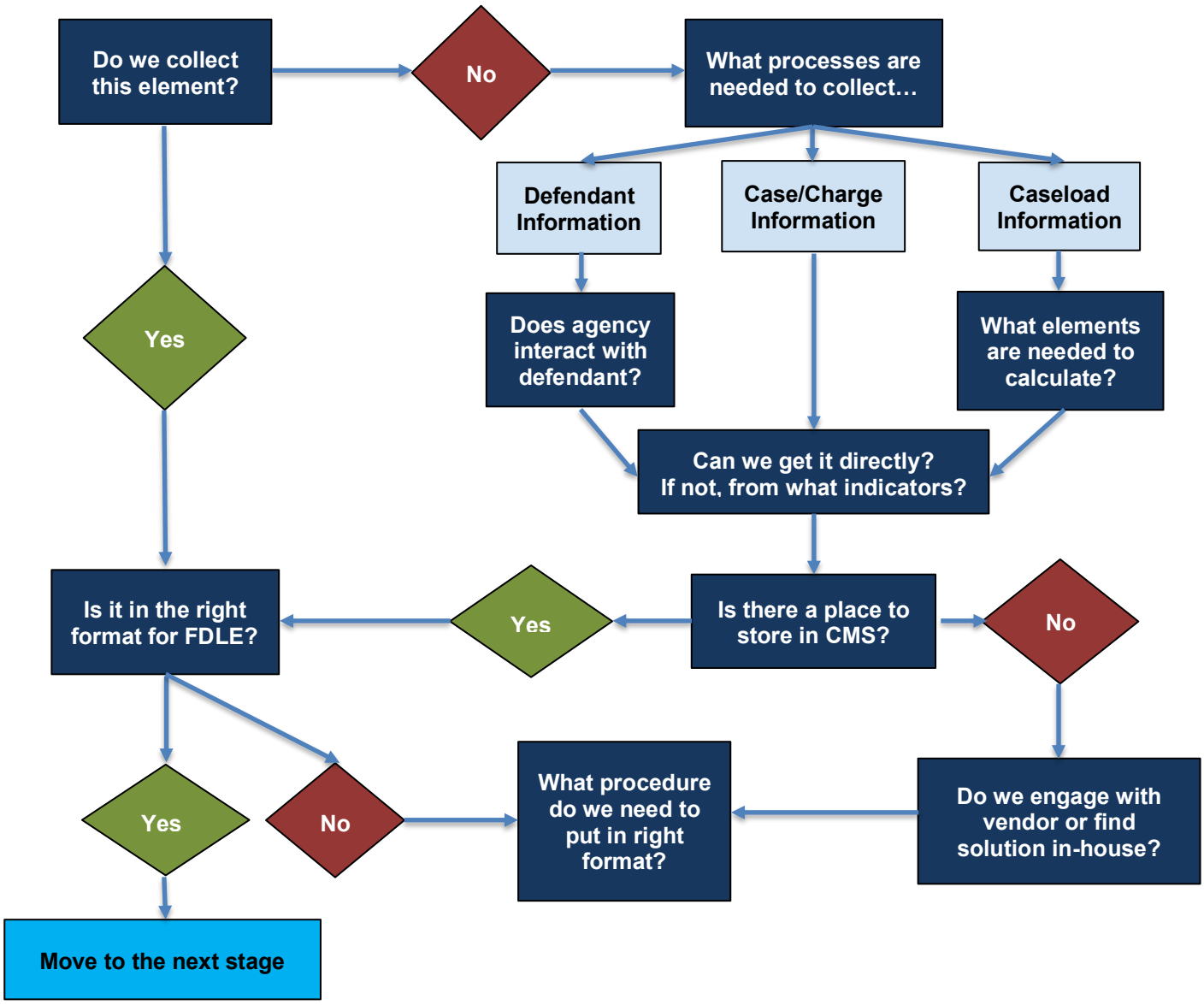
1.3. Are data elements in the required format for transmission to FDLE?

While many of the data elements were being reported through OBTS and CCIS, those reporting mechanisms used a different data format than the one specified by FLDE for CJDT reporting. Each agency needs to assess its capacity for data translation into the NIEM XML format for CJDT reporting and create a process to conform with that format. Further discussion about NIEM XML and its use is discussed in **Chapter 4: Data Extraction Translation & Validation**.

1.4. Self-Assessment Flowchart

Figure 1 illustrates the process each agency should undertake to assess its current ability to collect and report each of the CJDT data elements to FDLE.

Figure 1. Agency Self-Assessment Flowchart



Chapter 2: Data Element Definitions & Schemas

Each pilot county was tasked with creating uniformity in the data being collected from their county's criminal justice system to effectively track statistical data on criminal case processing. This chapter offers guidelines to achieve uniformity in the collection of criminal justice data on a county-by-county basis. It also offers a process to reach consensus around data element definitions, as well as the data structure needed to meet the mandate. This is not intended to be exhaustive, and the process will likely vary from county to county. However, it may serve as a starting point for any county to self-assess its own data collection process and begin preparing solutions for any foreseeable issues that may arise as the project continues to progress.

2.1. Individual Agency Meetings

To create an integrated system that effectively tracks statistical data from criminal case processing, it is important to understand what system each agency uses for its data collection and how those systems transmit data between agencies. That involves a review of how each required data element is defined and collected. In the pilot counties, meetings were held as soon as the legislation passed to discuss definitions and sources. Data fellows joined the meetings as soon as they were hired. Through that process, each agency identified known problem areas to avoid future complications and began crafting solutions. Below is a non-exhaustive list of questions that the agencies used to identify potential issues and solutions.

2.1.1. *What system does the agency use? Is it integrated or non-integrated?* For example, Pinellas County uses a centralized, integrated system that shares a common database between different justice domain agencies, and that is managed and maintained by the county's Business & Technology Services team ("BTS"). A data element such as "defendant's date of birth" is shared as a common database entry between multiple agencies (the offices of the Clerk, State Attorney, County Detention Center, and Public Defender). In practical terms, that means that rather than each agency's technology department reporting its required data elements separately, the county's BTS team will report the data on behalf of the Pinellas agencies from the centralized database source.

2.1.2. *Which data elements exist within the agency's system?* There are two issues to consider when addressing this question. First, which data elements are collected and stored within the agency's database?

Secondly, is the agency's database configured to store the data element? It was discovered that some data elements required by the legislation were not present in some agency systems, specifically, citizenship, primary language, and flag-designation elements.

2.1.3. *Which agency is the most appropriate source for collecting specific elements?* Considerations for addressing this question include whether the agency originated a particular data element, or whether the agency was closest to the source of the data element. One example discovered was the data element for defendant citizenship, as it was reported by both the Clerk and the County Detention Center. The Detention Center, however, is better equipped to source that information because it has direct access to the defendant and access to federal immigration databases and agencies.

2.2. Data Element Definition Workshops

Workshops were conducted with all partner agencies in the two counties to agree on uniform definitions for each data element. The feedback provided during those workshops was aggregated into a single document, which was provided to the Florida Legislature as the basis for the creation of a *clean-up bill*, and to the Florida Department of Law Enforcement, to assist with the data dictionary development process. The feedback addressed substantive issues that were identified by the agencies regarding the language of the original bill. These workshops were successful because the participating agencies were able to agree on unified project goals and worked together to achieve those goals throughout the project's life cycle. Below are some of the bigger takeaways resulting from the data element workshops.

2.2.1. *Statewide Uniformity***:** After speaking with each agency, it was clear that the greatest hurdles were the lack of uniformity in data collection practices and a lack of clarity in the definitions of data elements required by CJDT. All stakeholders benefitted from a consensus on the required data elements, baseline functionality, and workflow for each case type. Without uniformity, it would be difficult to accurately evaluate and compare caseload differentials and calculate specific trends and variations from county to county. Improvements in those areas are essential to collecting data that can be meaningfully evaluated and compared across counties. The feedback gathered during the data element definition workshops was

provided to the Florida State Legislature to aid in the creation of a Uniform Statute Table, Uniform Arrest Affidavit, and Uniform Data Definitions.

2.2.2. *Uniform Statute Table*: One of the issues encountered during the pilot were the differences between the State Attorney's Office (SAO) internal statute table and FDLE's statute table, which resulted in the collection of inconsistent data. That data inconsistency could be avoided with a uniform statute table. Those concerns were reported to the legislature with recommendations to amend the original bill and mandate the development of a uniform statute table and required usage by the SAOs across the state. Uniformity in the statute table is crucial to ensure the collection of consistent data.

2.2.3. *Uniform Arrest Affidavit*: For the data transparency initiative to succeed, business processes and the forms used statewide must be standardized before crafting an IT solution. In Florida, there is a Uniform Traffic Citation form, so all locations and law enforcement agencies issue tickets in the same format. However, Florida did not have a uniform arrest affidavit. For example, during the pilot we learned that the jails in Pasco and Pinellas were collecting different information on their inmates, making the data collection inconsistent between the counties. Since there was not a uniform arrest affidavit, agencies created their own internal, arrest affidavit forms. Each arresting agency in Florida collected what its team thought was most important when the forms were created, resulting in inconsistent data regarding incidents, arrests, and arrestees. Those concerns were reported to the legislature, which amended the legislation in 2019 to mandate that the FDLE develop a uniform arrest affidavit to be used by all law enforcement agencies in the state[4].

2.2.4. *Uniform Statewide Definitions*: Although each county was tasked with collecting the same data elements, many of the elements were not defined precisely, requiring agencies to individually interpret those definitions. As a result, inconsistent data has been collected. For example, in one county, gender was reported based on information listed on the individual's driver's license. However, in other counties gender was reported based on how the individual self-identified at the time of arrest. That was due to the lack of a set of uniform statewide definitions, causing inconsistencies in the availability and accuracy of data collected. Without the creation of uniform statewide definitions, data across the state will not be collected or reported uniformly. At a minimum, a common standard or baseline must

be defined in the areas of case processing and data collection for guidance to all counties.

2.2.5. *Scope:* For a data transparency initiative of this magnitude to succeed, there must be a clearly defined scope, from beginning to end. All parties involved must ensure the scope of the project is not painted with too broad a brush. Stakeholders and legislators need to work together to adequately define the scope, functionality, and main objectives for each data element required to be collected. Failure to do so can cause confusion when interpreting the intent for certain elements. For example, Clerks were tasked with the collection of the defendants' *arraignment dates*. The challenge with defining arraignment date scope was determining which date to capture. In some instances, defendants entered a plea to their charges *prior* to their listed arraignment dates, which created questions surrounding the scope and intent for collecting said element. Was it to capture only those cases that fit into the category as written in the bill ("arraignment date") or was it to capture any date in which the defendant was arraigned? Without a clear directive and consensus as to the scope, functionality, and main objectives, those particular elements could be interpreted differently by each agency involved, which was problematic. Furthermore, a project timeline must have interim milestones that can demonstrate progress and utility to everyone involved. This allows for less confusion regarding scope and intent and provides stakeholders and legislators the ability to collaborate effectively.

2.3. Schemas

After defining data elements required by the legislation, it was important to give stakeholders aid in finding and preparing those elements for transmission, i.e., to achieve *data readiness*. Specifically, stakeholders needed a way to prepare their infrastructures for data collection while the transmission specifications were being developed. That was where data schemas came into play.

2.3.1. *Data Schema Definition & Benefits*

Relative to criminal justice data transparency (CJDT), a data schema is a structure representing the logical view of the required data elements. The data schema shows how those data elements can be organized and associated in addition to any constraints applied to the data.

Since extraction was a key part of preparing data for reporting, the data schema serves as a blueprint to help Structured Query Language (SQL) developers design queries and for database administrators (DBA) to create a database. Resulting SQL extractions could be immediately processed into the required format for transmission or housed in a database for further tracking or processing.

The schemas developed for CJDT are legislation specific, and data system and infrastructure agnostic. They are designed as a tool to aid stakeholders in preparing data extractions from their systems using the data schema structure as a guide, allowing them to get one step closer to legislative compliance.

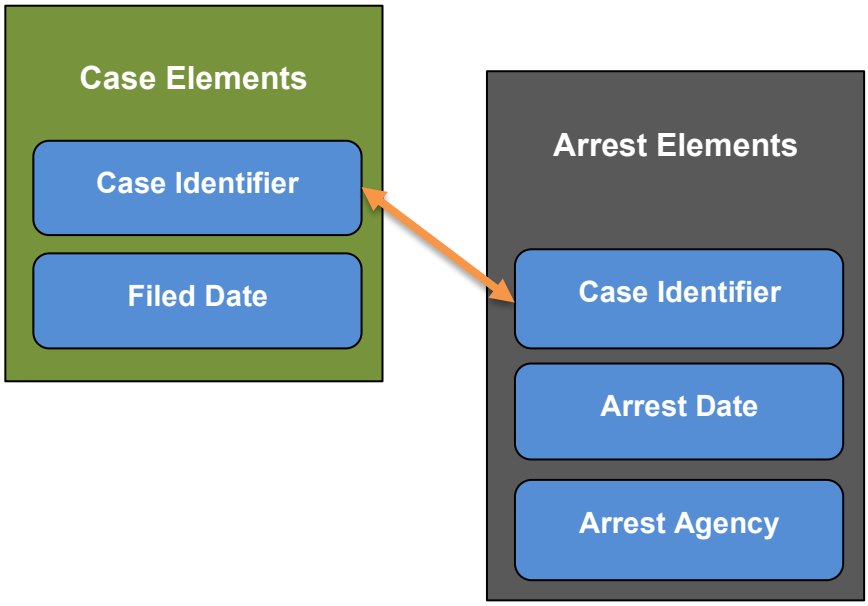
2.3.2. *Process*

The process for creating a data schema can be complex and is ultimately unique to the scenario for which it applies. The steps below are general guidelines for data schema creation and should help direct users in creating their own.

2.3.2.1. *Catalog and define data elements.*
As described earlier in this document, the first step requires a full catalog and corresponding definition for each data element designated for collection, in other words, a data dictionary. The data dictionary is essential for understanding what each data element means (especially if there is the potential for multiple interpretations) and usually provides insight into how those elements are related.

2.3.2.2. *Examine and map logical relationships.*
After the data elements have been defined, some logical linkages are relatively easy to discern based on those definitions and with some level of knowledge of the subject matter. The aim is to try to define logical groups of data and then determine how those groups relate. For example, with regards to criminal case data, information like the case identifier and the date the case was filed would logically be grouped together. Related data, like arrest information including the arrest date, the arresting agency, etc. could be grouped together and then related to an associated case by the case identifier. This example is illustrated in **Figure 2** below:

Figure 2. Relationships of Elements



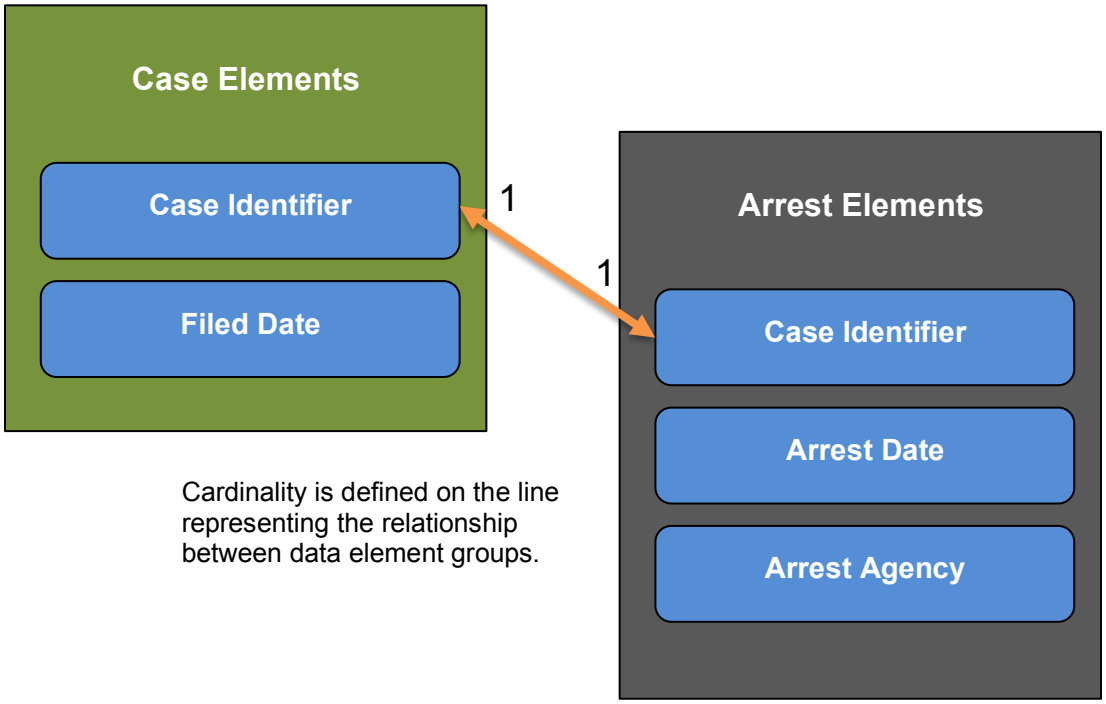
It is important to note that there are multiple ways to aggregate data and define how they relate. The key is to create a data schema structure that makes sense for the scenario and to the parties involved.

2.3.2.3. *Define relationship constraints: Cardinality & Ordinality.*
After elements are grouped and relationships are identified, it is helpful to then define their cardinality and ordinality. The emphasis at this stage is on quantifying the relationship between groups, different from grouping of the elements as defined in the previous step.

Cardinality is the maximum number of times an instance of one data element group can be related to instances of another group. Cardinality is usually expressed as being one-to-one (1-1), many-to-one (1-M), or many-to-many (M-M).

Continuing with the same example, one might define the cardinality of the relationship between cases and arrests as one-to-one; meaning, for every individual case, there is one corresponding arrest. The example is illustrated in **Figure 3.**

Figure 3 Cardinality Relationship

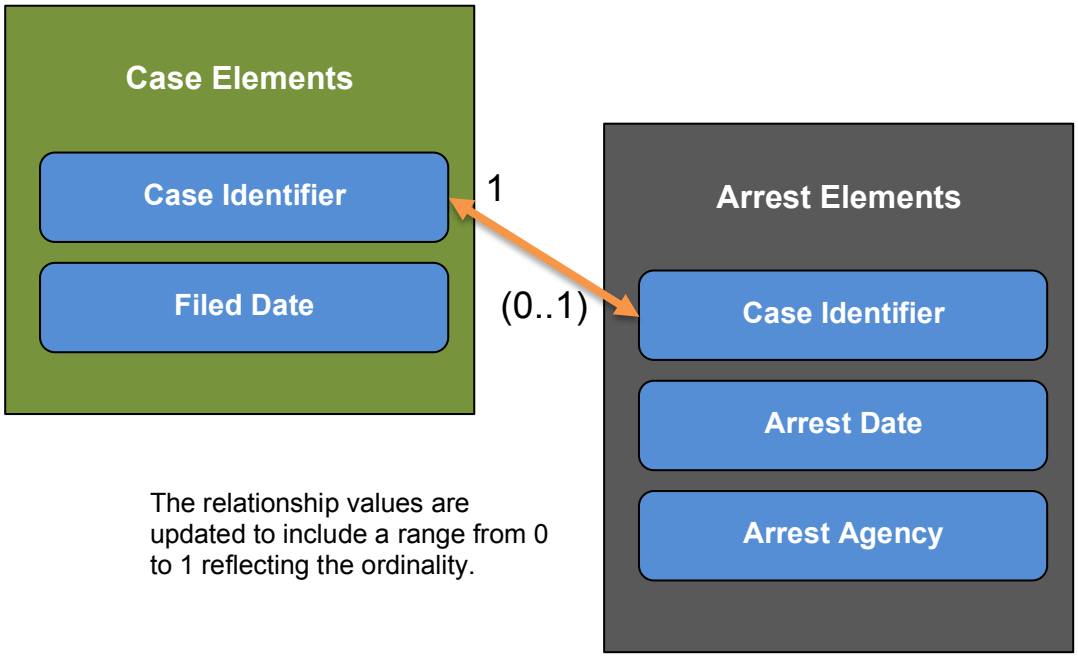


In addition to cardinality, data schemas can also be used to show a relationship's ordinality.

Ordinality is the minimum number of times an instance of one data element group can be related to an instance of another group: the converse of cardinality. In effect, it expresses whether an element group is optional or mandatory relative to another group.

Again, using the example, depending on the severity of an offense, a defendant could be issued a notice to appear (NTA) instead of being arrested. To model this scenario, combining cardinality and ordinality, one might define the relationship between cases and arrests as one-to-one, but also optional. The arrest element group is therefore dependent upon whether a defendant was arrested. This expanded example is illustrated in **Figure 4.**

Figure 4 Relationship Values



Combining cardinality and ordinality provides the flexibility to express a variety of relationships between data elements and their groups. When creating a data schema, use what best fits your scenario. Extensive examples can be found online. Some additional relationships are defined in **Table 1** below for reference.

Table 1 Relationship Definitions

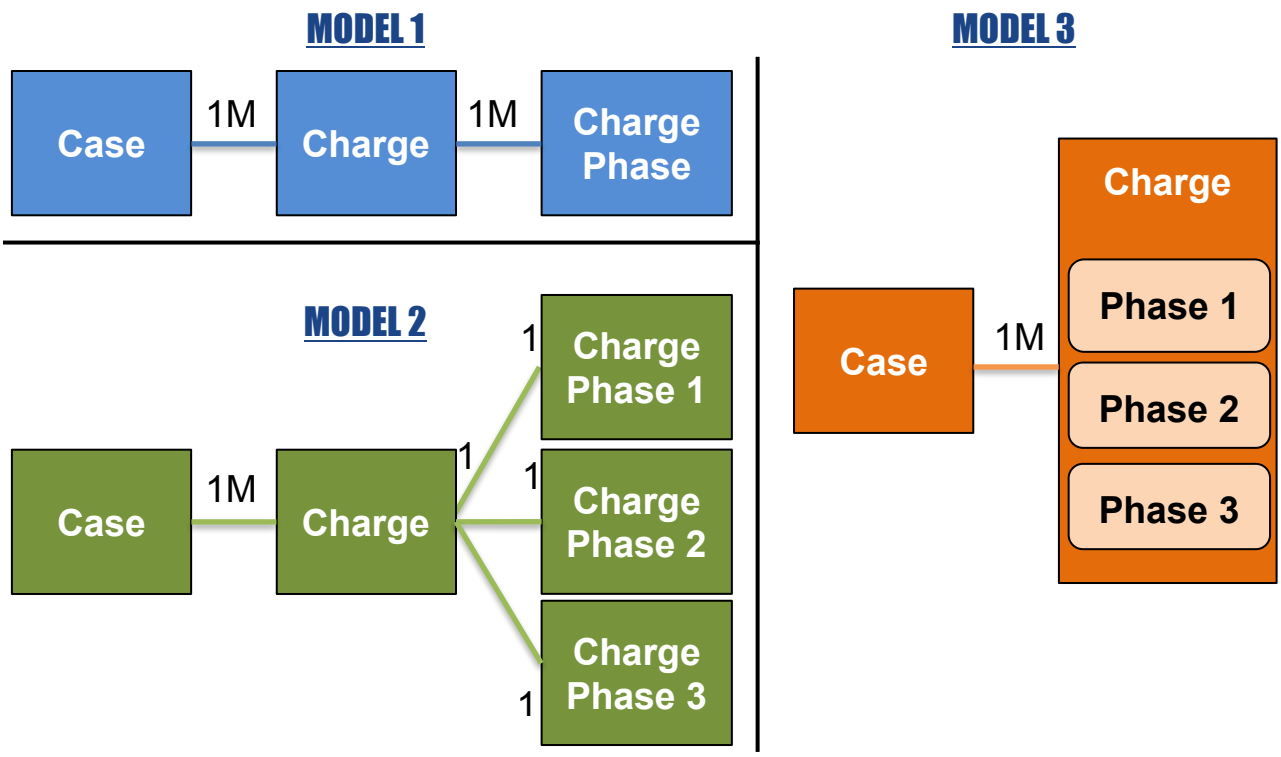
Relationship	Expression	Description
One-to-many, required	(1..M)	Indicates that a group of elements is required to have at least one instance but may have more related to another group.
Zero-to-many, optional	(0..M)	Indicates that a group of elements is optional but can have many instances related to another group.

2.3.2.4. *Resolve complicated data relationships.*
Some data relationships are less obvious and may require the aid

of one or more Subject Matter Experts (SME). An SME can provide valuable insight into practical aspects of how a data element is collected, how that element is stored and tracked in their systems, or what the data element means in the context of the criminal justice process. Understanding the purpose of the data can help identify the most effective ways to group data elements and draw relationships between groups.

A more complicated example relative to criminal case data could be expressing the relationship between cases, charges, and charge phases. This use case is interesting because it can be correctly expressed in multiple ways. A few examples are shown in **Figure 5** below.

Figure 5 Complicated Data Relationships



This is where working with an SME can help clarify which design is most appropriate given the project scenario.

2.3.2.5. *Iteration & Feedback*

Continue cycling through steps 2.3.2.2 to 2.3.2.4 until all elements have been grouped and associated. This exercise should result in a

network of connected data element groups with defined relationships. Throughout the process, be sure to solicit feedback from potential stakeholders and SMEs to help inform and mold the data schema structure. The goal is to create a logical view of the data required that is specific, flexible, and consistent with the project scope.

Florida Statute 900.05 outlines the data elements required by CJDT. Each agency type should develop its own schema that reflects its individual and unique data responsibilities.

2.3.3. *Practical Use*

As mentioned earlier, a completed data schema can serve as a blueprint for both SQL developers and DBAs.

SQL developers can use the data schema to create code that extracts data from their systems in a structure consistent with what was required for reporting. The schema groupings and associations provide valuable insight for developing queries and subqueries.

A DBA can use the data schema to create a separate database for housing the required data. The data required for each element could then be populated on a set schedule, independent of a core/production system, and queried when needed for transmission.

Additionally, if a data schema was developed early in the project process, it can be used to re-inform that process. Considering again the questions a data collection project is attempting to answer, the high-level view provided by a data schema can serve as a tool to validate project scope. Several validation examples include:

- Providing visibility to gaps in data collection where additional elements may need to be collected.
- Showing opportunities with data collection where the addition of a few elements may result in a more valuable dataset.
- Exposing a need to reduce collected data elements due to constraints or limitations with current ability for collection.

2.4. Chapter 2 Summary

After workshopping with agencies in the pilot counties, we found a need for

improvement in the definitions of required data elements and a need for standardization in data collection practices across the state. Feedback to the state was provided. Meanwhile, agencies can begin by addressing basic questions about the location of required data elements and whether another agency is better equipped to provide that data. Schema creation is an important step in developing a database that can transmit the required data elements to FDLE.

Chapter 3: Structuring for Data Extractions

Once the work of developing and defining data elements for reporting has been established, stakeholders must begin structuring their data for extraction. This is where data collection goals meet practical application.

To review, the business processes (from **Chapter 1: Business Process for Data Elements**), applied on an individual agency basis, in combination with the data element definitions and data schemas (from **Chapter 2: Data Element Definitions & Schemas**), inform the creation of a data extraction foundation.

The business processes provide a higher-level view of what information is collected. The definitions and schemas provide the blueprint for what information is needed. Studying the intersection of what is collected versus what is needed, defines the direction for achieving reporting compliance for an agency. Where the intersection overlaps, data elements should be reviewed for accuracy. Where there are no overlaps, known as data gaps (i.e., data not collected), an agency's processes and infrastructure should be reviewed for solutions. When all required data elements are reviewed and all data gaps are remedied, there is a solid foundation for data extraction.

This chapter serves as a guide on general concepts, methods, and tools for structuring the data to build that foundation. It also provides illustrations from circumstances experienced during the CJDT Pilot Initiative.

3.1 Data Structuring Goal

Data structuring, in the context of CJDT reporting, is the process(es) by which data are captured from an agency's CMS and conformed to comply with the provided data specification. After the data are extracted, they can be translated for transmission, a subject that will be addressed in the next chapter.

The success of data structuring for extraction is measured based on two factors:

- *Data Accuracy* or the extent to which data captured matches the definition and format provided in the specification.
- *Data Completeness* or the existence of data, in its entirety, as defined by the same data element specification.

Accordingly, an effective extraction is wholly dependent upon the data within an agency's CMS, the processes by which those data are populated and the ability

to format those data appropriately. Thus, the goal of data structuring is to capture a complete and accurate subset of data consistent with the reporting specification.

3.2. Firming the Foundation

The concepts below are a review of what was discussed in earlier chapters. To achieve our extraction goal, we need a solid foundation (i.e., complete and accurate data). To establish this, an agency can ask the following questions:

3.2.1. *Does the agency already collect this data element? (from* ***Chapter 1.1*** **Is the data element being collected?***)*

3.2.1.1. *Survey captured elements.*
Using the data element definitions and specifications provided, an agency should catalog the data within their CMS. This includes, but is not limited to backend information like associated databases, field names, data types and formats, data relationships and connections, etc. The goal is to document all relevant information that helps validate whether the data stored is what is required.

3.2.1.2. *Identify and classify missing or incomplete elements.*
Once the survey of elements is complete, one should have a good understanding of the data elements that are not captured by the exact definitions in the specification. To continue the analogy, these two categorizations in the survey are the holes in our foundation or the data gaps.

3.2.2. *What process can be created to collect this data element? (from* ***Chapter 1:2*** **What process could be created to collect data elements?***)*

3.2.2.1. *Create or update processes to capture missing data.*
This step is fundamental to establishing a stable foundation for extraction. This is where current processes are modified and/or created to fill those data gaps. This can also vary in complexity depending upon an agency's CMS and current business processes. Only when all the holes are filled will we have a solid foundation for extraction.

The above bullets make up the conceptual basis for preparing the technology environment for data extraction. The next section covers some of the ways in which these concepts are applied.

3.3. Data Preparation

It should be noted that the primary purpose of a CMS is to help an agency store and track information that facilitates daily operations. Because external reporting goals are not necessarily aligned with those operations, data must often be translated to comply with a reporting specification.

The methods below cover many of the ways in which data are prepared. They are interdependent and not necessarily sequential. Many of them are also common knowledge for technology professionals. The list is a guide, as application will ultimately be determined based on your needs and resources and with input from subject matter experts (SMEs).

3.3.1. Query Data

Querying a database is the best way to learn about the data. Through this mechanism we can create a survey of data elements and determine where and how to access them. This method usually requires knowledge of SQL and a base knowledge (or an ability to ascertain that knowledge) of how your system stores and relates data.

A general data element survey could consist of required element details and their counterparts in a CMS. This format is useful as an overall tracker of element analysis and as a communication tool across an agency. **Table 2** shows an example where the columns in blue describe required data elements and the columns in green describe their counterparts in a hypothetical CMS.

Table 2 Element to CMS Example

Element	Type	Description	CMS Table	CMS Field	CMS Field Type
Case Number	alphanumeric	The UCN...	case	caseNum	varchar
Filed Date	date	The date...	case	filedDate	date

In the pilot counties, similar spreadsheets were created and shared with team members to help find and verify elements. In some versions of these sheets, columns were expanded to include corresponding statute line numbers, data origin (or source), formatting, and discussion notes.

The pilot counties also used queries to perform raw data dumps (into spreadsheets) of field values to give teams a sense of the data stored, its format and completeness.

3.3.2. Validating Data & Auditing Collection Processes

This process is critical to determining the relevance of the data and finalizing the data element survey. Validation in this context not only applies to an element's value and format, but also its means of collection and structural compliance with what is required for reporting.

3.3.2.1. *Validating data value and format.*

This level of validation is straightforward. Agencies should compare the required data elements and formats with those in their system, making notes where there are discrepancies. The notes should identify if an element exists, whether it’s in the correct format and include details for elements needing resolution. A few examples encountered in the pilot counties are below.

3.3.2.1.1 *Compatible requirements*

Florida Court’s Uniform Case Number (UCN) is an example of a compliant match. All county agencies in the state of Florida are required to use this number to identify cases. The UCN format is also consistent across the state. Among the many benefits of UCN, is the ease of statewide reporting because of its use by all criminal justice agencies. CJDT requires this number for identifying cases, making it compatible with data stored in agency CMSs. Data elements like UCN are ready for extraction.

3.3.2.1.2. *Reconcilable requirements*

The frequent offenders of format violations in the pilot counties are generally date related fields. FDLE’s specification requires dates in the xs:date (NIEM) format which follows the pattern YYYY-MM-DD. The

most common date format in the US is usually some variation of MM-DD-YYYY. In an agency CMS, these fields require relatively minor conversions for extraction.

3.3.2.1.3. *Requirement disparities*

CJDT requires the collection of several data points including a defendant's immigration status, citizenship, ethnicity, primary language, and others that are not native to the pilot county CMSs. These are just a few examples of where validation revealed that there is a disparity and a need for resolution.

3.3.2.2. *Evaluate completeness of data.*

The purpose of this process is to evaluate how often information is populated for an existing data element. As a result, an agency may find that some data fields within their CMS may either not be used or are populated sporadically. Identifying these fields is the first step to redesigning workflows to ensure accurate population of data.

One of the more challenging elements for data collection about CJDT, deals with sentencing conditions. These conditions have a degree of variability and (at least in the pilot counties) are not typically captured and itemized digitally. While the conditions are documented with electronically scanned court forms, most of those conditions do not have corresponding CMS fields. For the minority of conditions that do, itemized recording of those values is unnecessary for operations and therefore populated inconsistently. Collection of this data point may require CMS modifications and workflow process changes.

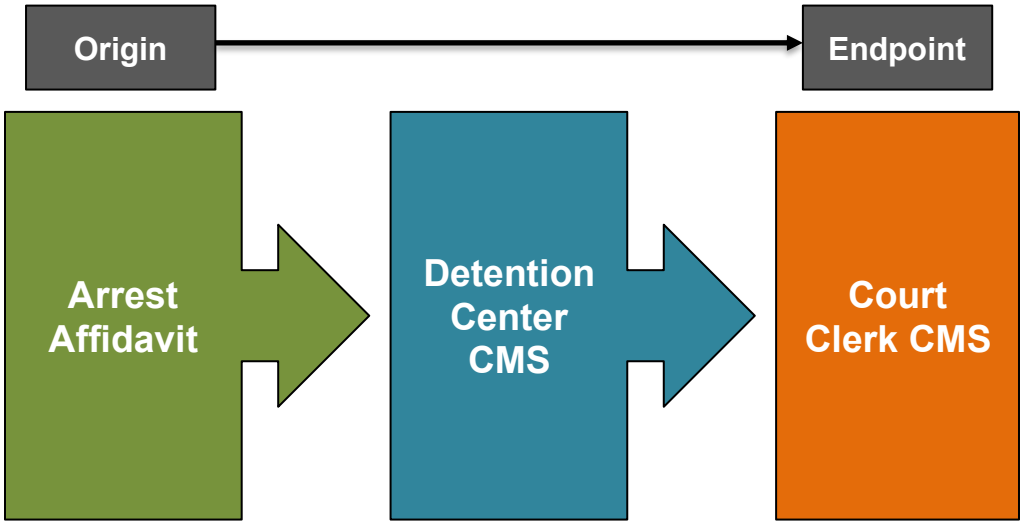
3.3.2.3. *Auditing means of collection.*

Sometimes, to accurately validate a data element, the collection path must be examined from origin to its input into a CMS. The collection process for some elements can be intricate, especially when accounting for aspects like timing and collection intention.

The advantage of this perspective is a complete understanding of a data point and its validity.

In the pilot counties, we found that the collection paths for most data elements engendered confidence in the validity of the data. For example, an arrest date in the Court Clerk system was populated by the information collected at the detention center, which was provided by the arresting officer(s). The chain of data collection was consistent with what CJDT reporting required and was the most accurate way to collect these data. See **Figure 6.**

Figure 6 Arrest Date Collection Path Example



In contrast, another CJDT requirement is to collect whether the primary language of a defendant is English. In the pilot counties, initial research indicated that Court Clerk CMSs captured that information. However, further inspection of the collection path revealed that the field was utilized only when a defendant asked for an interpreter. As a result, this field satisfies CJDT's requirements only if all defendants who do not use English as their primary language ask for an interpreter. Criminal justice personnel confirmed that the assumption was false, and that the data point was not compliant for use with CJDT. To restate, the intention of this field as used is to help agencies track when an interpreter is

needed and not specifically to track the defendant's primary language per CJDT.

3.3.2.4. *Evaluating structural compliance.*

This stage entails an analysis of an agency's data relationships to determine whether they align, or can be aligned, with the reporting specification. Recalling the data schemas discussed in Chapter Two, structural compliance is aided by understanding the cardinality and ordinality of data elements and their groups. The added task is to attempt to mold an agency's extracted data to mimic those relationships.

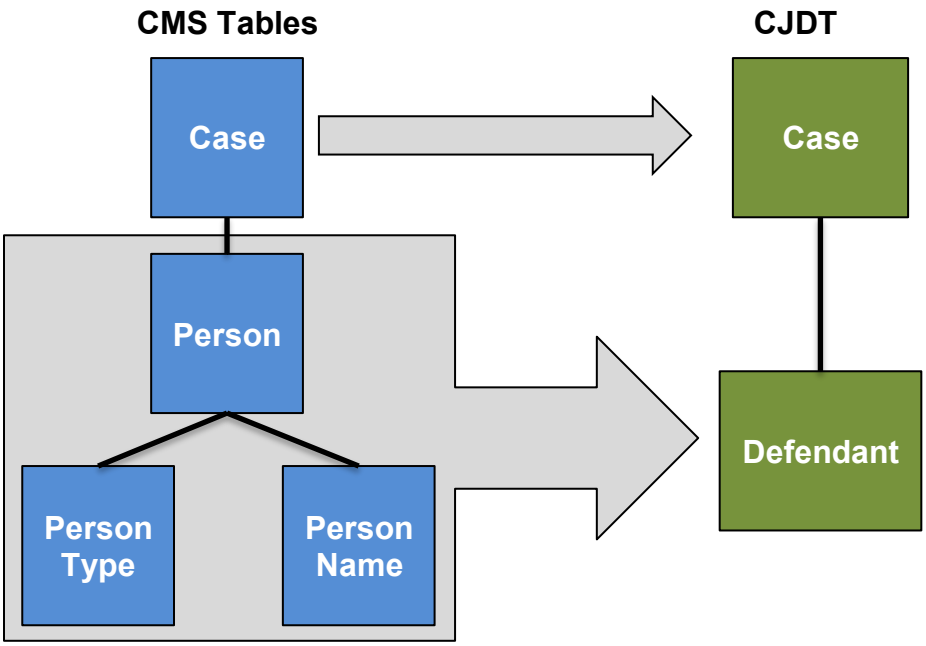
For example, one of the core relationships within CJDT is the connection between cases and defendants. This relationship is direct. However, it is highly likely that an agency's CMS may make that connection differently based on functional necessity.

In one of the pilot counties, the connection required engaging multiple tables, filtered by information related to defendants, because all persons related to a case were stored in the same table and classified using other values. On the surface, the ability to replicate the CJDT requirement using the CMS tables may seem unlikely. However, with careful examination and knowledge of SQL, developers can make the conversion work.
Figure 7 shows a high-level, generic view of this conversion.

The diagram illustrates how, despite the difference in structure, data in the CMS can still be reconfigured to comply. In some cases, this may not be possible and will therefore require more robust resolution using system modifications and/or workarounds.

In regular meetings (including the data element definition workshops), agency stakeholders diligently reviewed each data element applying a mix of the validation methods above. One of the lessons learned was that a universally agreed upon set of definitions was vital to the validation process. While most elements are concretely defined for CJDT, because of the novelty and complexity of this project, the definitions of other elements remain malleable.

Figure 7 Data Conversion



Validation identifies data elements requiring resolution. The methods below are ways in which those resolutions are realized.

3.3.4. Cleaning Data

Data elements that are a contextual match but require alteration will have to be cleaned. Data cleaning specifically refers to modifying an element's format and/or parsing through data values to include or exclude certain attributes. The end goal is to have a value that matches the reporting requirement in content and format. This process is simplified if the validation process is performed first, as the notes from validation can be used to determine the type of cleaning required.

Revisiting the date examples from 3.3.2.1.2., validation identified the format issue and the correct format to use (i.e., converting MM-DD-YYYY to YYYY-MM-DD). Therefore, when extracting a date type element, a developer will need to use code to modify the date appropriately.

A more complex example of data cleaning pertains to translating CMS values to preset code values mandated by the specification. With CJDT, counties provided input for some coded values including bond types, court appearance types, and others. Other code values in the specification were

predetermined, taking cues from existing reporting standards. To reconcile these values, a mapping system between CMS values and reporting values must be created. In the pilot counties current solutions use a mixture of hard code and dynamic tables depending upon the number of values being mapped.

3.3.5. Data Calculations

Some data elements are not directly available in an agency's CMS but could be derived via logical deduction or mathematical calculation. The element definitions should define the necessary formula. Agencies will have to identify the fields in their CMS needed to perform the calculation.

Regarding CJDT, most of the elements in this category were the aggregated counts required from the State Attorney's Office (SAO), the Public Defender's Office (PDO) and the county detention facility. With the SAO and PDO, CJDT required counts for attorneys and caseloads. For the county detention facilities, the spectrum was broader, including counts with classifications for inmates and personnel.

To ensure accurate calculations, uniform rules must be firmly established, especially for more involved statistics. For example, CJDT required county detention facilities to provide a per diem cost for a jail bed. For that calculation, there are a myriad of factors to consider, including salaries, benefits, supplies, medical costs, food, utilities, inmate programs, just to name a few. In addition, considering that there are multiple county detention centers and varying management arrangements in Florida, it is easy to see how the lack of a standard could lead to inconsistent calculations and inaccurate data.

Fortunately, in Florida, a per diem formula was developed by the Florida Model Jail Standards Committee. Briefly, the formula added annual costs from personnel, operations, and capital; subtracted revenue and divided the result by the number of days in a year. The formula also expounded upon the cost and revenue factors for calculation.

3.3.6. Data Collection Process Changes

This step refers to the specific modification of workflow processes to ensure the proper collection of data in an agency's CMS. Similar to data

cleaning, the validation process makes this step easier by identifying elements needing attention. Adjustments made at this stage may also require database modifications.

3.3.6.1. Minor changes:

Some process changes are relatively minor. For instance, CJDT required the collection of charge modifiers. In one of the pilot counties, we found that the existing CMS's preset values could be expanded to include more of those modifiers. This change also called for modifying data entry processes and retraining staff to use those new values.

3.3.6.2. Automation changes:

Some process changes are programmatic. For example, as part of their specification, FDLE requested additional identification numbers to help connect and validate information transmitted by multiple agencies. One of those is the warrant number requirement for the Clerks. While that information is not stored in their systems, it is stored by the Sheriff. The existing data feed between the Clerk and the detention facility could be programmatically adjusted to import that information from the Sheriff.

3.3.6.3. Complex changes:

One of the more complicated process changes required by CJDT deals with the collection of a defendant's immigration status by the Court Clerks. For the Court Clerks, there was no existing process to capture or validate that type of information. A resolution might require coordination with agencies who deal more directly with defendants and have resources to identify that information. This process might vary by county.

3.3.7. CMS Modifications & Workarounds

For data elements that are not captured in an agency's CMS and have no existing field with which to capture that information, modifications and/or workarounds are required. Modifications may take the form of code changes applied via upgrades or maintenance fixes. Workarounds may comprise external tools or databases used to collect data without having to modify an agency's system.

Depending on the type of CMS: vendor-based vs in-house tool, enterprise vs. medium capacity software, integrated vs. isolated database, etc., the specific modifications and workarounds are unique. Here are a few CJDT examples.

As discussed for charge modifiers, CMS modification means an expansion of the predefined values in the system. This was a relatively minor modification since it usually entailed adding rows to a core CMS table with new values. However, capturing a warrant number and defendant immigration status requires at least the addition of new database fields. Immigration status, because it introduced new predefined values, required a system table and structural changes. Other data elements in this category of complexity included defendant indigent status, ethnicity, and primary language.

Depending on the cost, time and effort of modifications, an agency may decide to use workarounds to access and/or store these data points instead. For instance, hypothetically, if immigration status existed in an external repository, the Court Clerk could import those values to a local database or establish a direct feed for use with extraction. In that way, they could achieve CJDT compliance without making modifications to their CMS.

3.3.8. Full Extraction Testing

The successful completion of all preceding preparations sets the stage for full extraction testing. This is the last step in evaluating the stability of your data foundation. Through initial querying, validating, cleaning, calculating, process changes, system modifications and workarounds, an agency should have a complete data source. The next step is to test extracting from that source.

Full extraction testing requires querying the entire dataset (based on the reporting specification) as the last level of validation and to test operational variables, like file sizes and runtimes, as well. This method is usually achieved via multi-level queries and/or programs developed by agency IT staff. The goal is to check the quality of these queries/programs by validating extractions and determining their impact on infrastructure.

Checking quality could involve manual verification of data using spreadsheets and a CMS user interface. It could also be accomplished by comparing smaller data subsets to the full extraction and checking for inconsistencies. An agency should also use the results from the validation exercises to double-check elements where adjustments were made for compliance.

Infrastructure testing is used to evaluate technology requirements and to determine the best times and methods for running an extraction procedure. This may be phase related. For example, with CJDT there is an initial data population phase and then a scheduled transmission phase. The initial phase involves creating a base level dataset in FDLE's repository consisting of cases from (at least) 2019 to current. The initial dataset is magnitudes larger than the subsequent scheduled transmissions, which is only to transmit records that had been updated. An agency should determine the best time to run queries based on factors like bandwidth, processing power and memory, server availability, runtime lengths, file sizes, etc.

Once an agency has completed that step, it would be compliant with the reporting specification from a data availability perspective; the foundation would be secure.

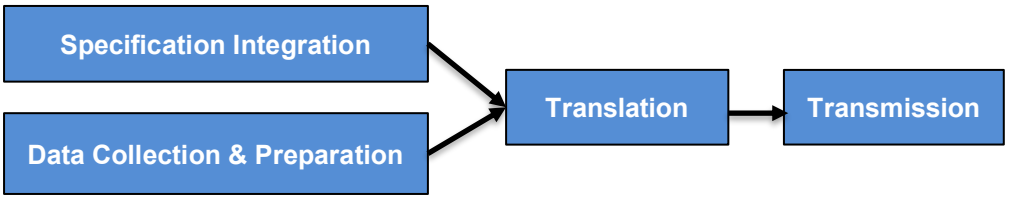
3.4 Automation & Standardization

Considering all the factors that may be necessary to effectuate the structuring of the data, the task can seem daunting. However, this challenge also presents an opportunity, as it will be explained below.

3.4.1. Breaking down reporting into process blocks

The first step to actualizing a reporting model is to break the process into chunks or blocks. Regarding CJDT, those blocks and the flow among them are diagrammed in **Figure 8.**

Figure 8 Process Blocks



3.4.1.1. *Specification integration* - refers to obtaining the latest report specification (i.e., XSD/XML based specification from FDLE for CJDT) and converting it into a usable format for an agency's system.

3.4.1.2. *Data collection and preparation* - includes all the process and methods used to prepare an agency's data to comply with a reporting specification (Section 3.3).

3.4.1.3. *Translation* - the convergence of the FDLE transmission specification with a complete dataset to create a report in XML format.

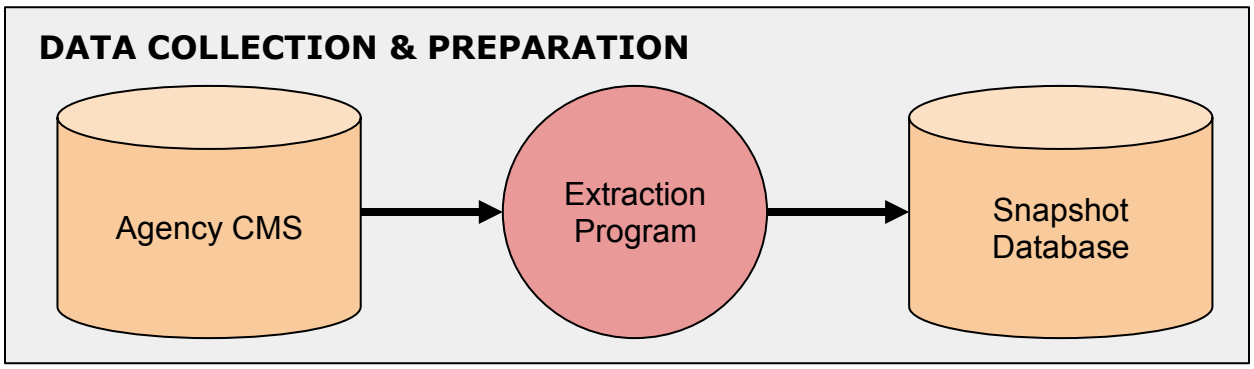
3.4.1.4. *Transmission* - sending the finalized report to FDLE.

Note, this process model can apply not only to CJDT but can also represent a standard for other reporting initiatives. The process blocks can also be divided into smaller parts or components, some of which will be discussed below.

3.4.2. Developing tools for the process blocks

Once the process blocks have been identified the next step is to develop tools to automate those processes. As this chapter deals with structuring data for extractions, emphasis is on the data collection and preparation block. The diagram in **Figure 9** breaks down this block into components.

Figure 9 Data Collection & Preparation



The starting point is an agency's CMS. An extraction program is then used to manage the cleaning, coding, and extraction of the required data. Because each CMS is unique, the extraction program must be customized for each agency. Finally, a snapshot database, or a local repository for extracted data, is created. While not required, the snapshot database has the added benefit of allowing agencies to further audit and validate the data.

The extraction program represents the beginning of automation and an agency's interface to a standardized tool. The snapshot database, if generalized, becomes the entry point to standardization of which other components pivot their function to further the reporting process.

One of the pilot counties developed a tool based on this approach called Transparency Initiative Data Extraction (TIDE) tool.

3.4.3. Introduction to the Transparency Initiative Data Extraction (TIDE) tool

The work done in the pilot counties uncovered the possibilities for automation and tool generation, resulting in TIDE. TIDE, a solution designed by the Pasco County Clerk & Comptroller IT department to achieve CJDT compliance, uses a combination of infrastructure agnostic components, addressing all workflow process blocks from the model. While TIDE is introduced here, supplemental documentation and future chapters will discuss it further.

<u>Micro to macro approach history of TIDE:</u>
The Clerks in the pilot counties, Pinellas and Pasco, use Odyssey and Clericus, respectively, for their CMS solutions. In developing a CJDT solution for Pasco County, it was discovered that with modification, the tool could be integrated into Clericus and thus used to help other counties using the same CMS. Additionally, after conversations between the Pasco Clerk IT and Pinellas Business Technology Services departments, it was found that with adaptations favoring generalization, Pinellas could use a similar solution as well. As with Clericus users, a solution developed for Pinellas could inherently benefit other Odyssey counties. Furthermore, if the solution, as applied to Pinellas, was developed to be system agnostic, it could have the potential of helping all Florida counties achieve CJDT compliance and become a reporting standard.

With this tool, agencies would be responsible for preparing their data and developing a means for extraction (via an extraction program component) into TIDE's snapshot database. This extraction function, as outlined in the previous section, is consistent with current levels of data reporting and data requests. The tool model would also manage the integration of FDLE's specification and translation of the data into the proper format for transmission. Although the tool is still in development, it has shown promise for becoming a statewide reporting standard for CJDT and perhaps future reporting projects.

3.5 Chapter 3 Summary

The goal of data structuring for extraction is to prepare an agency's technology environment to accurately collect a complete dataset in accordance with a reporting specification. The differences between an agency's data and the elements required by CJDT must be fully understood. Modification of the data, the CMS, or data collection practices must be implemented to meet CJDT standards and prepare the data for extraction. Automation and standardization are keys in developing a means for extraction that can be replicated across counties. In the end with preparation, an agency will have a firm data foundation from which to extract information.

Chapter 4: Data Extraction Translation & Validation

Once a complete set of data elements exists, the next step is to create a report for transmission in the required format. With regards to CJDT, this means translating the dataset into the National Information Exchange Model (NIEM) XML specification provided by FDLE. While **Chapter 3: Structuring for Data Extractions** focused on preparing data elements based on their values and formats, this chapter delves into understanding the FDLE spec, translating data into XML, techniques for validating translations and transmitting data to FDLE. This chapter also continues the discussion on automation, detailing how the Transparency Initiative Data Extraction (TIDE) tool manages translation, validation, and transmission.

4.1. Understanding the Specification

To specify the CJDT transmission format, FDLE provided local agencies with guidance in the form of Data Element Dictionaries (DED) and Information Exchange Package Documentation (IEPD), both available on CJNet. The DEDs define data elements for each agency and provide sample values and general syntax rules in an easily read form. The Information Exchange Package (IEP) is more advanced consisting of a collection of XML Schema Definition (XSD) files and documents that provide extensive protocol requirements including formatting, code values, object associations and much more. Used together, they give agencies and their technology professionals all the information they need to translate their data.

This section provides a brief and high-level overview on using these documents. It assumes a moderate level of technical knowledge and ability with XML. For more in-depth information, please refer to the original documents or call the FDLE Helpdesk for support.

4.1.1. Objects and elements

To send data in XML, it is important to understand the structure. The basic structure of the FDLE NIEM syntax consists of objects and data elements grouped by criminal justice components and expressed using markup language. FDLE's IEP contains a complete listing of all objects and elements used to transmit data for all its reporting responsibilities. A subset of that information, outlined in the DEDs, is relative to CJDT. To illustrate this, and help explain how the DED and IEP work, it is helpful to look at how a defendant's information is formatted.

Because XML is hierarchical, it allows nesting of objects and elements as a means of grouping data, conveying relationships, and standardizing information transmission. FDLE has documented its complete hierarchy of reporting elements in the *FDLE High Level Model.pdf* file within the IEP. The highest level object for a CJDT report is the FDLE Message noted by the <fdle:Submission> tag. Beneath that is <fdle:Report>, followed by the <nc:Person> tag used to classify information related to defendants. CJDT requires the collection of several different data points associated with defendants including, but not limited to: birth date, name, gender, race, etc. The majority of these elements are listed under the <nc:Person> tag. **Figure 10** uses a tree diagram to depict this hierarchy of objects and elements.

Figure 10 Sample Hierarchy of CJDT XML Objects*

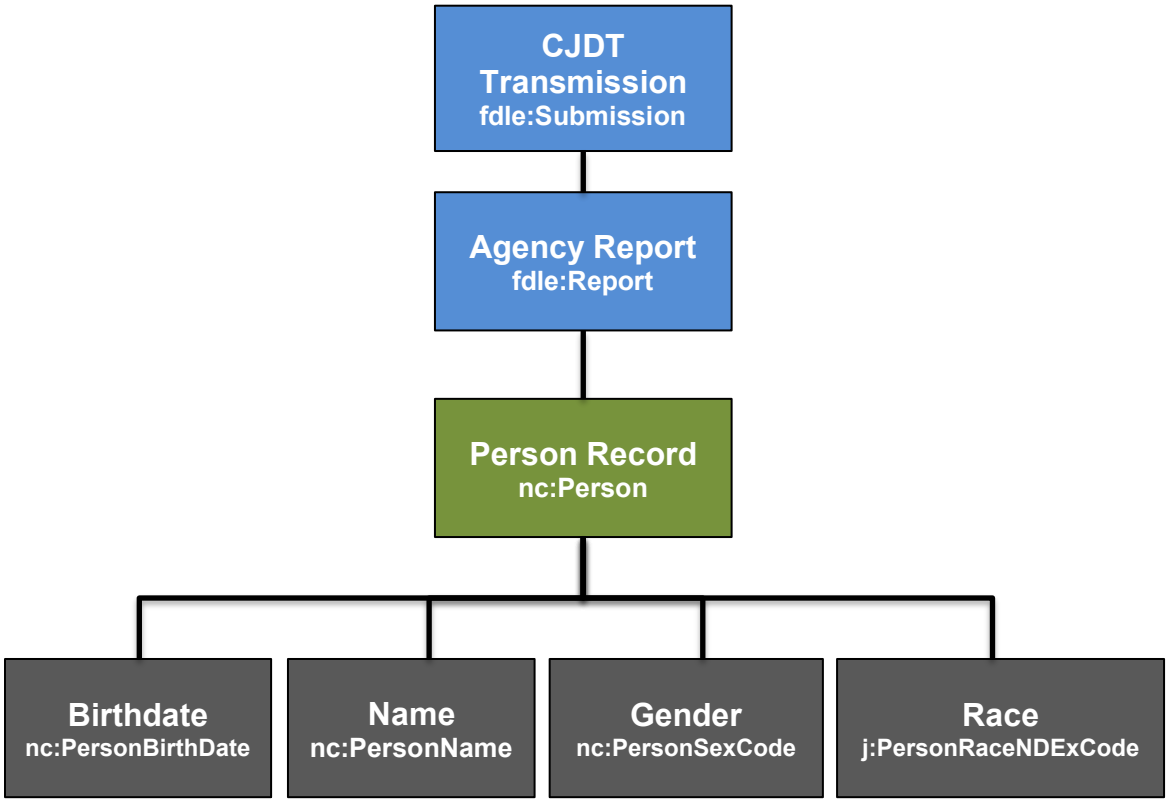


*Diagram based on IEPD v1.2. Subtext shows syntax for associated XML

In the diagram, the labels indicate the object names and underneath their corresponding XML tags. The example shown is limited as a full person record contains many more elements and additional levels of elements.

For instance, within the Name object (i.e., nc:PersonName) there are additional elements for the parts of a person's name including their first name, last name, suffix, etc.

Notice that each XML tag has a prefix, followed by a colon, and then the element's name. This naming system is a feature of the NIEM XML protocol. Before continuing further, we will introduce NIEM and explain how it is incorporated into transmissions for CJDT.

4.1.2. NIEM Introduction

"The National Information Exchange Model (NIEM) is a common vocabulary that enables efficient information exchange across diverse public and private organizations."[5] Think of it as an XML catalog of predefined tags each with definitions, relationships, and formats. NIEM's XML tags are organized by domains that classify both general and community specific elements. The domains are indicated by the previously mentioned tag prefixes. More information about NIEM can be found at NIEM.gov.

Relative to CJDT, the majority of native NIEM tags in the specification fall under the NIEM core (nc) and justice (j) domains. FDLE has also expanded the protocol to capture the uniqueness of its reporting requirements (including but not limited to CJDT) via its own domain using the "fdle" and "cjadmin" prefixes. The CJDT transmission protocol is a combination of tags from these native and custom NIEM domains to create a comprehensive format for reporting.

4.1.3. Base Structure Using the DEDs

To convey the NIEM XML hierarchy from **Figure 10**, FDLE's DEDs define xpaths for each data element. Xpaths are expressions, much like file paths for computers, that can be used to navigate XML. These expressions give insight into the XML structure required for submitting a data point. The xpath for Person records begin:

/fdle:Submission/fdle:Report/nc:Person…

following the hierarchy from the diagram. All subsequent Person elements will begin with the same appellative. Specifically, the xpaths for some of the example elements are:

- Birthdate:
  /fdle:Submission/fdle:Report/nc:Person/nc:PersonBirthDate

- Gender:
  /fdle:Submission/fdle:Report/nc:Person/nc:PersonSexCode

Reviewing the xpaths for each element provides a good understanding of the base XML structure. The next step is to convert that structure into XML code. Using the example, the xpath of each element can be traced by observing the hierarchy indicated by each forward slash ("/") and noting the elements terminating the xpath strings. Using this method, **Figure 11** demonstrates the base XML code for our example Person record, as follows:

Figure 11 DED Person Record

```
<fdle:Submission>
    <fdle:Report>
        <nc:Person>
            <nc:PersonBirthDate>...</nc:PersonBirthDate>
            <nc:PersonName>...</nc:PersonName>
            <nc:PersonSexCode>...</nc:PersonSexCode>
            <j:PersonRaceNDExCode>...</j:PersonRaceNDExCode>
        </nc:Person>
    </fdle:Report>
</fdle:Submission>
```

In practice, the instances of "..." above would be replaced with a defendant's information. This simple example demonstrates how to build a base XML structure corresponding to the FDLE specification. However, to fully translate CJDT data, users will need to use the IEP.

4.1.4. Validating the structure using the IEPD

The IEP documentation provided by FDLE is extensive. As mentioned, it includes a wide variety of documentation with various specifications for all

their reporting initiatives. A good place to start is with the *readme.pdf* file. This file contains descriptions for all of the IEP support documents.

Essential to understanding the XML structure of CJDT are the *FDLE_Mapping_Spreadsheet* and the agency associated IEP samples. The mapping spreadsheet contains the definition, classification and rules for every data element, in addition to xpaths, attributes, relationships and much more. The CJDT sample documents, XML files with names beginning *Data Transparency Case…*, provide practical examples of the actual XML code.

Without going too deep, here is a continuation of the example above by looking at how *Person* elements are defined in the mapping spreadsheet. The spreadsheet has multiple tabs and built-in documentation. We will focus, for now, on the "Data" tab which has the element list.

Table 3 below is a pseudo-sample of the mapping spreadsheet modified to show only CJDT information relative to the *Person* example and color-coded to ease explanation.

Table 3 Mapping Spreadsheet Sample for Sample Data Elements

Source Class	Source Element	Source Definition	CJDT	Target Class	Target Element
Person		A human being...	X		
Person	Birth Date	A date when the person was born.	X	nc:Person	nc:PersonBirthDate/nc:Date
Person	Sex Code	A gender or sex of a person.	X	nc:Person	j:PersonSexCode

Legend:

Source Class | Element Info | Reporting Protocol | Xpath

Data elements are grouped together according to their parent objects called source classes. The beginning of source class groupings is indicated by yellow rows. The source class for our example (as shown

above) is Person. After the yellow divider, all data elements associated with that source class are listed.

The first set of columns with blue headers are an element's identification information: source class, name, and definition. The next set of columns in green, mark the reporting protocols that an element is a part of using an "X" to indicate inclusion. The actual spreadsheet has several protocols/columns listed including NIBRS, Use of Force (UoF) and CJDT breakdowns by agency. Agencies can use these columns for filtering the spreadsheet's list of source classes and elements to show information specifically relevant to them.

The last column set shown in purple provides the tag information using xpath. "Target Class" refers to the XML tag for the source class, where the "Target Element" shows the xpath for an element within that target class. This information should look familiar since it is like the DED xpaths and follows the hierarchy from **Figure 10**.

We can affirm the knowledge gained from the DED and the mapping spreadsheet by looking at an IEP sample. **Figure 12** below is truncated XML from the IEP sample document of the Person source class and the elements in the following example.

The advantage of the sample document is that it shows the XML in practice with sample data. The elements themselves are labeled using XML comments and the sample data follows the formatting and value restrictions detailed in the DEDs. The combination of DED and IEP documents correlate and can be used to understand and create the basic CJDT XML structure. The subsequent sections will build upon that structure.

Figure 12 Truncated XML from the IEP Sample Document of the Person Source Class

```
<fdle:Submission… >
    <fdle:Report… >
    ...

        <!-- Person 1 -->
        <nc:Person structures:id="Person1">

            <!-- Birth Date -->
            <nc:PersonBirthDate>
                <nc:Date>1960-01-01</nc:Date>
            </nc:PersonBirthDate>
            ...

            <!-- Race Code -->
            <j:PersonRaceNDExCode>W</j:PersonRaceNDExCode>
            …

        </nc:Person>
    ...
    </fdle:Report>
</fdle:Submission>
```

4.1.5. IDs & Object Associations

CJDT transmissions will usually consist of multiple objects and related elements. Because the structure of those elements is similarly codified, the CJDT XML specification uses object IDs to uniquely identify and relate data. In the previous code example (and in future examples) the object/element ID, an XML attribute, is highlighted in blue. This example can be expanded by explaining object relationships starting with a defendant's alias. **Figure 13** is the IEP sample code for an alias:

Notice that to define an alias for a defendant, a new person object must be created with a unique identifier. Notice also that the <j:PersonNameCategoryCode> tag and corresponding "alias" value are used to signify the name provided as an alias. Although only one alias is used in the example, the specification does allow for multiple aliases.

Figure 13 IEP Sample Code for an Alias

```
...
<!-- Person 1 Alias -->
<nc:Person structures:id="Person1Alias">
     ...

     <!-- Person Name and Name Category -->
     <nc:PersonName>
          <nc:PersonGivenName>John</nc:PersonGivenName>
          <nc:PersonSurName>Doe</nc:PersonSurName>
          <j:PersonNameCategoryCode>alias</j:PersonNameCategoryCode>
     </nc:PersonName>

</nc:Person>
...
```

Appending the above code structurally creates a defendant (or person) object and a separate person object with that defendant's alias. To link the example defendant to the alias, CJDT uses associations. As the name implies, an association explicitly identifies the relationship or association between two objects, as shown in **Figure 14**.

Figure 14 Visual Representation of an Alias Association

An association is like a chain that links objects together by defining their relationship. The keys to that relationship definition are the object IDs. **Figure 15** shows the code for an alias association, as follows:

Figure 15 Code for an Alias Association

```
...
<!-- Person Alias -->
<nc:PersonAliasIdentityAssociation>
    <nc:Person structures:ref="Person1"/>
    <nc:Identity>
        <nc:IdentityCharacteristics structures:ref="Person1Alias"/>
    </nc:Identity>
</nc:PersonAliasIdentityAssociation>
...
```

The association root tag, <nc:PersonAliasIdentityAssociation>, identifies the association object and its intention: to associate a person and an alias. The subsequent tags use the IDs of the defendant object (within the <nc:Person> tag) and the alias person object (within the <nc:IdentityCharacteristics> tag) to specifically identify the linked objects. Multiple aliases for one person can be defined with additional alias associations, provided the referenced alias IDs are unique.

There are various types of associations used in the CJDT specification. The complete list can be found on the “Associations” tab of the mapping spreadsheet. Its format is similar to the “Data” tab and also details similar information for each association.

4.1.6. Business Rules: Cardinality

Business rules and restrictions refer to the limitations on data objects and elements within the CJDT specification. More specifically, it refers to requirements on the allowed number of object and element instances based on the context of the data being transmitted.

This is a good time to revisit the statement made in the previous section regarding associating multiple aliases. Within the mapping spreadsheet, all the elements and objects have a “Cardinality” column that provides an instance ratio. **Table 4** below is an abbreviated table showing the cardinality for elements of an alias association.

Table 4 Cardinality of an Alias Association and Related Data Elements

Source Class	Source Element	Cardinality
Person Alias Identity Association		0-n
Person Alias Identity Association	Person	Person 1 ref 1 Person...
Person Alias Identity Association	Alias Identity	Identity 1-n Characteristics 1 ref 1 Identity...

The cardinality rules are like those discussed in Chapter 2 when developing data schemas. Here is the logic expressed by the cardinality rules one row at a time.

0-n (Association Source Class)

Note that this first cardinality statement applies to the parent alias association object. Because a defendant may have zero or many aliases they consequently can have (0) or many (n) alias associations. Based on this rule aliases are optional for transmission.

Person 1 ref 1 Person (Person Element)

If a defendant has an alias, the alias association object must correspond to 1 (and only 1) person (or defendant per the example). This person object is identified or referenced by its unique ID (ref).

Identity 1-n Characteristics 1 ref 1 Identity (Alias Identity Element)

Additionally, if a defendant has an alias, the alias association must also have at least 1, but may have many (n), related alias characteristics. However, all these characteristics must relate to 1 (and only 1) defendant. For this association, the only characteristic required is the alias' reference ID (ref).

To summarize, based on these cardinality rules, to relate any number of aliases to a single defendant, the specification requires one to multiple alias associations. Each association is required to have the original defendant's ID and a unique ID for each alias.

4.1.7. Business Rules Continued

Continuing the discussion on business rules, there are a few other places within the DEDs and mapping spreadsheet that also describe the cardinality rules discussed earlier in addition to value limitations and dependencies. Within the DEDs this place is in the "Additional Information" section for some data elements. The mapping spreadsheet uses the "Comments" and "Mapping Notes" columns for this purpose.

For instance, relative to the *Alias Identity* element in the example, the mapping spreadsheet describes the identity characteristics allowed when defining an alias. Other elements have detailed descriptions addressing origins of values, dependencies on other elements and even usage notes for other FDLE reporting initiatives. Reviewing this information can be helpful in understanding data requirements and validating your XML.

Please note that object associations are not specifically defined in the DEDs; however, there are inferred object association references in the xpaths for certain elements.

4.1.8. Data Formatting & Code Values

After the data preparation step (from **Data Preparation**), an agency should have a strong sense of what types of values and formats are required for CJDT data transmissions. Formats are specific for every data element and are defined in several documents. The best place to start is with the data element dictionary relative to a specific agency. Each dictionary has a "Format Types" section (towards the beginning of the document) defining the common data formats and permitted alphanumeric sequences.

Some data elements are further restricted to certain keywords or code values. Most of these code values are also available in the DEDs. The codes are listed with their associated elements and in the appendix for longer lists. FDLE has also included an FDLE_CodeTables spreadsheet

as a reference for certain data types. Additionally, each FDLE defined NIEM domain has a *<domain name>-codes.xsd* file. These files and the native NIEM domain XSDs (in the IEP *niem-core* directory) have complete code indexes relative to their elements as well as descriptions in XML format.

For our example, the code values for <j:PersonNameCategoryCode>, used to classify the alias name, can be found under the "Value Options" for the Name Category element in the DED. Note by the XML prefix that this is part of the justice domain, therefore the same code values can also be found in the justice domain's XSD (i.e., the *jxdm.xsd*). However, to find the values within that file, there is another column in the IEP's mapping spreadsheet: "Target Type," as shown in **Table 5**.

Table 5 Mapping Spreadsheet Sample of Target Information

Source Class	Source Element	Target Class	Target Element	Target Type
Person	Name Category	nc:Person	nc:PersonName/j:PersonNameCategoryCode	j:PersonNameCategoryCodeType

The "Target Type" is the data type for the element storing the code values. As shown above the data type for Name Category is *j:PersonNameCategoryCodeType*. To find the corresponding code values in the XSDs, search for this data type within the *jxdm.xsd* file. **Figure 16** below is an excerpt from that file.

Notice the text in bold. Based on the XML hierarchy, it is clear that *PersonNameCategoryCodeType* is representative of another data type — *PersonNameCategoryCodeSimpleType*. This data type is defined below in the same XSD.

Figure 16 Finding Corresponding Code Values

```
...
<xs:complexType name="PersonNameCategoryCodeType">
  <xs:annotation>
    <xs:documentation>A data type for possible kinds of names.
    </xs:documentation>
  </xs:annotation>
  <xs:simpleContent>
    <xs:extension base="j:PersonNameCategoryCodeSimpleType">
      <xs:attributeGroup ref="structures:SimpleObjectAttributeGroup"/>
    </xs:extension>
  </xs:simpleContent>
</xs:complexType>
...
```

Figure 17 Code Values and Descriptions

```
...
<xs:simpleType name="PersonNameCategoryCodeSimpleType">
  <xs:annotation>
    <xs:documentation>A data type for possible kinds of names.
    </xs:documentation>
  </xs:annotation>
  <xs:restriction base="xs:token">
    <xs:enumeration value="aka">
      <xs:annotation>
        <xs:documentation>Also known as, e.g., a stage name
        </xs:documentation>
      </xs:annotation>
    </xs:enumeration>
    <xs:enumeration value="alias">
      <xs:annotation>
        <xs:documentation>An assumed or alternate name suspected to be in
use for deception; usually involves criminal intent.  A term used in legal
proceedings to connect the different names of anyone who has gone by two or
more, and whose true name is for any cause doubtful.
        </xs:documentation>
      </xs:annotation>
    </xs:enumeration>
  ...
  </xs:restriction>
</xs:simpleType>
...
```

Figure 17 above enumerates the code values and contains their descriptions (between <xs:documentation> tags). Not all data types will require tracing through hierarchies. However, this example shows that investigating an element's "Target Type" will yield its admissible values.

4.1.9. Metadata & Header Values

At the beginning of each transmission file there are a set of required values that include a transmission's metadata. These elements will be used by FDLE to identify and classify a transmission, in addition to informing additional processing steps. Once again, the mapping spreadsheet and the relative IEP sample are especially useful for researching these elements. In the mapping spreadsheet the element information is on the "Metadata & Header" tab.

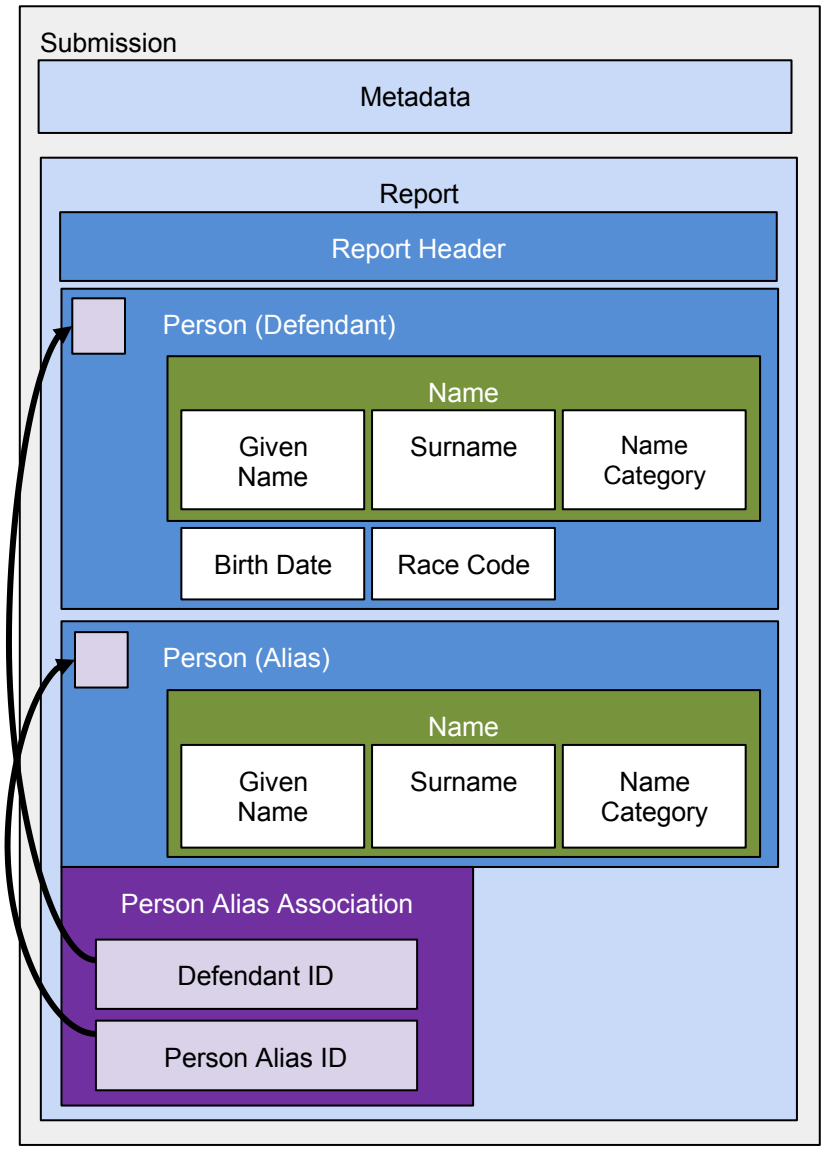
4.1.10. Full Example

All the objects and elements discussed in our example are representative of the basic structure of a CJDT transmission. **Figure 18** illustrates that example, using shapes to portray the XML framework. The data elements are depicted in ▢ white. The object IDs (attributes) are shown in ■ purple with arrows illustrating the relationship defined by the alias association.

All transmissions will consist of a *Submission* with metadata, a *Report* with header information, and CJDT data objects with corresponding elements and attributes. One can codify the full example by essentially substituting the placeholders in the figure with the previously discussed code snippets. More elements can be added by including additional object and element structures. **Figure 18** consolidates concepts discussed to depict a CJDT transmission's fundamental aspects.

More extensive samples can be found within the IEP documentation. Those examples have more code specific instances of objects and their elements. Combining this information with that provided by the DEDs, XSDs and other supporting documents, provides a complete understanding of the specification.

Figure 18 Visual Representation of CJDT XML Basic Structure



4.2. Translating Data into XML

With a firm grasp of the specifications, data can be translated into the appropriate XML format. There are various methods to perform this and two are discussed below.

4.2.1. Hard-coded translation

This method refers to manually adding code to an SQL query to format exports. More specifically, each query and subquery would have XML strings concatenated with data values according to the requirements. An

advantage to this method is that it is relatively straightforward; however, it does require a strong grasp of the XML specification and considerable, manual code maintenance. Additionally, depending on how the data are stored, matching XML tags to data values may present some structural challenges.

4.2.2. Dynamic Link Library (DLL) Translation

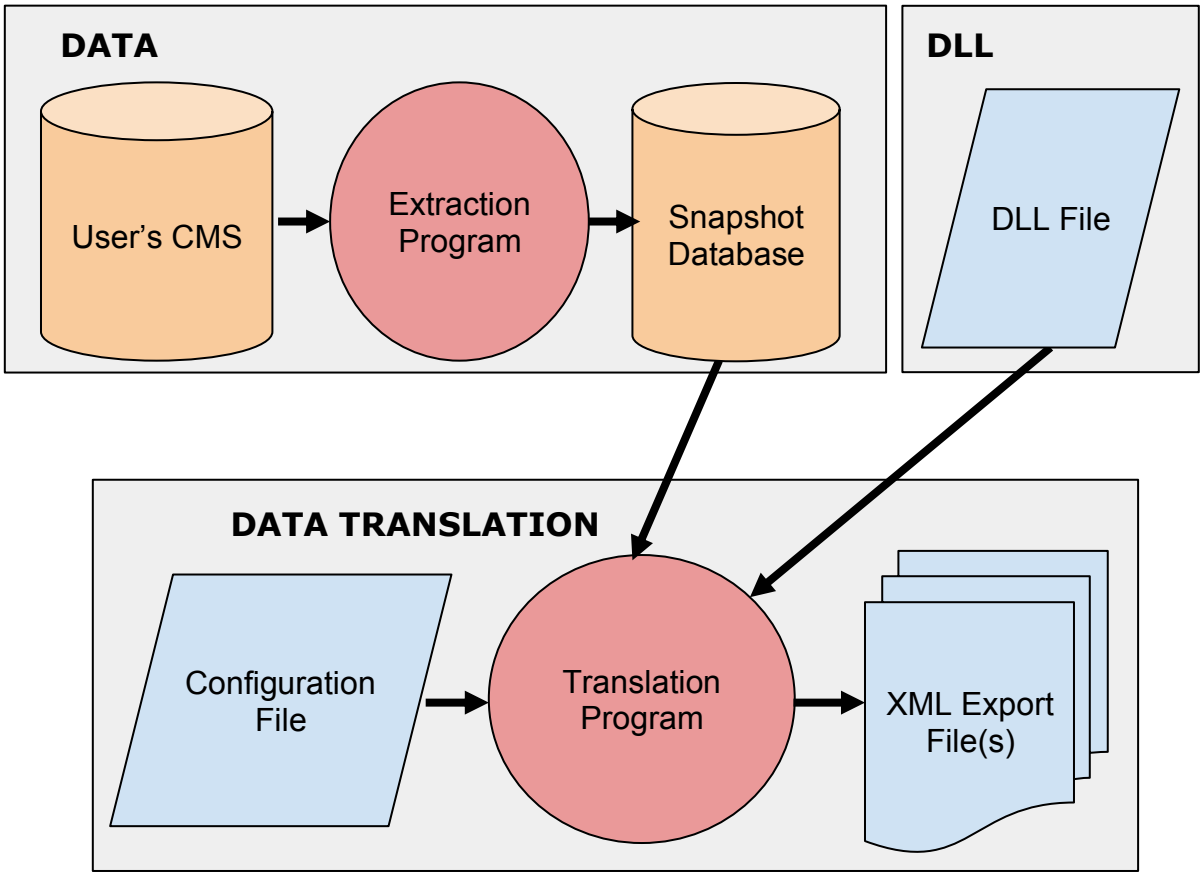
The use of a DLL file simplifies the hard-coded method by serving as a mapping reference for code translation. This reference lowers the XML specification knowledge burden by allowing programmers to map data values more easily to objects and have those mappings automatically translated when code is executed. Although a DLL does not necessarily remedy potential structural challenges, it does make querying and coding much easier to maintain.

4.2.3. Translation with TIDE

As described in **Chapter 3**: Structuring for Data Extractions, TIDE first extracts and stores an agency's data in a snapshot database. Then, TIDE translates the data into XML using DLL translation. With this translation step, the previous TIDE workflow diagram is expanded with two additional processes: DLL creation and data translation. See **Figure 19 Expansion of TIDE Process Illustration**.

There are a few additional steps included in the DLL creation process but for the sake of brevity and to focus on relative information, those steps are not included above. The translation process adds a small additional responsibility to also maintain the configuration file. The configuration file allows the translation program to find the connection(s) to an agency’s local snapshot database and specifies the location where CJDT export files will be stored. The translation program then queries the data from the snapshot database and creates XML files per case and in accordance with specification.

Figure 19 Expansion of TIDE Process Illustration



4.3. Local Validation

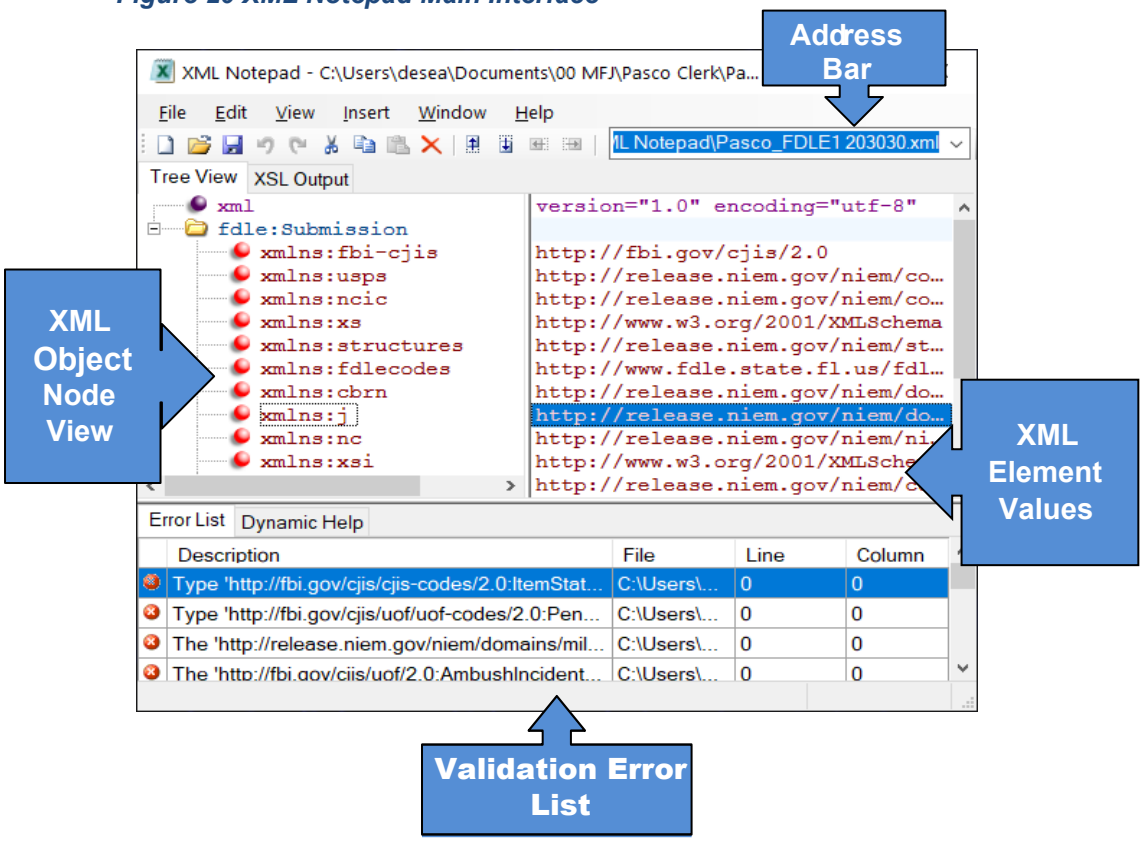
As part of development, agencies can perform local validations to ensure the accuracy of submission files before they are transmitted to FDLE.

Although tedious, validation could be done via manual review. However, this method is inefficient requiring an extensive knowledge of the XML specification and a keen eye. Alternatively, there are a variety of software options available to perform validation; the most robust of which are expensive. To find cheaper alternatives, validation testing was performed with success in the pilot counties using XML Notepad. XML Notepad is a free and open-source tool originally designed by Microsoft and now available for download on GitHub. The general procedure for validation using XML Notepad is relatively intuitive and outlined below.

4.3.1. Feature Overview

XML Notepad contains a strong base of features for manipulating and validating XML files. For starters, it provides a hierarchical node-based view of your XML code. This interface presents the XML objects and elements as selectable nodes, making it easy to add, delete, and modify structures and values associated with each node. For validation, the tool has an error log with associated file location information. Additionally, the software is lightweight and quick, with help documentation available both locally and online. A complete detail of features is also available from the same sources. **Figure 20** shows several features mentioned with sample validation information.

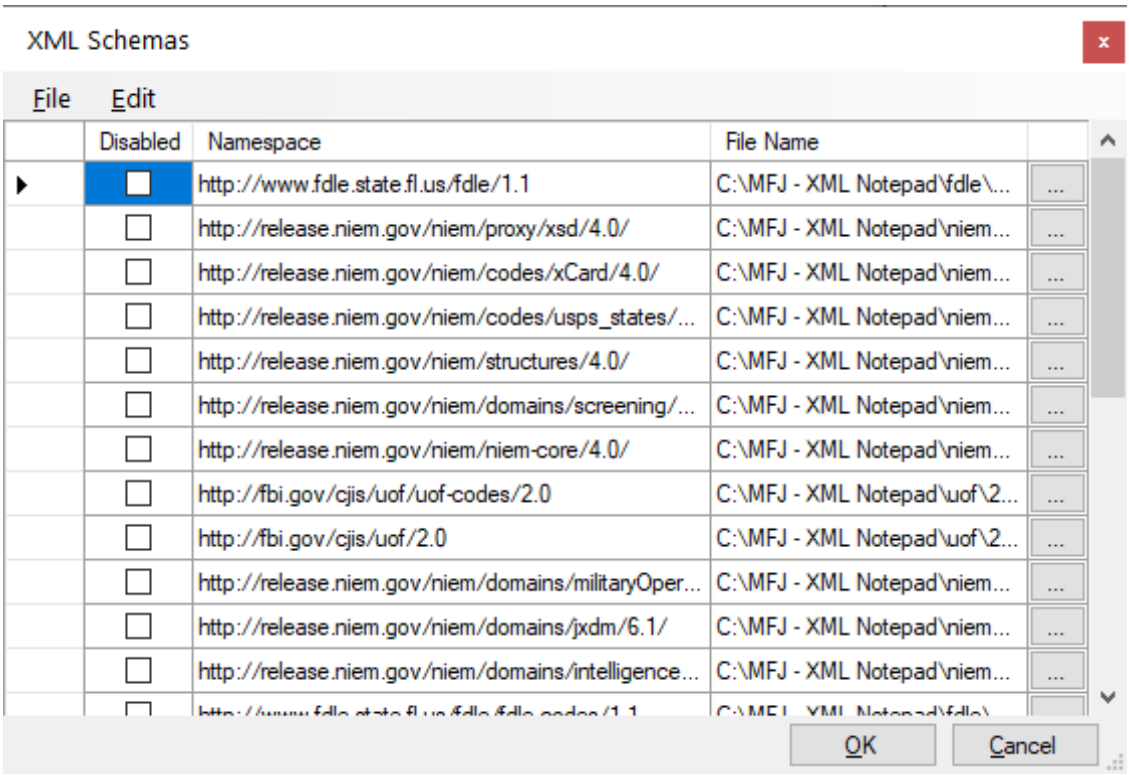
Figure 20 XML Notepad Main Interface



4.3.2. Load the Specification

The validation tests performed were done with local versions of the export/request files and FDLE's NIEM specification. To perform the validation, XML Notepad needs to know where these files are located. shows a sample of the XSDs loaded into the software, which can be done by clicking *View* on the menu bar, followed by *Schemas*.

Figure 21 XML Notepad Schemas View



Once the specification is loaded the next step is to indicate where the export file exists by either dragging and dropping the file into XML Notepad's main screen or by using the address bar. Once the necessary files are loaded, validation can begin.

4.3.3. Validating & Resolving Errors

When the request is loaded, XML Notepad triggers the validation process automatically, populating any applicable errors. Refer again to the validation error list section from **Figure 20**. There are four fields: description, file, line, and column. The description field provides detail on the error message(s), while the remaining fields refer to specific files and code locations within those files that are causing the validation issues.

To resolve, one can go through each message and make changes within the software. As changes are applied the tool will make real-time updates to the XML code and error log.

4.4. Transmitting Data to FDLE

The minutiae on how and where to send CJDT transmissions to FDLE is contained in the Interface Control Document (ICD). The ICD is comprehensive, containing details for all aspects of transmission. As a result, the following information is designed to provide users with a conceptual understanding as specifics can be found in the ICD.

4.4.1. Applying for Credentials

To submit information to FDLE, all agencies are required to create a service account. This account will provide access to FDLE's test system and, following successful test transmissions, the production system. The procedure for creating this account is found in the ICD.

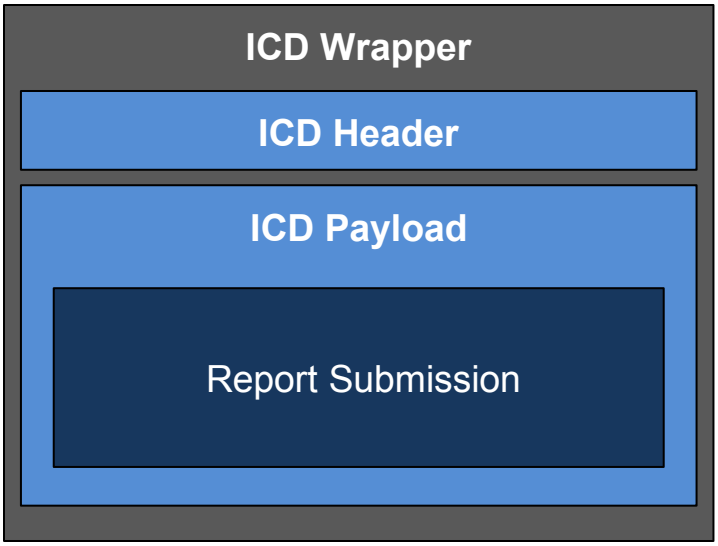
4.4.2. ICD Schema/XML - Request Messages

According to the documentation the "ICD schema can be described as a wrapper layer around the IEPDs." Specifically, the ICD specifies an additional XML hierarchy that precedes the code containing the CJDT report. This additional XML layer consists of header and payload element groupings, varying based on the type of report being sent. With this addition, the submission information visualized in **Figure 18** falls under the payload group as shown in **Figure 2.**

Please note, while the ICD structure is similar for the CJDT report types, i.e., FDLE IEPD and CJAdmin IEPD, there are small variations in XML tag names and data provided. Be sure to follow the ICD reference relative to your agency.

The combination of the CJDT report wrapped in the ICD schema constitutes a complete CJDT transmission file. In the ICD this is referred to as a request. It is this completed request file that agencies will send to FDLE.

Figure 22 ICD Wrapper Hierarchy



4.4.3. Response Messages

After the transmission of a CJDT request FDLE systems will send a reciprocal message called the response. Like a request, CJDT responses are written in XML following a format specified in the ICD.

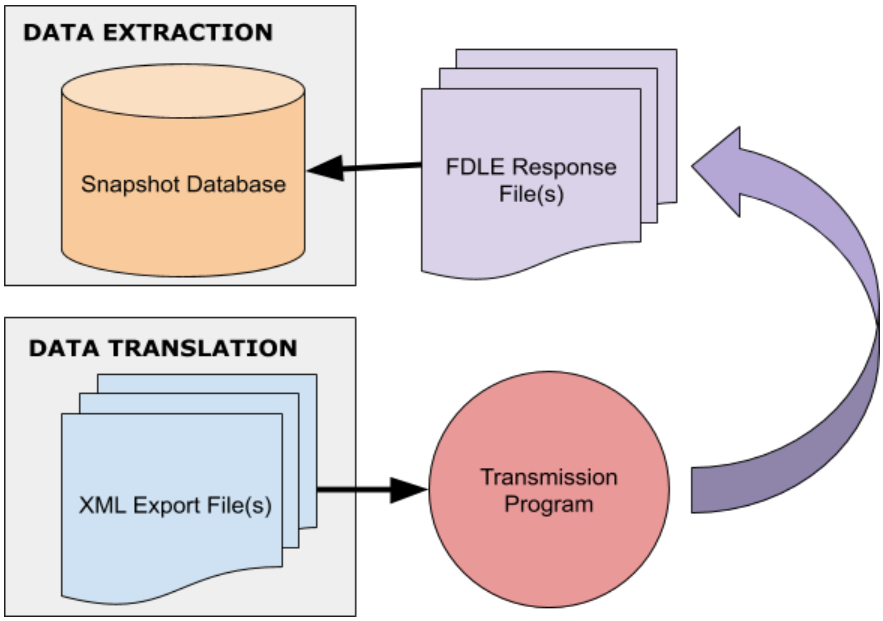
4.4.4. Transmission with TIDE

TIDE uses the combination of the transmission program and snapshot database for sending transmissions and receiving responses.
 illustrates this process.

In addition to translating the data into XML, the translation program also applies the ICD Wrapper. The transmission program then uses an agency's credentials (i.e., the service account) to transmit the export/request files to FDLE and to receive the responses. The ratio of requests to responses is 1:1. As each export file represents a case and an agency will typically send hundreds of case files per transmission, the number of responses received can quickly add up. To simplify this process, the translation program is designed to receive and parse the XML responses and store the values in the snapshot database for review.

If there are any errors, a user can use the logs to pinpoint the issue and reprocess and retransmit data for the offending request files. With this step, the final processing for TIDE is complete.

Figure 23 TIDE Transmission Components



4.5. Chapter 4 Summary

FDLE has provided several documents detailing the NIEM XML specification and mechanisms involved in transmitting data for CJDT. This chapter is meant to provide guidance to agencies as they review these reference materials.

Key references include the Data Element Dictionaries (DED), Information Exchange Package Documentation (IEPD) and the Interface Control Document (ICD). The DEDs are useful as references for all users, outlining data element responsibilities for each agency as well as providing essential data attributes, values, and formatting information. The IEPD is a more advanced collection of documents that instruct technical staff on the practical implementation of the NIEM specification. The ICD explains the infrastructure in place for receiving agency files and the protocols for their transmission. It is the combination of these documents that will give users the basis of CJDT compliance.

In addition to the reference material there are also tools available to help translate, transmit, and validate an agency's data. TIDE's translation and transmission components build upon its extraction capability by leveraging the snapshot database. It inherently provides a level of validation through its use of a

DLL file for structuring data. The features of XML Notepad allow users to perform further validation of export files locally, if desired.

This concludes the discussion on CJDT implementation. The next chapter shifts gears, focusing on the spirit of data transparency by establishing key principles for a healthy data culture.

Chapter 5: Improving Data Culture

5.1 Introduction

Data is fundamental to solving societal problems. However, data do not exist in isolation, and must be understood as part of a wider ecosystem that includes science, technology, economics, and society-driven innovations.[6] Organizations must develop a data culture that allows them to navigate this ecosystem to facilitate lasting change.

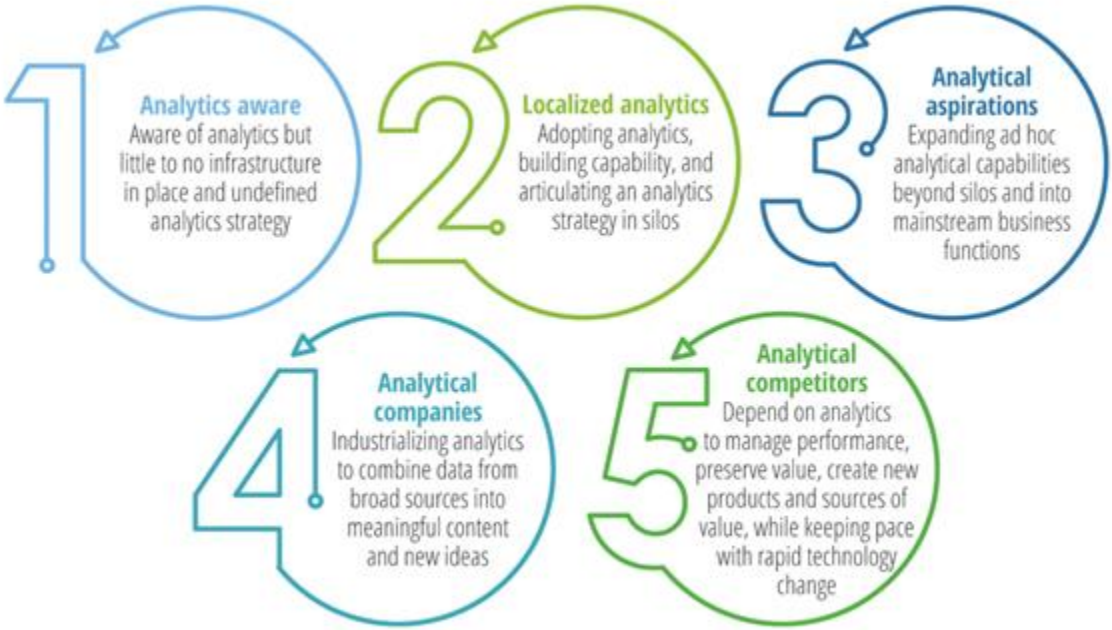
The previous chapters introduced proper business processes for data collection, the pilot project and consensus built around data element definitions and schemas, and practical guidance on data extraction and transmission to FDLE. This chapter focuses on the organizational culture needed to support the effective use and consumption of data to drive decision-making.

5.2 Definition of Data Culture

While data are not a natural resource, they are everywhere and are constantly collected. Think of data as a cultural resource that needs to be generated, protected, and interpreted.[7] Data culture should lead to decision-making that is consistent and repeatable, yet adaptable to new insights, using tactical and strategic processes based on empirical evidence.[8] Selecting data pathways that support the strategic alignment sets the stage for organizational trust, the foundation of leadership.[9] Organizations that operate under strong data cultures make decisions that are informed by evidence. Executives feel confident to act on analytically derived insights, not just on intuition or personal experience.[10]

Evaluating a team's level of analytical awareness is important to determine how analytically strong the organization is, and its readiness for data culture improvement. Organizations that want to prepare for a successful future must consider its analytical awareness and how improvements can be made. Deloitte identified five levels of analytic maturity, as described in **Figure 24**.[11]

Figure 24 The Insight-Driven Organization Maturity Scale[12]



Data is a combination of ideas, practices, and knowledge that, taken together, generate insights.[13] A positive data culture allows and encourages all levels of employees to make decisions based on data. It is not an easy task, as it is more involved than acquiring the right tools or hiring the right talent. But the investment pays off. An organization is twice as likely to exceed goals when it has a strong cultural orientation to data-driven decision-making.[14] While a company or government agency can have the best technology available, data strategies must also consider the effects of human behavior on the generation of data.[15] Developing and nurturing data culture requires the standardization of behaviors among those gathering, recording, analyzing, and using the data. The hardest part of the data culture journey is changing behaviors.[16]

5.3 Ten Commandments for Creating and Improving Data Culture

David Waller outlined ten “commandments” to create a data-driven culture.[17] While these ten principles are directed to large consumer-facing corporations, they are also instructive in setting the stage for creating a data culture at government agencies, as shown in **Table 6**.

Table 6 Ten Commandments for Creating and Improving Data Culture

Ten Commandments for Creating and Improving Data Culture	
1	**Data-driven culture starts at the (very) top.**
	Ensure that proposals reference data and the follow-up includes taking evidence-based actions that will encourage all team members to do the same.
2	**Choose metrics with care--and cunning.**
	Carefully choose what to measure and how to use the metrics. This allows leaders to identify priorities for gathering data and measuring performance.
3	**Don't pigeonhole your data scientists.**
	Pull leaders toward data science, train employees to be, in some cases, code-literate and conceptually fluent in quantitative topics.
4	**Fix basic data-access issues quickly.**
	Start small by granting universal access to a few key measures at a time, rather than a grand program to reorganize all data. Demonstrating dependence on those key measures will lead team members to request additional metrics.
5	**Quantify uncertainty.**
	Requiring team members to make decisions when there is uncertainty does three things: a. Allows them to express confidence in the reliability, or lack of reliability, of the data. b. When team members understand the metric models, they have opportunities to improve the measures. c. Understanding uncertainty allows team members to run experiments and if/then scenarios.
6	**Make proofs of concept simple and robust, not fancy and brittle.**
	Expensive and complicated processes may not be financially beneficial. Start by building something that is industrial grade but simple before moving on to sophisticated concepts.

7	**Specialized training should be offered in a timely fashion.**
	Offer training and education just before the skills are to be used. Specialized analytical concepts and tooling are more effective right before they are needed.
8	**Use analytics to help employees, not just customers.**
	Immediate goals of saving time and automating easy tasks inspire team members to work with the data themselves.
9	**Be willing to trade flexibility for consistency—at least in the short term.**
	Pick canonical metrics and programming languages. Inconsistencies across an organization requires more training for analysts and takes additional translation time.
10	**Get in the habit of explaining analytical choices.**
	Establish a pattern of requesting teams to explain their approaches, alternatives considered, and tradeoffs, followed by their reasoning for choosing one approach over another. This allows team members to walk through their cognitive processes.

5.4 Case Studies in Implementing and Improving Data Culture

Many federal, state, and local governmental agencies have committed to transitioning to data-driven organizations. Here are a few examples.

Table 7 U.S. Government Data Commitment Example

U.S. Government	
Challenge	**Approach**
To promote job growth, government efficiency, and public access to government data, President Obama signed Executive Order 13642 in 2013, requiring all federal agencies to comply with an Open Data Policy.[18] The Order also required agencies to set incremental performance goals, identifying metrics and milestones.	The Department of Transportation (DOT), as one agency example, uses many transparency portals to provide information on railroad safety, commercial motor vehicle safety, vehicle safety, and pipeline safety.[19]The DOT proactively adds content online in consistent, open formats, while assuring accuracy and protecting privacy, security, and confidentiality.

U.S. Government
Results The DOT released 765 datasets to Data.gov. In May of 2012, DOT, in partnership with the Department of Justice, Department of Labor, and the Consumer Product Safety Commission, launched Safety.Data.gov with 713 datasets, 4 mobile apps, 14 resources and public domain software tools, and 3 competitions. To educate the public, DOT hosts safety transportation *datapalooza* events. The community now has 869 datasets, 10 mobile apps, and 6 competitions encouraging public use of safety data from around the government.[20]

Table 8 State of Colorado Data Commitment Example

State of Colorado	
Challenge To facilitate public access to department budgeting, citizen input into strategic planning, and data-driven decision-making, the State of Colorado created several process improvement initiatives over the course of more than ten years.	**Approach** Several programs were created to address these challenges, including the State Measurement for Accountable, Responsive, and Transparent (SMART) Government Act, the Performance Management Academy, Colorado Results First, Vision 2018, and the Pay for Success program.[21]
Results The Performance Improvement Team supported more than 500 projects in 18 agencies and trained more than 3,000 employees in Lean tools and practices. Results include decreasing driver license waits from 40 minutes to 20 minutes; reducing pharmacist license waits from 188 days to 14 days; and addressing 70% (from 22%) of oil and gas complaints within 30 days. Vision 2018 established a public-facing dashboard so citizens could understand the progress made toward goals. Colorado now conducts cost-benefit analyses for all programs and adjusts to ensure that new requests are grounded in research and sound evaluation designs.	

Table 9 State of Connecticut Data Commitment Example

State of Connecticut	
Challenge	**Approach**
To be more responsive to the public and implement decision-making by data analysis, the Connecticut Legislature codified a state data-driven government and decision-making process for its leadership.[22]	The legislation required a state data plan to establish management and data analysis standards, make recommendations to enhance standardization and integration of data systems and data management practices, provide a timeline and review of any obstacles, and set goals for improving online repository.[23]
Results	
The State of Connecticut set three focal points for 2021-2022, including COVID-19 response and recovery, equity in the data lifecycle, and using data for decision-making. It experienced higher traffic on the data portal due to COVID-19 resources, and included daily updates, dashboards, and data stories. It had more data requests in a six-month COVID-19 period than the first six years of the portal. To determine potential vulnerable populations, data demonstrated that the average citizen was involved with 4-6 agencies. These data helped it identify many service gaps and areas for improvement.[24]	

Table 10 State of Pennsylvania Data Commitment Example

State of Pennsylvania	
Challenge	**Approach**
With a desire for government to be more open and accountable to taxpayers, Governor Tom Wolf identified three specific goals: "Jobs that pay. Schools that teach. Government that works."[25]	The Commonwealth of Pennsylvania established OpenDataPA to allow citizens, entrepreneurs, developers, researchers, and policymakers to make data-driven decisions.[26] The data availability creates transparency, accountability, modernization, and innovation to the government by making government data open and centralized.
Results	
Online reports and key data points are available from 30 commonwealth agencies. One specific dashboard is the Opioid Data Dashboard. It contains sections with data and charts on *Preventing Addiction, Saving Lives, Getting People into Treatment, Community Impacts*,	

State of Pennsylvania
and *Learn More About Opioids*. Another dashboard contains information on transportation, including airports, bridges, traffic patterns, road construction and complications, and related data.

Table 11 County of Miami-Dade Data Commitment Example

County of Miami-Dade	
Challenge	**Approach**
Experiencing urbanization, globalization, and climate change, Miami-Dade County needed to build resiliency from external shocks and stresses stemming from natural disasters, economic challenges, and failing infrastructure.[27]	The County coordinated with cities of Miami and Miami Beach to create a plan to address urbanization, globalization, and climate change. Thousands of stakeholders, including a wide range of expertise, age groups, ethnicities, cultures, income levels, and geographic areas, worked together to identify opportunities for improvement.[28] The biggest concerns were affordable housing, mobility, income inequality, sea level rise, climate change, and health and safety.[29]
Results	
Miami-Dade County's Resilience Strategy identified resilience opportunities and discovery areas, including Living with Water, Advancing & Adapting, Building Prosperity, A Thriving Community, Robust Recovery, and Leadership for Tomorrow.[30] There are 59 Action Items on the list including the *Better Bus Project*, making bus service more frequent and efficient, *Pilot an Arrest Diversion for Opioid Users* (treatment and diversion), *Buy Local* (*Buy Miami* program with website and free advertising for small businesses), and *Teach Kids to Save (*Savings incentives targeting kindergarten children).	

Table 12 City of Boston Data Commitment Example

City of Boston	
Challenge	**Approach**
Looking for better ways to serve his community, Mayor Martin J. Walsh wanted to make civic data easily	Taking advantage of the established library system, *Analyze Boston* was designed as a new open data platform. The goals were to inventory, catalog, and

available to his residents. He built from previous Boston programs like CityScore and 311 but wanted to reimagine the open data experience.	build a comprehensive data catalog, redevelop Boston's open data portal, and develop and deliver an open data curriculum to Boston area libraries.[31] The public was involved in the planning process.
Results Analyze Boston offers search opportunities from 253 datasets from geospatial, city services, finance, environment, permitting, economy, public safety, facilities, and others. In addition to city services, there are 43 organizations with datasets within the data portal. One example is *Crime Incident Reports*, which outline offenses and reporting areas on an annual basis. Another popular dataset is the *Food Establishment Inspections.*	

Table 13 City of Cape Coral Data Commitment Example

City of Cape Coral	
Challenge The City of Cape Coral, Florida was hit hard by the real estate bust and the Great Recession around 2008 and needed to focus on economic development.[32]	**Approach** Although the city had established performance measures and metrics in its day-to-day operations, it was not shared with the public. The City of Cape Coral established open data practices and improved performance management programs.[33]
Results The city enhanced its open data program, established an open data inventory, and centralized the data in an easy-to-find location. It also established internal and external dashboards to continually track progress.[34] Data repositories open to the public include the *311 Call Center, Business & Budget, Planning & Zoning, Public Works, GIS,* and *Fire.* An example of a dataset is the contractors by permit type, showing a list of active contractors and the number of permits issued. The Business dataset includes map visualization and charts with many details on local businesses.	

Table 14 City of Detroit Data Commitment Example

City of Detroit	
Challenge	**Approach**
To increase public access to valuable data and information concerning government operations and services, Mayor Mike Duggan established the City of Detroit Open Data Initiative by Executive Order 2015-2.[35]	Detroit GO DATA was designed to provide a transparent, open, collaborative, participatory, and accountable government for the citizens of Detroit.[36] A Task Force and Advisory Commission were created to evaluate and determine the best methods to design, implement, and monitor the project.
Results	
The citizens of the City of Detroit have online access to government, public safety, property and parcels, education, transportation, public health, boundaries, and permit data. Open data tools include *open checkbook* with easy access to vendor spending; restaurant inspections, crime viewer, COVID-19 spending, neighborhood tracker, parcel viewer, rental map, and waste pickup reminders.[37]	

Table 15 City of Fort Lauderdale Data Commitment Example

City of Fort Lauderdale	
Challenge	**Approach**
To prepare for upcoming challenges and anticipate new opportunities, the City of Fort Lauderdale created a vision for its future following its Centennial in 2012 and called it *Fast Forward Fort Lauderdale*.	The city reached out to the public and collected 1,562 unique citizen ideas obtained through Stakeholder Interviews, Open Houses, Telephone Town Hall Meetings, Meetings-In-A-Box, its OurVisionFTL.com website, Big Ideas Fort Lauderdale 2012 event, and a Neighborhood Summit."[38]
Results	
As part of *Fast Forward Fort Lauderdale 2035,* Fort Lauderdale improved performance management programs and used open data for progress.[39] The public has online access to the Vision Scorecard, a living document that tracks progress continually. The six high-level key indicators are "we are connected, we are ready, we are united, we are prosperous, we are community, and we are here,"[40] and are strategically aligned with measures that citizens can track.	

As these cases illustrate, efforts by government agencies to communicate with the public are essential. These examples include federal (**Table 7**), state (**Table 8, Table 9, Table 10**), county (**Table 11**), and city (**Table 12, Table 13, Table 14, Table 15)** examples of valuable data made available to the public. Many of the agencies involved citizens in their decision-making process, ensuring that the public could benefit from the efforts made by all parties.

5.5 Importance and Value of Data

Streamlining government services, stimulating economic opportunities, encouraging innovation, improving public safety, and reducing poverty are just a few of the benefits of open data.[41] Organizational culture allows team members to shape those concerns into their strategies,[42] and it influences many aspects of organizational leadership, including human behavior, motivation, knowledge transfer, team work, and collaboration.[43] All of these factors are necessary to create a successful organization and facilitate authentic and sustainable change. These facts apply to private corporations and government agencies, both of which are concerned with budgets and spending, sustainability, stakeholder value-adding, and fulfilling social and public needs.[44]

Investing in technology is necessary, but the investment cannot stop there. Data needs to be seen as a sociotechnical subject. Organizational survival today mandates that staff at all levels of the organization understand the role and importance of data. Furthermore, it is imperative to interact with end-users of data, who too often are not a priority, leaving out key components of a successful process.

As residents are dependent on the government for public safety, economic development, and critical health services, it is important to include, from the beginning of the process, the voice of the community members who will ultimately be impacted by the resulting decisions.[45] Data-rich information shared between government agencies, business leaders, and community members helps everyone understand how to better manage social and economic challenges. A Pew Research Center survey found that Americans are more likely to trust research and findings when the data is openly available.[46]

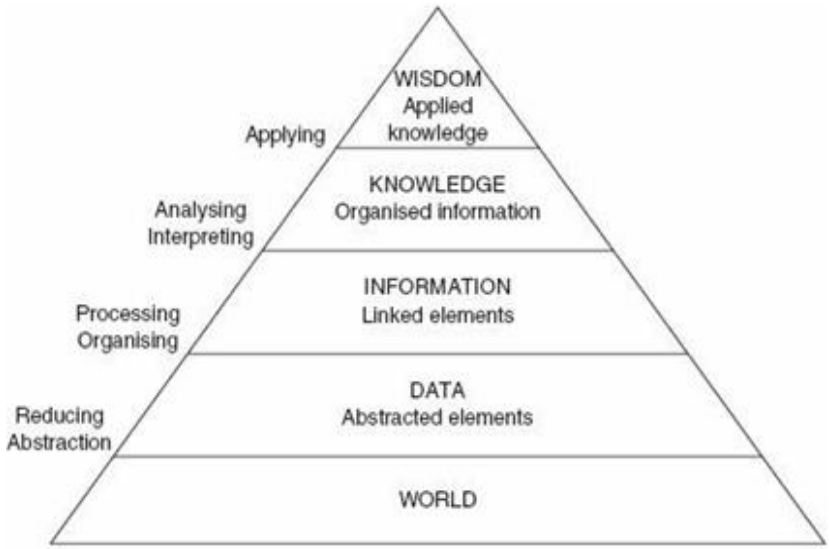
Data have been used to create innovations, build products, establish policies, and contribute to knowledge that shapes how people live their lives.[47] However, it is the effective use of information that creates the genuine victory. Determining effective measures and knowing how each measure fits within the organizational

schematic, allows leaders to work toward realistic data dependence. Moreover, data reliability permits leaders to accurately review the past, compare the present, and plan better for the future, across time, space, and scale.[48]

Informed decision-making is facilitated by the collection of data and the capabilities of technology. As decisions are made throughout the organization, insight from all levels is needed. Success relies on a willingness to act on analytically derived insights, not intuition.[49] This mindset needs to be applied by everyone in the organization, as this approach requires a shared vision. While this may seem like an insurmountable task, there are successful analytical tools and processes to help organizations develop more robust data cultures. For too long, organizations have depended on *the way it's always been done*, even though technology has enabled them access to statistical information. A data culture will support, encourage, and authorize sincere analytical processes.

The *Knowledge Pyramid* illustrated in **Figure 25** demonstrates the progression of data. Data elements that were previously determined to be meaningful in some ways are transferred into information. Using that information, knowledge is created. That knowledge transfers into organizational wisdom, as leaders make significant decisions for the successful future of the organization. The pyramid's layers demonstrate the process of distillation, reduction, abstraction, processing, organization, analysis, and interpretation, revealing data relationships and truths.[50]

Figure 25 The Knowledge Pyramid



5.6 Avoiding Biases and Establishing Trust

John Maxwell, author and advisor to Fortune 500 companies, declared that change is inevitable, but growth is optional.[51] As changes happen in society and technology, it is imperative to create a positive data culture so leaders can make fact-based decisions. Unfortunately, organizations often fall back on habit, especially when the alternatives look too risky.[52] Establishing a data culture where team members understand data collection and metrics allows organizations to move forward in the direction the data is pointing. Using data properly can provide evidence to back up hypotheses, allowing managers to have confidence to expand into new areas and processes.[53]

Biases, whether explicit or implicit, have been shown to affect decision-making, even by experts in many fields.[54] Decision makers are not objective when they make choices because biases interfere with objective rationality.[55] Here are a few ways biases can interfere with effective management.

5.6.1. Illusion of Control

Sometimes people believe they can influence an event when they have no control over the outcome. This overconfidence allows people to ignore risks while avoiding objective evaluation. Managers need to be realistic about their strengths and weaknesses and listen to advisers who can point out realistic consequences.[56]

5.6.2. Framing Effects

Some decisions are made based on whether the options are presented positively or negatively, e.g., as a loss or as a gain. Leaders need to be able to clarify subjective influences and evaluate the organization's willingness to avoid or seek risks.[57]

5.6.3. Discounting the Future

Relying too heavily on short-term gains without considering the benefits of longer-term decisions can allow leaders to harm their teams and organizations. Investing in research and development, updated technology, and strategic planning are essential to the success of every organization.

5.6.4. Confirmation Bias

The tendency to favor information that supports personal beliefs may deflect a leader's objectivity and direct attention to search for and interpret information to support existing beliefs. It is hard to challenge long-held

> beliefs and natural intuition, but facts allow leaders to have a more realistic understanding of the organization's needs.

The biases listed are just a few examples of situations that can prevent objective decision-making. However, biases can be mitigated with a committed reliance on data and analytics. True, committed reliance on data and analytics has the potential to counteract biases and facilitate a more solid foundation for the organization.

It cannot be emphasized enough that this data culture must be implemented organization-wide. Davenport, a professor and author specializing in analytics, business process innovation, knowledge management, and artificial intelligence, wrote:

> In creating a data-driven culture, there's no rest for the weary. We know of organizations that were hugely focused on data and analytics, but when the CEO champion left, they drifted back to their old gut-based thinking and decision-making. From boards of directors to CEOs to analytics and AI leaders, everyone who believes in this focus should work to persuade others to adopt and maintain it. No one should assume that software and hardware alone will lead the organization to the cultural promised land.[58]

It is unreasonable to discuss data and not include the value of trust. Neider, Professor and Chair of the Management Department at the University of Miami, described this era as an age of pervasive distrust with a *trust deficit* that affects all organizations.[59] Opinion polls illustrate deep distrust towards political leaders, government organizations, and businesses.[60] Governments operate in the public arena where actions are criticized by the public, with or without facts. This level of societal frustration demands the attention of leaders on every detail, as understanding motivation and perceptions is essential.[61]

Organizational trust includes trusting data and the analytics powering its technology. Trust is considered a defining factor in an organization's success or failure, relating to reputation, customer satisfaction, and loyalty. The presence of trust can inspire employees, reduce uncertainty, and build resilience. The need for trust is so critical that more than half of CEOs believe building trust is one of the top priorities for their organizations.[62]

5.7 How to Develop and Nurture Data Culture in Your Agency

In consideration of the theory and concepts presented, **Table 16** offers a survey for evaluating and moving your organization from present-day old habits to a more engaged data-driven culture.

Table 16 Developing and Nurturing Data Culture

A.	**Assess current data culture and capability.**
	First, do you and your team understand the importance and value of data?
	Can you identify cultural influences on leadership, individuals, and individual behavior, motivation, knowledge transfer, teamwork, and collaboration?
	What positive strides have you already made?
	What opportunities for improvement are there?
B.	**Appraise analytical awareness.**
	Can you identify how *analytically aware*[63] your organization is by using **Figure 24**? It may be helpful for leadership team members to perform this exercise separately, and you may want to look at different areas independently.
	Divide the organization systematically on paper and review specific areas. Where do you want to start?
	Regarding specific team members, what are the training needs? It may prove enlightening to review team members' thought processes as you discuss the results.
C.	**Establish data infrastructure: Apply the 10 data commandments.**
	There are many examples of organizations, public and private, that realize the necessity of a strong, positive data culture. In today's transparent world, it is critical that every person in the organization commit to a data-driven culture. Take time to go through the *10 data commandments*[64] for your organization.
	1. Data-driven culture: a. Is your organization committed to fostering a strong data-driven culture? b. Are your leadership team members committed?
	2. Metrics: a. What metrics do you need to measure? b. What predictions do you have and how will you track the predictions and the results?

3. Data Scientists:

a. Are you willing to allow your data scientists and staff members to communicate directly with users?
b. Would it benefit your team to become code-literate?
c. Is training needed?

4. Data Distribution:

a. How is data distributed?
b. Do team members have access to basic data?
c. What data is most meaningful to team members?
d. How can you improve access for all team members?

5. Quantify Uncertainty:

a. Require your team to be explicit and quantitative about their levels of uncertainty. Force decision-makers to determine if the data is reliable. What could be the missteps? What are the dynamics?
b. How can analysts gain a deeper understanding of their models so they can tweak where necessary?
c. Emphasizing uncertainty pushes team members to run experiments and look at different angles. What is an example that you have found?

6. Proofs of Concept:

a. Considering the metrics, remember that promising ideas outnumber practical ones. Can you create *proofs of concept* that are simple and robust, saving fancier versions until after all requirements have been gathered from all main stakeholders, including end users?
b. Once a simple process works from the first step to the final step, improvements can be made to create more complicated runs, greater volumes, and faster runtimes. Identify the first project.

7. Specialized Training:

a. Can you plan to offer specialized training when team members can use it right away to avoid analytical concept retention problems?
b. How can you prioritize the various team members or divisions?

8. Analytics:

a. Could you use analytics to help your internal customers, as well as your external customers?
b. Empowering team members to become more data-fluent, including understanding coding, could help them better comprehend the insights provided by the data and lead to a more robust technical base.

9. Consistency:

a. Are you willing to trade flexibility for consistency?
b. Will your team agree to new programs for the betterment of the organization?
c. Does your team realize that working as a team, rather than in silos with competing lists, allows a more consistent collection and analysis of meaningful information?

10: Develop New Habits:

	a. Can you create a business process that makes explaining analytical choices habitual and routine? b. Can you make this habit routine so others can understand thought processes and participate in deeper discussions?
D.	**Review the Case Studies.**
	Do you have similar challenges to any of the government agencies identified in the case studies section? More details are available online for each of those jurisdictions and discussing challenges with colleagues who have travelled the journey is always helpful.
E.	**Evaluate data progression and pathways.**
	Considering the Knowledge Pyramid[65], can you measure how your data moves from the real world up the pyramid through collection, organization, and processing?
	Can you identify the relationships that happen along the way to create beneficial information?
	Can you determine how to analyze the data and interpret the information to generate organizational knowledge?
	Can you identify the organizational wisdom and see how it leads to better processes and decisions? You may choose to perform this process for specific areas separately.
F.	**Set expectations, making data-driven decisions routine.**
	What will it take to be able to rely on your data, and become a true data-driven culture?
	Can you avoid the temptation to return to old habits or to rely solely on experience?
	How can you create expectations for all team members to make decisions based on data?
G.	**Perform data quality assessments.**
	Can you avoid biases and establish trust?
	Can you put measures in place to avoid returning to old habits, leading by example?

5.8 Chapter 5 Summary

Improving data culture and establishing data-driven methodologies can be overwhelming, but avoidance can be toxic to an organization. Data cannot be kept in a sequestered room with data scientists, as research demonstrates significant advantages to involving all levels of the organization. Taking data from raw numbers to information to knowledge to wisdom allows organizations to apply invaluable advances to its workflow and business processes. Moving from *analytics aware* to *analytical competitors* will assist leaders in managing performance, preserving value, making decisions, and safeguarding sustainability.

Building a data-driven culture can be guided by Waller's Ten Commandments.[66] As in any organizational change, the message has to be consistent throughout the organization, keeping in mind that human behavior, motivation, knowledge transfer, teamwork, and collaboration must be valued and handled with care. This process is not an overnight change, but, rather, a journey of discovery for all team members. Finally, trust is critical to the success of all organizations. Establishing a safe environment with committed values will help facilitate data-driven decisions and successful pathways.

Chapter 6: Data Visualization and Publication

The volume of data used in business, research, and technological development is massive and continues to grow. However, as access to data grows exponentially, it becomes harder for users to quickly grasp the core message that the data is attempting to convey. When presenting complex information, choosing the right visualization is critical to the success of a presentation. The best data in the world will not matter to the audience unless it is presented in a way that people can comprehend and appreciate. In this chapter, considerations are discussed for creating effective data visualizations and establishing rules for the publication and suppression of data. These guidelines are suggested to help readers summarize and present large datasets in simple and easy-to-understand visualizations that provide insightful information for the target audience. Topics include data visualization and its importance, general guidance for data visualization and publication, and recommendations.

6.1. Why is Data Visualization Important?

Data visualization is important because it allows individuals to visually process large amounts of complex data using graphical tools. By doing so, it makes the data easier and, more importantly, quicker to digest. This is because data visualization allows for the data to be presented in a manner that allows the audience analyzing it to sift through large datasets simply by looking at visuals such as charts, graphs, or infographics.

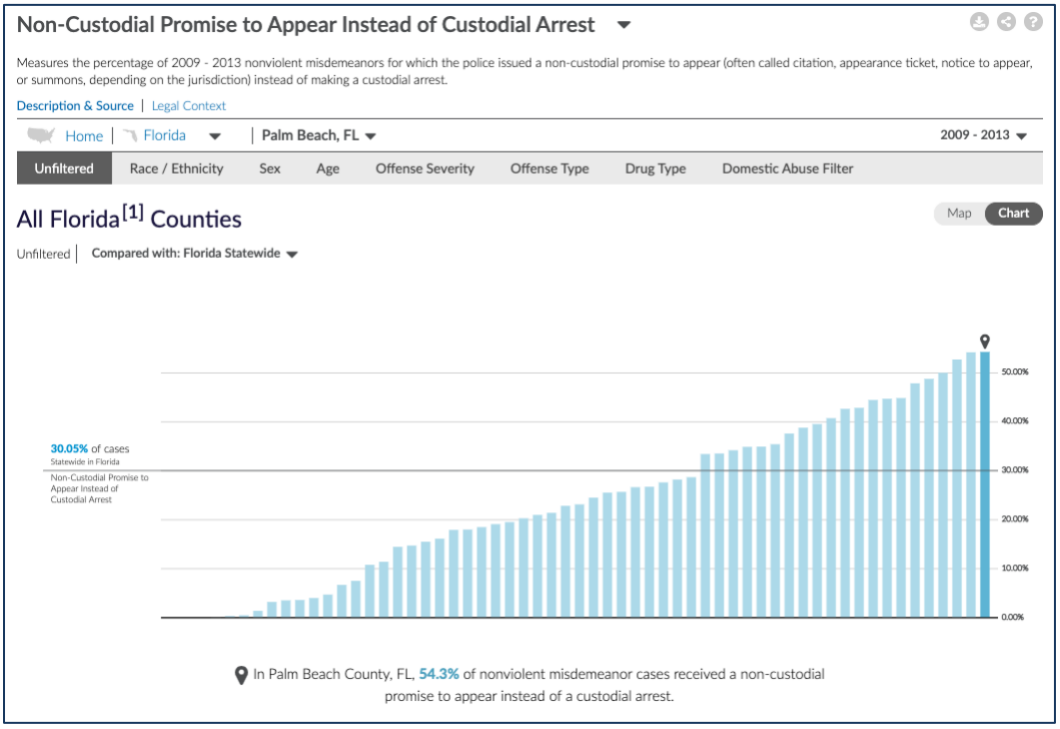
By using these visualizations, the audience will be able to quickly spot a pattern or trend in the data rather than being overwhelmed with long descriptions or multiple pages of data on a spreadsheet. A good visualization should summarize and organize information in such a way that anyone reading it can focus on the relevant points that the data is trying to convey. Simply put, data visualization creates a way to represent data that allows for the data's meaning to be communicated clearly and effectively. Below are some considerations as to why data visualization is important.

6.1.1. Visualizations Amplify the Message

As previously mentioned, a good visualization takes complex data and organizes it in a way that allows its audience to understand what is being presented. Without visualizations, the presenter may run the risk of the data not being understood or perceived as meaningless as the message could be lost on the reader. The purpose of gathering data is to gain valuable insight and facilitate better results for the reader.

Figure 26 is an example showing a comparison of one county in Florida, Palm Beach, to the rest of the state in terms of its use of citations instead of arrests for low-level crimes. In a single chart, the user can see that Palm Beach had the highest rate of citations instead of arrests for nonviolent misdemeanors among all Florida counties for the time period 2009-2013, that this rate varies dramatically across counties in the state, and that Palm Beach surpasses the state average by about 24 percentage points.

Figure 26 Visualization Example[67]

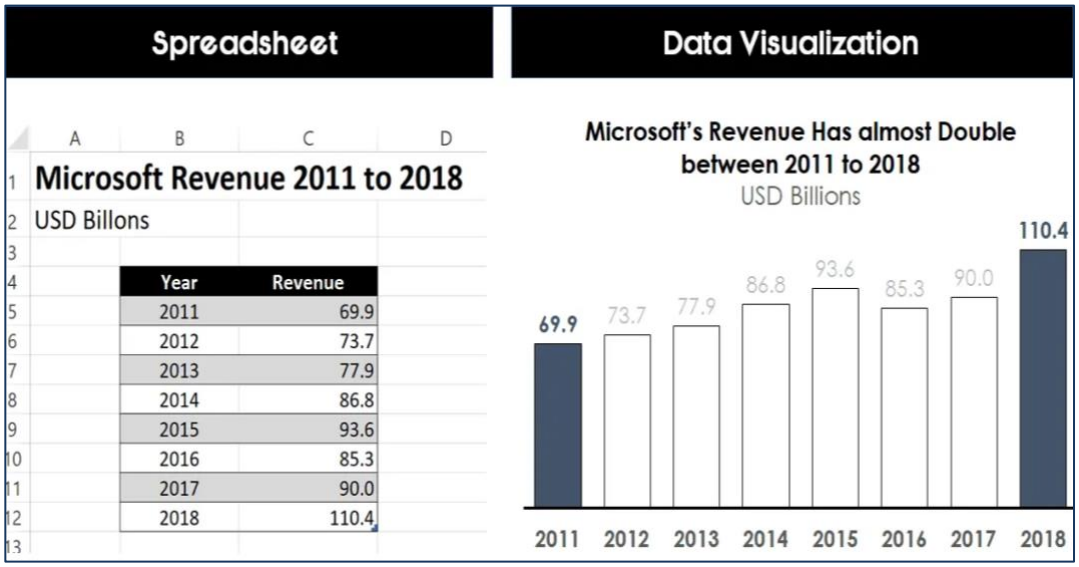


6.1.2. Visualizations Save Time

By using a graphical representation, organizations can view and share large amounts of data in clear, cohesive ways with their stakeholders, and quickly identify trends from the information being presented. Because it is significantly faster to analyze information in graphical form, individuals analyzing the data can address problems and answer questions in a timelier manner.

For example, a spreadsheet showing that revenue has almost doubled between 2011 and 2018 is not nearly as visually insightful as that same data in a simple bar graph with some formatting, as shown in **Figure 27**.

Figure 27 Spreadsheet Compared to Data Visualization[68]

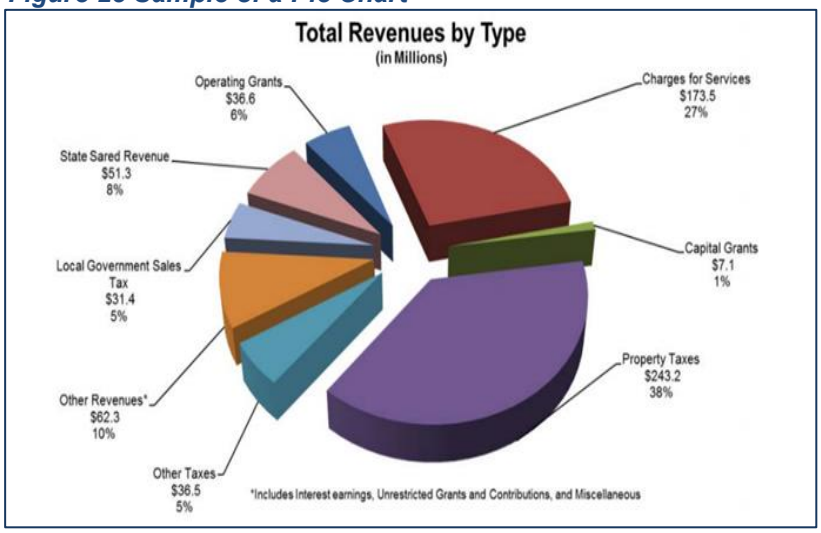


6.1.3. Visualizations Assist with Decision Making

If data visualization is done properly, an organization and its stakeholders will be positioned to make meaningful decisions quickly based on data that is better understood.

For example, the pie chart in shows the total revenue received by a governmental organization by individual revenue type. This chart helps provide a clear picture as to which revenue type is providing the greatest return so that business leaders can leverage this information to produce more revenue.

Figure 28 Sample of a Pie Chart[69]

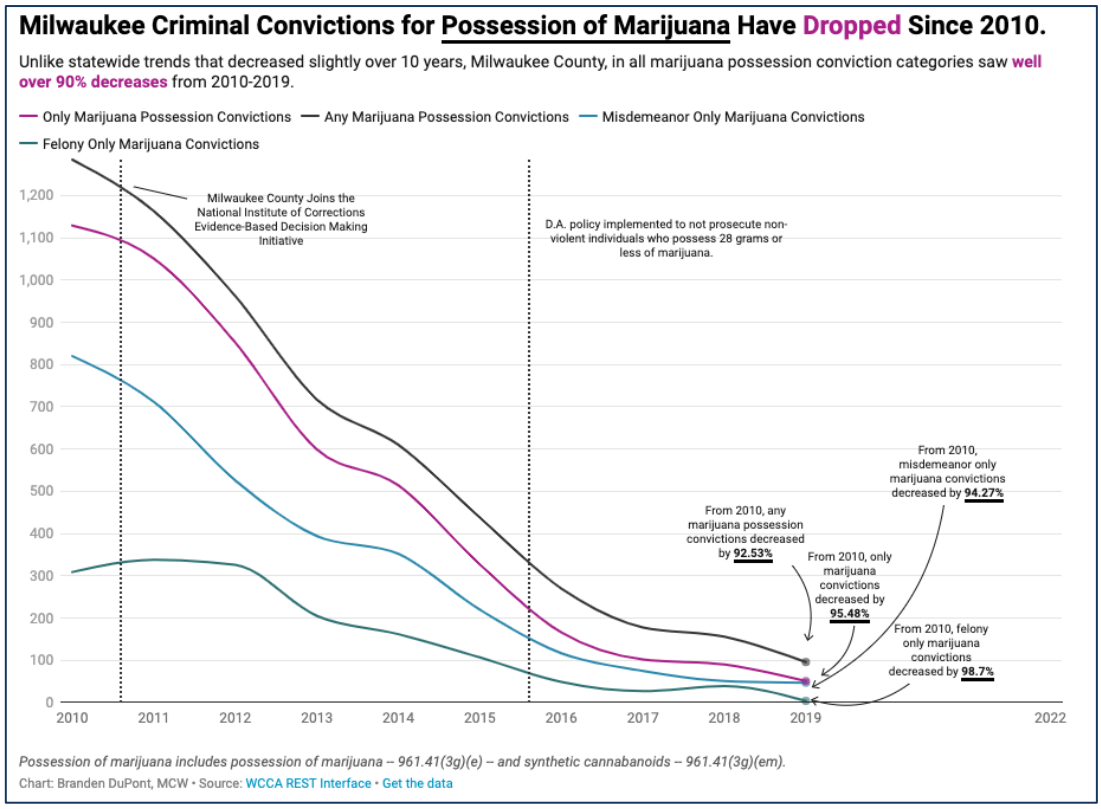


6.1.4. Visualizations Communicate the Data Story

Visualized data allows new insights to be gathered from data that may have otherwise gone unnoticed. The next step is to communicate those insights to others, and visualizations can make it easier to communicate the data's story across large audiences. Charts, graphs, infographics, maps, and other visually impactful representations of data are crucial when communicating the story behind the data.

Figure 29, below, is an example of a time series trend graphic showing the decrease in marijuana possession convictions in Milwaukee County since 2010. In addition to the trends, the graph is also depicting policy inflection points that had an effect on marijuana possession convictions, telling a more complete story about the trends and patterns observed.

Figure 29 Example of Time Series Trends[70]



6.2 General Guidance for Data Visualization and Publication

The specific visualizations presented below are intended to be a demonstration of different ways to implement this guidance with a focus on accessibility and transparency.

6.2.1 Getting started with data visualization

When preparing to implement any type of data visualization, there are steps that should first be considered. Not only is a solid grasp of the data required, but it is also crucial to understand the goals, requirements, and the audience that will be consuming the data. Preparing for data visualization implementation begins with the following:

- Understand the data, its size, and level of individuality (the uniqueness of data values in a column). A high level of individuality means there is a large percentage of unique values (e.g., defendants' case numbers, because each item should be unique). Low levels of individuality means a column of data contains a large percentage of repeat values (as might be seen in a *gender* column).

- Know who the audience consuming the data will be and understand how they process visual information. By doing so, the organization will be able to determine what the visualization should portray and how to communicate the message to their target audience.

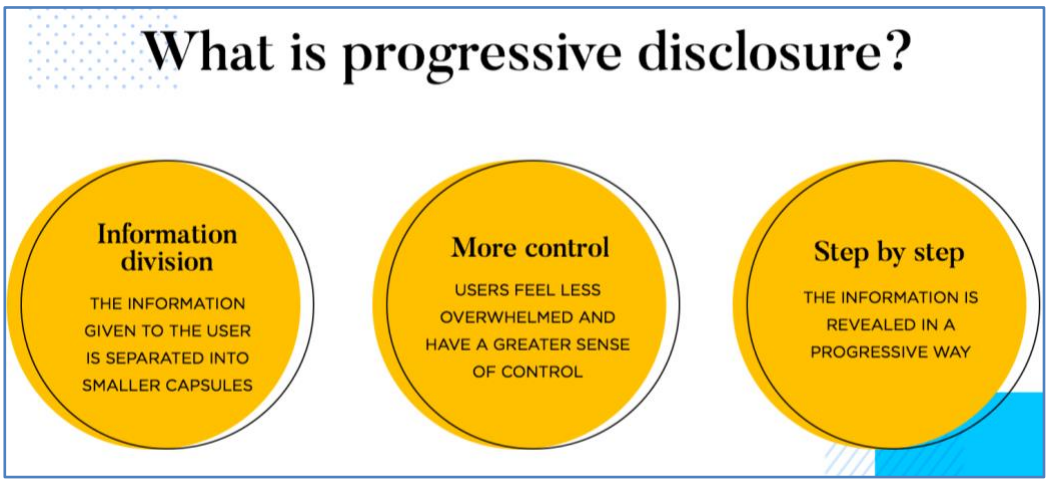
- Use simple visuals that express the information the data is trying to communicate.

- Determine the amount of data needed to present the points desired.

6.2.1. Simplifying the data using progressive disclosure

Websites today are more complex with new content and features being added every day. However, at times it may seem that these new complex features act as more of a hindrance than a convenience to the user. This is because the more complicated an interface becomes, the more effort a user must invest when interacting with it. It is vital that users are led through the process of interpreting the data, as too much information or too many choices can make the user feel overwhelmed in the process. Users do not want to spend their time trying to learn how the multitude of features work; they simply want to complete their tasks with little to no effort. In these instances, simplicity always wins. Fortunately, that content can still be demonstrated to users in a step-by-step fashion using the process of progressive disclosure as shown in **Figure 30**.

Figure 30 Progressive Disclosure[71]



Progressive disclosure makes the user's experience easier by only showing the user data that is relevant to them, while hiding other details until the user needs to see them. There is a general assumption that, by

initially focusing a user's attention on a few core features that they may be unable to understand all the data's value. However, research makes it evident that this assumption is incorrect.[72] When content and features are prioritized, the users will spend more time on the most important items leading them to form a better understanding of the data and its message.

Essentially, progressive disclosure helps by prioritizing a reader's attention and allows the reader to take more control of advanced aspects of the data. It makes the experience more efficient for the user and reduces the learning curve needed to navigate the data.[73]

6.3. General Guidelines for Progressive Disclosure

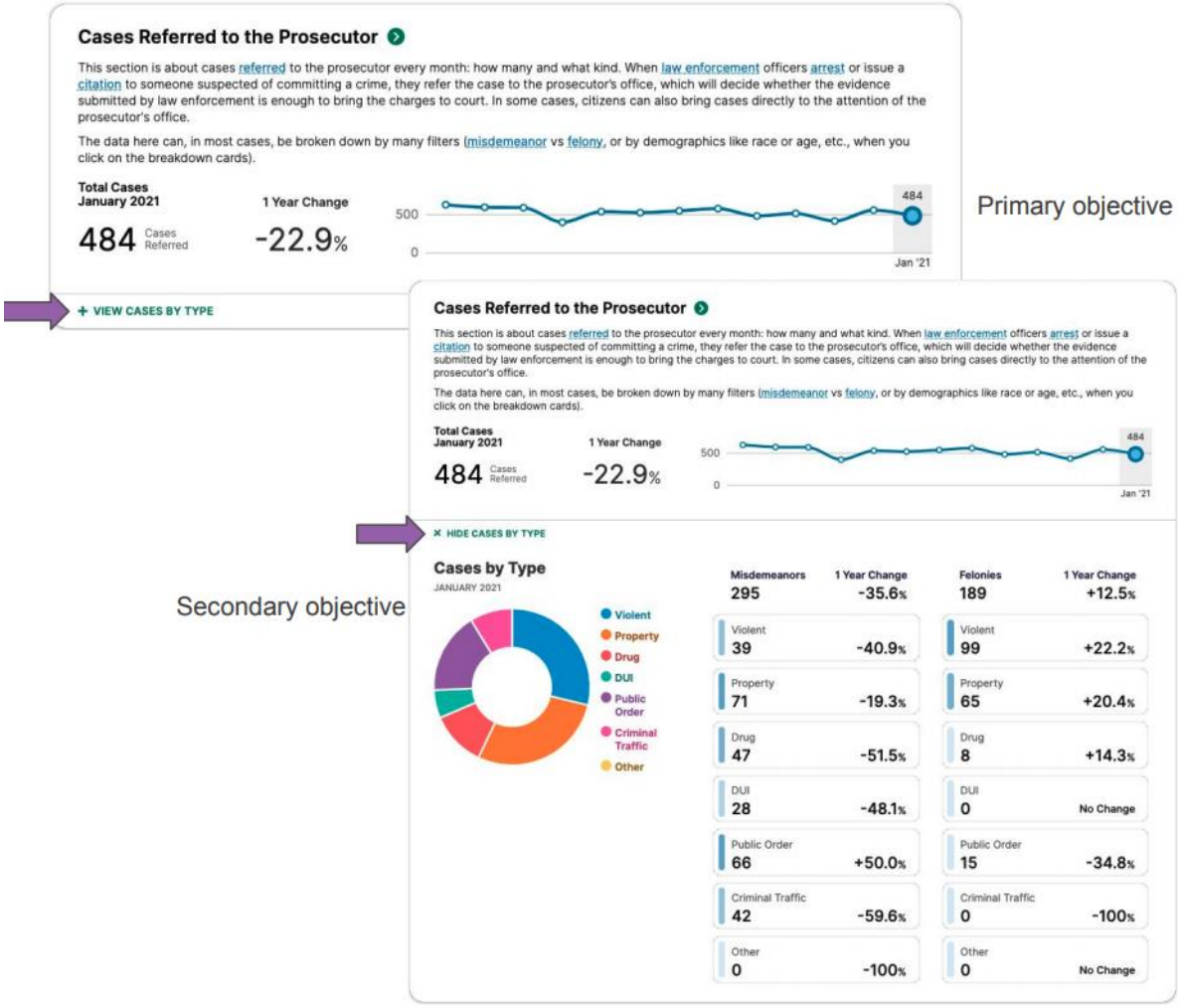
Along with providing users with an opportunity to only see the relevant information at each step of their analytical journey, the general interface of the visualization should also have a clean appearance and provide its users with the basic tools necessary to comprehend the data they are reviewing. Below is a non-exhaustive list of guidelines to consider when exploring progressive disclosure:

6.3.1. Primary and Secondary Objectives

When creating a visualization using progressive disclosure, content should be prioritized into primary and secondary objectives. This is to ensure that the user's attention is focused on the data that matters most to them. By prioritizing information into these two objectives, it allows for the user to be presented with the core information first. From there, users are presented with the option to further venture into more complex data through the use of secondary objectives, if needed.

For instance, **Figure 31** shows the progressive disclosure of cases referred to a prosecutor's office by law enforcement. The primary objective is to present the user with a monthly trend of the number of cases received by the prosecutor's office, while also providing information about changes in the volume of cases compared to the same month a year prior. If the user wants to see more details about the types of cases referred to the office, they can click on the "+ view cases by type" feature to discover the secondary objective. In this view, users can see more details about the specific type and severity of the cases the prosecutor received from law enforcement.

Figure 31 Primary and Secondary Objectives[74]



This way, rather than overwhelming the user with a wall of data at the first point of interaction, the user is afforded the opportunity to first digest the primary information while actively deciding if more information or content is required.

6.3.2. Transitions between objectives should be effortless.

As discussed, primary and secondary information should be separated and only the information that is necessary or relevant to the task at hand should be presented. However, it is also important to provide a simple and easy to use mechanism that will allow users to obtain secondary information effortlessly upon request, as shown by the "+ view cases by type" feature in **Figure 31.**

It is crucial that users are able to quickly find the secondary level of disclosure. Items like buttons, links, or modal windows, can help the user anticipate what information they will be able to obtain once they venture into the secondary level.

6.3.3. *No more than two levels of disclosure.*

To stay true to the concept of progressive disclosure, it is best to have no more than two levels of disclosure. Having more than two levels of disclosure can generally cause problems when a user is analyzing data. This is because including multiple levels of new information can ultimately lead to the user becoming disoriented, or worse, to lose sight of the message the data is trying to convey.

6.4. Deciding Which Visual is Best

One of the biggest challenges for an organization is deciding which visual should be used to best represent the information, as the decision-making process can be daunting. Knowing who the target audience is and the message that needs to be communicated to them will help pinpoint the appropriate data visualization type to use. Asking the right questions ahead of time can help with this decision-making process.

To go further into detail, here are five concepts to consider:

6.4.1. *What story is the data trying to tell?*

Data visualization is the process of taking data and presenting it in a way that tells its audience a story. By using the data to tell a story, the user will gain more valuable insight that they would have by simply looking over a spreadsheet. When data-driven storytelling is considered, it can be a powerful tool because it uses a narrative to put numbers and analytics into context in a quick and easy to read format.

Furthermore, by asking what message the data needs to convey to its audience, the presentation of the data will allow the audience to engage with it in a way that goes beyond what simple facts can show.

6.4.2. *Who is the audience?*

Understanding who the audience analyzing the data will be is another important factor to consider when picking the proper data visualization

type. By doing so, data visualization may be used to communicate a set of trends or forecast insights that would be specific to the audience analyzing it. Knowing who the target audience of the data is will result in a visualization type that will yield the biggest impact on its audience.

6.4.3. What type of data are being visualized?

Because each data visualization endeavor is unique, different chart types will suit different goals and objectives. Bar charts are good to compare a given data point across different categories (see **Figure 26**). Pie charts, donut charts, and treemaps can be used to show the composition of a given data point by breaking it down by groups (see). Line or trend charts are used to display time series data, such as, say, the yearly percentage of cases resulting in conviction across 10 years (see **Figure 29**). Maps are the best way to display geographic information. Scatterplots help visualizing relationships between two data points. In sum, effective visualization doesn't just require a deep understanding of the data, but also of the best ways to graph any given data type.

6.4.4. What is the composition of the data?

Data in the aggregate often times hides more nuanced patterns. In the interest of full disclosure, users should be provided with the ability to "slice and dice" the data in multiple ways to understand its composition. The visualizations for the "secondary objective" in

Figure 31, demonstrate how to provide the users with more information about the types and severity of cases received by the prosecutor's office, allow the user to look beyond just the monthly volume of cases.

6.4.5. What are the Key Performance Indicators (KPIs)?

Key performance indicators are useful tools for tracking progress toward an intended goal. They serve as a tool to help focus on operational improvements by offering a foundation for making well informed decisions and helping focus attention on the most critical issues.

Determining how to present KPIs is crucial as it will help determine how clear the visualizations or data-driven stories will resonate with the audience. It helps to consider what information will be needed from specific KPIs within the project.

6.5. Data Publication and Suppression Rules

Data misinterpretation and misuse is a real risk of data transparency. Organizations putting out their data for public use should be mindful of when NOT to publish information that could mislead the public or that is unreliable. Below is a list of rules to consider when deciding whether to publish particular data points or not:

6.5.1 Sample Size

When the amount of available data to be collected and used for statistical analysis is small, the indicators created using that data tend to have poor reliability. A common rule of law is to only publish measurements, such as percentages, rates, averages, and medians, when there are at least 30 cases to generate the calculation.

6.5.2 Privacy

Another important reason for suppressing particular data points is to protect the privacy of individuals whose data may be included in a report by reducing or eliminating the risk of disclosing their identity and personal information. Aggregating data to higher units, for instance a city or county, is a common practice used to preserve the privacy of individuals included in the data.

6.5.3. Missingness

Datasets or data elements that have a large relative proportion of missing values are problematic, especially when such missing values are not randomly distributed and are caused by unseen factors. High missingness reduces the reliability of the data. When publishing data, organizations should suppress any indicator that is missing more than 10 percent of values, and should always provide users with information on the missingness rate for each data point that is visualized.

6.5.3 Missingness Bias

Missingness Bias refers to the estimate amount of bias that may result from missing data. The bias depends both on the percentage of missing data and the actual value of the measure being estimated. For example, in a county where the pretrial diversion rate is low (e.g., 3%) and there is a considerable proportion of cases missing data (e.g., 7%), the estimate of the pretrial diversion rate could be inaccurate.[75] When an organization

finds a significant missingness bias in their data, they should suppress that data point from publication as well.

6.6. Chapter 6 Summary

In today's "big data" world, there are great opportunities for using data visualization to help diagnose and solve all kinds of social and business problems. Data visualization lets users see things that may otherwise go unnoticed. That is because effective data visualization is the crucial final step of data analysis. It is not a matter of creating a single, optimal visualization; it is a matter of choosing a group of visualizations that will help provide more information for the individuals viewing them. By looking at data from a new point of view, it can more quickly reveal the root causes behind issues that an organization may be encountering on a regular basis. When data visualizations can be used to plan and proactively solve issues within an organization, they can truly make a difference in the presentation of data to the public. Visualization of data can demonstrate patterns and relationships, allowing organizations to identify successes and challenges.

Conclusion

The Florida Legislature took an ambitious step towards data transparency with the passage of the Criminal Justice Data Transparency (CJDT) bills. Lawmakers showed the nation their commitment to improving transparency within Florida's criminal justice system by dedicating themselves to improving its efficiency, accountability, and transparency. Their goal was to equip policymakers with the information necessary to make impactful changes in the lives of their constituents and communities.

This bold action has been met with steady resolve on the part of the statewide and local agencies tasked with the challenge of implementing the legislation. Through workshops and workgroups focused on self-assessment, as well as coordination with the Florida Department of Law Enforcement (FDLE), the pilot counties were able to help progress the initiative and share knowledge and tools with peers. This standard operating procedure (SOP) documents those lessons, supporting the effort to help others within the state achieve CJDT compliance. Additionally, this SOP can also serve as a guide for others aspiring to initiate similar projects within their own communities.

With that in mind, here is a recap of the main lessons covered within this document. Regardless of the magnitude of a data collection project, one must work hard to address the project scope and establish a base level of uniformity. These objectives can be achieved through modeling and consensus building with stakeholders. The effectiveness of these early steps heavily influences the success of the project by identifying challenges and defining the desired outcome.

Once a project is clearly defined, the heavy lifting can begin. Data must be audited to ensure quality, with processes in place to maintain their accuracy and completeness. The specifics of data collection detailed throughout these chapters are not necessarily unique to CJDT. They can be applied to any reporting requirement and thus create a standard for collection for current and future projects. Additionally, investing in tools that automate the collection process not only makes data collection easier, but, when agnostic, furthers the uniformity objective.

Collected data is the foundation. The conversion of that data into information actuates change. Strong data culture within an organization breeds ingenuity, viability, and availability of information to decision makers. Individuals and teams well versed in data will create visual tools that inform decisions, justify perceptions, and educate the audience. These are the ultimate goals of all data collection projects and the metrics upon which success is measured.

CJDT strives to turn data into information, to turn information into understanding and to leverage understanding to make decisions that affect change from a place of discernment. However, it all begins with data. No data, no change.

References

Allaire, Y. & Firsirotu, M. E. (1984). Theories of organizational culture. *Organization Studies*, 5, 193-226.

Avila, A. (2018, January-February). Rocky Mountain Evidence: Using data to drive Colorado State Government. *Public Administration Review*.

Bateman, T. S., Snell, S. A., & Konopaske, R. (2020). M: Management (6th ed.). McGraw Hill.

Bloomberg. (2020). *What Works Cities.* https://whatworkscities.bloomberg.org/cities/

Bloomberg-Cape Coral. (2020). *What Works Cities*. https://whatworkscities.bloomberg.org/works-city-feature-cape-coral-fl/

Bowker. L. (2005, March). Productivity vs quality? A pilot study on the impact of translation memory systems. *Localisation Focus*, 4.

Chisholm, J., & B. DuPont (2021, March 22). *10 Years of Wisconsin Possession of Marijuana Convictions.* https://data.mkedao.com/marijuana-convictions/

City of Boston. (2017, April 11). *Open data to open knowledge.* https://www.boston.gov/innovation-and-technology/open-data-open-knowledge

City of Detroit. (2019). *City of Detroit Open Data Portal*. https://data.detroitmi.gov/pages/about-detroit-open-data

City of Fort Lauderdale. (2013). *Fast Forward Fort Lauderdale: Our City, Our Vision 2035.* City of Fort Lauderdale. https://www.fortlauderdale.gov/home/showdocument?id=4202#:~:text=Fast%20Forward%20to%20Fort%20Lauderdale,visualization%20of%20life%20in%202035.

City of Fort Lauderdale. (n.d.). *City of Fort Lauderdale website. Vision Scorecard.* https://www.fortlauderdale.gov/departments/city-manager-s-office/structural-innovation-division/vision-plan/vision-scorecard

Commonwealth of Pennsylvania. (2020). *Governor Tom Wolf: OpenData.PA.* https://www.governor.pa.gov/newsroom/what-is-open-data-and-common-questions/

Connecticut Public Act No. 19-153.(2019). *An act concerning interagency data sharing*. https://www.cga.ct.gov/2019/ACT/pa/pdf/2019PA-00153-R00SB-01103-PA.pdf

Connecticut. (2020, December 17). *Connecticut state data plan.* https://portal.ct.gov/-/media/CT-Data/Connecticut-State-Data-Plan-summary-2020.pdf

Criminal justice data collection. 900.05 §4 et seq. (2018). Fla. Stat.

Criminal justice data transparency. 943.6871 §(1) et seq. (2018). Fla. Stat. Statutes & Constitution :View Statutes : Online Sunshine (state.fl.us)

Data Driven Detroit. (2020, November 5). *Open data portals.* https://datadrivendetroit.org/blog/2020/11/05/open-data-portals/

Davenport, T., Guszcza, J., Simth, T., & Stiller, B. (2019, July 25). *Analytics and AI-driven enterprises thrive in the Age of With*. Deloitte Insights. Retrieved from https://www2.deloitte.com/us/en/insights/topics/analytics/insight-driven-organization.html

Davenport, T. & Mittal, N. (2020, March 23). How CEOs can lead a data-driven culture. *Harvard Business Review*.

Demirbag, M., Glaister, K.W., & Tatoglu, E. (2007). Institutional and transaction cost influences on MNEs' ownership strategies of their affiliates: evidence from an emerging market. *Journal of World Business, 42,* 418-434.

Dupey, R., Gunasekaran, A., Childe, S. J., Papadopoulos, T., Hazen, B., Giannakis, M., & Roubaud, D. (2017). Examining the effect of external pressures and organizational culture on shaping performance measurement systems (PMS) for sustainability benchmarking: some empirical findings. *International Journal of Production Economics, 193*, 63-76.

Dupey, R.; Gunasekaran, A.; Childe, S. J.; Blome, C.; & Papadopoulos, T. (2019). Big data and predictive analytics and manufacturing performance: Integrating institutional theory, resource-based view and big data culture. *British Journal of Management, 30*, 341-361.

Fisher, B. (2018, February 7). *In a digital world, do you trust the data?* KPMG. Retrieved from https://home.kpmg/xx/en/home/insights/2018/02/in-digital-world-do-you-trust-the-data.html

Funk, C., Hefferon, M., Kennedy, B. & Johnson, C. (2019, August 2). Trust and mistrust in Americans' view of scientific experts. Pew Research. https://www.pewresearch.org/science/2019/08/02/americans-say-open-access-to-data-and-independent-review-inspire-more-trust-in-research-findings/

Garcia-Perez, A. (2018). Living with data: Scale, time, and space dimensions in a data-driven culture. *Social Business, 8*(1), 87-93

Gitelman, L. & Jackson, V. (2013). Raw data is an oxymoron. *Digital Archiving Resources. h*ttps://dar.cah.ucf.edu/items/show/254

JustinMind (2021, January 27). *All You Need to Know About Progressive Disclosure.* https://www.justinmind.com/blog/progressive-disclosure/

KPMG. (2020). *Guardians of trust: Who is responsible for trusted analytics in the digital age?* KPMG.

Keller, S., Nusser, S., Shipp, S., & Woteki, C. E. (2018, Spring). Helping communities use data to make better decisions. *Issues in Science and Technology, XXXIV*(3).

Khazanchi, S., Lewis, M. W., & Boyer, K. K. (2007). Innovation-supportive culture: the impact of organizational values on process innovation. *Journal of Operations Management, 25*, 871-884.

Kitchin, R. (2014). *The data revolution.* London: Sage.

Kostova, T., Roth, K. & Dacin, M.T. (2008) Institutional theory in the study of multinational corporations: a critique and new directions. *Academy of Management Review, 33*, 994-1006.

Kozuch, B. & Jablonski, A. (2018). *Trust management in public organizations: New concepts and contemporary trends*. Nova.

Lapowsly, I. (2018, March 13). Florida could start a criminal-justice data revolution. Wired. https://www.wired.com/story/florida-criminal-justice-data-sharing/

Liu, H., Ke, W., Wei, K. K., Gu, J., & Chen, H. (2010). The role of institutional pressures and organization culture in the firm's intention to adopt Internet-enabled supply chain management systems. *Journal of Operations Management, 28*, 372-384.

Marcoulides, G. A. & Heck, R. H. (1993). Organizational culture and performance: Proposing and testing a model. *Organization Science, 4,* 209-225.

Maxwell, J. C. (2007). *The 21 irrefutable laws of leadership, 10th anniversary ed.* Nashville: Thomas Nelson, Inc.

Measures for Justice. (2021). "*About us*." Retrieved from https://measuresforjustice.org/about

Measures for Justice (2021). *Measures for Justice Data Portal.* (Data Release: 3.7.8). https://www.measuresforjustice.org

Measures for Justice (2021). "Methodology." https://measuresforjustice.org/portal/methodology

Miami-Dade County. (2021). *Strategic Planning and Resilience 2020*. Miami-Dade County. https://www.miamidade.gov/global/management/strategic-plan/home.page

Miami-Dade County. (n.d.). *Resilient305.* Miami County. https://www.miamidade.gov/global/economy/resilience/resilient305.page

National Information Exchange Model (NIEM). (2020). "About NIEM." https://www.niem.gov/about-niem

Neider, L. L. (2012). *Perspectives on justice and trust in organizations*. Information Age Publishing.

Nielsen, J. (2006, December 4). Training wheels user interface. Nielsen Norman Group. https://www.nngroup.com/articles/training-wheels-user-interface/

Nuts and Bolts. (n.d.). Data visualization: Understanding how to visualize your data. https://nutsandboltsspeedtraining.com/powerpoint-tutorials/data-visualization/.

Obama, B. (2013, May 9). Executive order: Making open and machine readable the new default for government information. The White House. https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-

Pasco County Clerk. (2020). Annual comprehensive annual financial report. https://www.pascoclerk.com/ArchiveCenter/ViewFile/Item/2487

Quaintance, Z. (2018, December 27). What's new in civic tech: Connecticut's first state data plan. *Emergency Management*.

Ramaswamy, P. (2015, June). *How to create a data culture*. Cognizant 20-20 Insights.

Resilient305. (2019). "*Our work."* https://resilient305.com/our-work/

U.S. Government. (2021). Department of Transportation Case Study: Enterprise data inventory. Department of Transportation Case Study: Enterprise Data Inventory | resources.data.gov

Waller, D. (2020, February 6). Ten steps to creating a data-driven culture. *Harvard Business Review.*

Wolf, T. (n.d.). *Governor's goals*. https://data.pa.gov/stories/s/kcju-zzfu

The World Bank. (2019). Open data in 60 seconds. http://opendatatoolkit.worldbank.org/en/open-data-in-60-seconds.html

Yolo District Attorney. (2021). Commons Dashboard. https://yoloda.org/

Yong, K. T. & Pheng, L. S. (2008). Organizational culture and TQM implementation in construction firms in Singapore. Construction Management and Economics, 26, 237-248.

Zu, X., Robbins, T. L., & Fredendall, L. D. (2010). Mapping the critical links between organizational culture and TQM/Six Sigma practices. *International Journal of Production Economics, 123,* 86-106.

End Notes

[1] Measures for Justice, “About us,” 2021
[2] Lapowsly, 2018
[3] Criminal justice data collection, 2018
[4] Criminal justice data transparency, 2018
[5] National Information Exchange Model (NIEM), 2020
[6] Garcia-Perez, 2018, p. 89
[7] Gitelman & Jackson, 2013
[8] Ramaswamy, 2015
[9] Maxwell, 2007
[10] Davenport & Mittla, 2020
[11] Davenport et al., 2019
[12] Ibid
[13] Bowker, 2005; Gitelman & Jackson, 2013
[14] Davenport et al., 2019
[15] Ramaswamy, 2015
[16] Davenport et al., 2019
[17] Waller, 2020
[18] Obama, 2013
[19] U.S. Government, 2021
[20] Ibid
[21] Avila, 2018
[22] Connecticut Public Act No. 19-153, 2019; Quaintance, 2018
[23] Connecticut, 2020
[24] Connecticut, 2020
[25] Wolf, n.d.
[26] Commonwealth of Pennsylvania, 2020
[27] Miami-Dade County, 2021
[28] Miami-Dade County, n.d.
[29] Resilient305, 2019
[30] Ibid
[31] City of Boston, 2017
[32] Bloomberg-Cape Coral, 2020
[33] Ibid
[34] Ibid
[35] City of Detroit, 2019
[36] Data Driven Detroit, 2020
[37] City of Detroit, 2019
[38] City of Fort Lauderdale, 2013
[39] Bloomberg, 2020
[40] City of Fort Lauderdale, n.d.
[41] The World Bank, 2019
[42] Demirbag et al., 2007; Dupey et al., 2017; Khazanchi et al., 2007; Liu et al., 2010
[43] Allaire & Firsirotu, 1984; Kostova et al., 2008; Marcoulides & Heck, 1993; Yong & Pheng, 2008; Zu et al., 2010
[44] Kozuch & Jablonski, 2018
[45] Keller et al., 2018
[46] Funk et al., 2019
[47] Kitchin, 2014
[48] Ibid
[49] Davenport et al., 2019
[50] Kitchin, 2014 (adapted from Adler (1986) & McCandless (2010))

[51] Maxwell, 2007
[52] Waller, 2020
[53] Ibid
[54] KPMG, 2020, p. 125
[55] Bateman et al., 2020
[56] Ibid
[57] Ibid
[58] Davenport et al., 2019
[59] Neider, 2012
[60] Ibid
[61] Ibid
[62] Fisher, 2018
[63] Davenport et al., 2019
[64] Waller, 2020
[65] Kitchin, 2014 (adapted from Adler (1986) & McCandless (2010))
[66] Waller, 2020
[67] Measures for Justice, MFJ Data Portal, 2021
[68] Nuts and Bolts, n.d.
[69] Pasco County Clerk, 2020, p. C-6
[70] Chisholm and DuPont, 2021
[71] JustinMind, 2021
[72] Nielsen, 2006
[73] JustinMind, 2021
[74] Yolo District Attorney, 2021
[75] Measures For Justice, “Methodology,” 2021