Office of the Child Advocate
Juvenile Justice Policy and Data Board - Data Subcommittee Meeting Minutes
July 18th, 2019

Council Members or Designees Present:

- Abigail Averbach (DPH)
- Laura Lempicki (Probation)
- Rachel Wallack (Juvenile Court)
- Matthew Broderick (DMH)
- David Chandler (DYS)
- Lydia Todd (Children's League)
- Leon Smith (CfJJ)
- Barbara Kaban (CPCS)
- Elizabeth Walk (Office of Representative Dykema)
- Kim Occhiuti (DCF)
- Kristina Johnson (EOTTS) Cristina Tedstone (DCF)
- Melissa Threadgill (OCA)

1.

Other Attendees:

- Alexis Yohros (OCA)
- Dan Finkel (EOTTS)
- Jaclyn Cirinna
- Rowan Curran (DYS)

Meeting Commenced: 2:08 pm

Welcome and Introduction from the Child Advocate:

Ms. Threadgill called the meeting to order and welcomed the subcommittee and guests. The group went around the table to introduce themselves.

Ms. Threadgill asked if anyone had any questions or feedback regarding the May and June meeting minutes. The group did not have any additions or corrections. Ms. Threadgill asked for a motion to approve. Ms. Kaban motioned to approve with no objections. The minutes for May and June were approved.

Update re: Fall Implementation Report

Ms. Threadgill reminded the group that at a previous meeting, they discussed specific data metrics that would help meet statutory requirements and measure the post-implementation of the juvenile justice law. These changes include 1) changes in the age of the youth involved with the juvenile justice system, 2) changes in the way youth are being matched to and provided services, 3) youth outcomes as a result of these changes in service delivery, 4) changes in delinquency

charging patterns, and 5) any remaining or emerging gaps in educational supports and other needed services. Ms. Threadgill then reminded the group that the JJPAD had decided to report data from July 1, 2017 through June 30, 2019 (one-year pre/post implementation) in the report they were developing for this fall. She said that she would follow up with agencies regarding the specific data requests next week.

Juvenile Justice Data Site Presentation

Dan Finkel from the Executive Office of Technology Services and Security introduced himself and explained that the presentation would cover the recent work that EOTSS and OCA had done on the juvenile justice data website. He also said that the plan is to present the big picture in charts and pull up dashboards to discuss feedback and ways to improve the work thus far.

Mr. Finkel described the objectives of the juvenile justice data website. These goals are to publish a public-facing, interactive data site on the state of juvenile justice in Massachusetts, incorporate data from many different MA State Agencies, and when possible join data from different agencies to build cross-agency insights. Mr. Finkel and Ms. Threadgill emphasized that joining data from different juvenile justice agencies is a long-term goal. The goal is to also utilize tools and processes that can be transitioned to Office of Child Advocate (OCA) for site maintenance and sustainability. The hope is that OCA can maintain the juvenile justice data website using these tools.

Mr. Finkel explained the process of creating the juvenile justice data website. This first involves ingesting the data or receiving it from agencies such as EOPPS, DYS and the Trial Courts/Probation in the form of an excel sheet or CSV file. The next step is to stage the data, for example, aggregate the data, join it to the census, join it to GIS, and device calculated fields. This is accomplished in Python. The third step is to visualize the data by integrating charts, graphs, and maps, as well as developing user filters. This is done in Tableau. The final step is to report the data by integrating copy and context, layer in UX, and develop it to Mass.gov standards.

Ms. Averbach asked whether embedding Tableau into Mass.gov is a new feature. Mr. Finkel replied that is not new. Ms. Johnson said that one has always been able to imbed visuals on Mass.Gov, however, they must be stored somewhere. She also said that they are storing the data on a Tableau server, a public server that then feeds into Mass.gov. Mr. Finkel said that there is a very specific way to tell Mass.gov to make this happen.

Ms. Threadgill asked Mr. Finkel to provide more detail on the census and GIS data. Mr. Finkel explained that he pulled in census data to link to the arrest data from the community survey over the past ten years. He also said that more recently, he began to pull median income data to start thinking about its impact. He noted that Mass GIS published shapefiles which can be downloaded and joined to the data in order to power the maps.

Ms. Todd asked what census data Mr. Finkel is considering. She also asked whether age is being used as well. Mr. Finkel replied that he is incorporating age and that he bucketed the age categories to 12-14 and 15-17 in order to be consistent with the census data.

Mr. Finkel provided screen shots of the data overview visualizations. He explained that the first set of visualizations provide a high-level overview of the data. This includes a map-based overview, a demographic breakdown, and arrests over time. The second set of visualizations break down the impact of race. It examines arrest counts normalized by the census as well as type of arrest breakdowns.

Mr. Finkel presented on the first interactive dashboard, examining arrest counts by county. He explained that these are MA juvenile arrest counts using 9 years of EOPPS data. He also said that presented are the raw numbers of counts from 2009-2017. Mr. Finkel highlighted that, for example, Natick has more arrests than Framingham while Springfield has the highest number of arrests. He also noted that EOPPS shared NIBRS format data, meaning there are a lot of towns without data, including Boston. He also explained that the data can be broken down by year, gender, age, and race/ethnicity. The data can also be broken down by arrest type and offense types using the NIBRS definitions.

Mr. Smith asked whether there is a way to differentiate between where an arrest takes place and where a youth is from. He noted that mall towns are always skewed because there are youth from multiple towns that come in. Mr. Finkel replied that this data was not provided.

Ms. Todd asked whether the rate used by the census can be used instead of raw numbers. Mr. Finkel replied that it would be simple to divide by the population to get the rate.

Ms. Kaban asked whether the data includes seventeen-year-old youth. Mr. Finkel replied that it does include seventeen-year-old youth.

Ms. Threadgill noted that she plans to add a layer on to define different terms and have pop-ups that provide further definitions for each term.

Mr. Finkel presented the interactive dashboard showing the demographic breakdown of arrests. He noted that the breakdown can be seen as gender by age as well as race by age. He also created tree maps to highlight race and ethnicity. He noted that one could click on individual groups and further interact with the site.

Mr. Finkel then presented the time-based interactive dashboard showing arrests over time. He first presented counts by year of arrest and showed the filters in the right-hand corner to further breakdown the data by the different categories. He noted the downward trend and the seasonality component.

Ms. Walk asked how arrests are defined and whether it is defined as an instance or number of people arrested. Mr. Chandler also asked if it is person-based. Mr. Finkel replied that it is event-based. Ms. Walk also asked how it is recorded if a person is arrested for five different charges. Mr. Chandler replied that an event is considered one instance, but it can have multiple offenses.

Ms. Johnson noted that the leading charge is the one recorded. Ms. Threadgill stated that this would need to be clarified on the website.

Mr. Finkel presented on the interactive dashboards highlighting the racial components on arrest data. Mr. Finkel showed the number of arrests per population over time. He also incorporated ethnicity as white-Hispanic and white non-Hispanic. He noted that the white-Hispanic and Black arrests look like they are falling flat but that it might instead be a visual trick because the decay rate is actually constant across the races.

Ms. Threadgill asked to clarify: did Mr. Finkel mean that the percent change is similar, but it looks more dramatic because the base rate for some groups is higher? Mr. Finkel said yes, that is correct. Ms. Threadgill said that it might be helpful to think of ways to properly present that information given that it could be confusing for the public. Mr. Finkel noted that it is a difficult topic to tackle.

Ms. Averbach said that it seemed strange to categorize race and ethnicity the way it is currently on the dashboards. She also said that often times Hispanic youth won't identify themselves as a specific race (white or black), so normally there is a category for Hispanic (under ethnicity) and then categories for racial groups (white, black, Asian, etc.). Mr. Finkel replied that he was questioning the integrity of the data because the white-Hispanic line is so high. He noted that it might have something to do with the reporting.

The group discussed the different ways that agencies report race and ethnicity data. Ms. Todd said that at Department of Youth Services, questions are asked in the following order: non-Hispanic or Hispanic, then race, then country of origin. She asked whether police have the same standard. She noted that the trial court might also be converting to the same standard, as well as other agencies such as the Department of Education. Ms. Occhitui noted that the reporting for DCF is Hispanic or non-Hispanic, followed by race. The group continued this discussion on race/ethnicity reporting. Ms. Threadgill noted that the dashboards would need to be restructured and categorized in the same way that the majority of agencies are reporting it.

Ms. Cirinna expressed that the various race and ethnicity categories look confusing. Mr. Finkel replied that the white-Hispanic category was kept because the trend stood out. The group engaged in a discussion on how race and ethnicity is reported via checking the different column boxes. Mr. Smith asked what happens when they don't check both columns. Mr. Chandler replied that at DYS, 98% of those who answer yes for Hispanic chose not to self-identify for race and skip the question. He noted that this might be different for police because they are instead making an observation. Ms. Averbach said that it might be interesting to look at the differences between those who identify as white-Hispanic and those who do not. Mr. Finkel said that looking at consistency in race and ethnicity reporting across different data is good feedback that will be incorporated into the next steps.

Mr. Finkel presented on the "type of arrest" dashboard and reiterated the different definitions of arrest: on-view, summoned/cited, and taken into custody. He noted that without any filters, the white non-Hispanic group has the highest ratio of summoned/cited versus on-view arrest. He also

said that the discrepancies stand out even more when you use different filters, such as age. He noted that the data can also be broken down by Group A and Group B offenses.

Ms. Threadgill said that it is important to mention before presenting that there are concerns about data quality for some of these categories. Mr. Finkel explained that there is an odd discrepancy in the "all other offenses" category. Ms. Johnson said that when an arrest is made, there are a list of offenses and that very often the arrest is listed under "all other offenses" instead of having a specific offense. She said that typically it is used for city ordinances but that she believes it is used as a catch-all or utilized when police/or departments are busy. She noted that in some municipalities, this category can make up as high as 70% of all arrests, leading to questions about data quality. She also said that arrests by warrants in the "all other offenses" are high, making her believe that they probably are not city ordinances.

Mr. Chandler asked whether this could be a default warrant or probation violation and that police don't know the offense because they are picking people up on warrants. Ms. Threadgill asked whether it's possible that 70% of arrests would be for warrants. Ms. Todd replied that in some places, detention centers can be filled mostly with youth picked up on warrants. She asked the rough percentage of arrests that fall into the category of "all other offenses." Mr. Finkel replied that he would estimate this to be approximately 10-20% of the data but that the variation across place is incredibly high. Mr. Chandler explained that gateway cities do a lot of warrant clean ups and that for places like New Bedford these offenses can be coming in as warrant sweeps.

Ms. Kaban asked if there is a way to distinguish between misdemeanor and felony offenses. Ms. Johnson replied that she thought this can be done. Ms. Threadgill said that this is something that would be helpful to do if possible.

Ms. Averbach noted that she does not believe the colors on the dashboard are particularly helpful. Mr. Finkel replied that they are working on figuring out the best way to present the data. Ms. Todd noted that the best way to present it would be in a way that the public will understand. Mr. Finkel also noted that Ms. Yohros mentioned that the arrest categories can be collapsed. Ms. Yohros replied that some researchers dichotomize the category into fewer options, and that it might help visualize the patterns more clearly.

Ms. Kaban said it might be helpful to compare what the juvenile population looks like to begin with versus percentage of youth who are arrested. She said that this can make the data jump out in a different way. For example, if a particular racial category is 1% of the juvenile population, but 15% of youth committed to DYS. Ms. Threadgill noted that while they can't join different datasets, they can begin to do some comparisons like that.

Ms. Todd asked whether the dashboard allows for county breakdowns. Mr. Finkel replied that yes, there is the ability to filter by county.

Ms. Todd asked whether the data shows city breakdowns. Mr. Finkel replied that city can be added in and asked for JDAI purposes whether the city would be important.

Ms. Lempicki said that one of the issues is that there is no Boston arrest data. Ms. Threadgill replied that they are working on getting that data.

Ms. Kaban suggested that they can also plot the top ten cities of juvenile crime.

Mr. Smith asked where the Cape Verdean population fits in, as they tend not to identify as Black. He emphasized that this is important in areas such as Brockton. Ms. Todd noted that this has been a discussion in the trial court. Mr. Chandler said that in DYS, it is up to how the youth identify themselves. He also said that they can refuse to self-identify and then pick their specific country of origin. Mr. Leon noted that this is an interesting nuance to consider. Ms. Threadgill said that the dashboards are bound by the categories that are in the data. NIBRS data does not have this level of nuance. She also said this can be put in the footnotes for the arrest data and then they can figure out how to best present it in other datasets such as DYS.

Mr. Smith noted that offense type is very important to consider. Mr. Chandler asked Mr. Finkel where the definitions come from for what constitutes offense type. Mr. Finkel responded that they are rolled-up NIBRS categories. The group discussed the importance of the different categories. Mr. Smith noted that it is an important layer in viewing disparities.

Ms. Todd said that the probation violation question is important to consider. She asked whether it is possible to tease out whether the offense is an original crime or a violation. Ms. Threadgill said that it is important to think through these questions with the probation and detention data. She also said that these nuances are not available in the arrest dataset however.

Ms. Threadgill asked whether it would be helpful to have the specific offenses. Ms. Lempicki replied that it would be helpful to reduce the categories into the common charges. She also said that it would be interesting to look at the nuances between offenses such as drug offenses that have class A, B, C, D, and E. Ms. Johnson replied that as long as it is in the offense category of the data then it can be done. She also said that the information on what type of substance was found is not in the data.

The group spent time looking through the original dataset on excel. Some members of the group agreed that the categories are different because they are federal categories, and that it may be easier to show the top ten offenses.

Ms. Threadgill asked the members of the group if there were any other ideas, questions, or things they'd like to see done with the data.

Mr. Curran asked if there is information on the arresting party. Ms. Johnson replied that there is an associated agency name, but that state police is not included in the data.

Ms. Threadgill noted that they are working on getting the UCR data as well as the NIBRS data, but that this data is much less detailed.

Ms. Lempicki noted that it is easier to look at the data by county. Ms. Johnson replied that it can already be broken down by county but asked whether it is also worth breaking it down by town. She said that the numbers begin to get small when broken down by town.

Ms. Occhitui asked whether it is possible to calculate the relative rate index, which may jump out a bit more, particularly when it comes to the topic of racial disproportionality. Mr. Finkel asked for clarification on the relative rate index. Ms. Occhitui stated that this would mean normalizing the counts. Mr. Finkel said that this could be done. Ms. Threadgill noted that this number may not be as accessible to those who don't spend much time with data and that, therefore, it might be helpful to display both numbers.

Mr. Finkel showed the group a visualization that they have been working on in python. He explained that similar to what was mentioned with the relative rate index, the plot looks at population rate and examines three different race/ethnicity categories. He explained that the contours represent those racial and ethnic categories that are overrepresented in the towns.

Ms. Kaban asked what towns are being represented. Ms. Johnson replied that all towns are represented in the data. Ms. Walk asked for clarification on the towns in the data. Ms. Johnson replied that all towns are represented, and she said the plot is just showing some of what is being done when playing around with the trends, patterns, and visualizations in the data.

Ms. Threadgill said that this is an ongoing project for the summer and welcomed any other feedback and ideas.

EOTTS Data Sharing and Services Presentation

Ms. Johnson introduced herself and said she would be going over the legal framework for data sharing that went into effect last October. This also includes EOTSS's role in the data sharing space and the other services they offer. Ms. Johnson laid out the agenda for the presentation, which included an overview of the data sharing framework, EOTSS's data services, and their key products such as data prep/secure storage, data science, and data visualization.

Ms. Johnson went over the data sharing problem statement. The first challenge is that data is not shared across state agencies in a cost-effective, replicable manner. For example, there have been over 287 unique sharing agreements and it takes an average of 133 days to create a data-sharing agreement. The second challenge is that there is confusion over rules and regulation limits for data sharing; therefore, *not* sharing data becomes the default. There is lack of clarity over what can be shared and with whom and there is no common process or support system for data sharing. Ms. Johnson explained that the Data Steward Counsel created a new framework for data sharing to be used across agencies.

Ms. Johnson explained the new legal framework put forth by the counsel. She defined an MOU as a statewide agreement broadly governing the sharing of protected data between Secretariats. She defined a Data Use Licensing Agreement (DULA) as an agreement between a data owner and a data recipient(s) specifying the details of how data will be shared for a specified purpose/project. The MOU covers the following areas: justification for data sharing, data

access/confidentiality, data transfer/storage, security requirements/breaches, requirements under a DULA. The parts of a DULA include value extraction and learning, which includes a purpose/question and plan for dissemination. It goes over technology and security, including the infrastructure to combine and protect data. Governance is also part of the DULA and involves the framework for active management of sharing data, including legal compliance and term/termination.

Ms. Johnson explained the role of the Data Steward Council, a peer forum to support data-sharing. It 1) assists with the timely execution of the DULA process, before and after signatures, 2) manages the MOU, including the addition of new signatories, 3) mediates disagreements around data sharing, and 4) provides general support for data sharing projects.

Ms. Johnson also explained the DULA process. It begins by identifying the question/problem to be addressed, then identifying the data and owning agency, and reach out to the owning agency's contact. If the data can be shared, the owner or recipient initiates DULA online. Both parties complete/ review DULA online and then both parties sign DULA online. If the data cannot be shared, the recipient reworks the question/problem or the project ends.

Ms. Todd asked who the DULA applies to and whether it is just for state agencies. Ms. Johnson replied that it applies to all state agencies and others such as the Children's Trust are in the process of signing on. Ms. Kaban asked if this would apply in the case of a private party wanting access to data. Ms. Johnson said that at the moment this would not apply.

Ms. Johnson went over the data sharing resource site. She said that the site will provide resources for Data Sharing Coordinators, such as DocuSign resources for initiating/using DULAs, quick access to the Data Sharing MOU, and listing of Data Sharing Coordinators. The resources for other data users will include the introduction to the Data Steward Council, an application to join the MOU, instructional resources for signing DULAs, and data sharing FAQs.

Ms. Johnson explained EOTSS's data services. This includes data sharing, which supports the work of the Data Steward Council, manages the statewide MOU, facilitates the electronic DULA system, and develops resources for Data Sharing Coordinators. In terms of data analytics, services include data matching, integrated data systems, data science/analytics, and machine learning. The services also include open data such as data sites and imbedding into Mass.gov.

Ms. Johnson presented on the data prep and storage of an integrated data systems (IDS). The agencies ingest the data or pull data from agencies into a secure environment. Then the data is processed, merged, and identifiers are suppressed. The data is then transferred, and anonymity is verified. In the stage process, the anonymized data is stored and prepped for reporting. Last, in the report phase, the results of the data are delivered to stakeholders. Ms. Johnson used the example of the Learn to Earn project that integrates data across transitional assistance programs with the goal of getting the data out of silos and integrating it. This will also include data such as wage, income, and educational attainment. The process involved ingesting datasets from the nine agencies, securing it,, and matching and anonymizing it. Ms. Johnson noted that the projects are flexible and can be spun up or spun down accordingly.

Ms. Johnson went into detail about the project, explaining that analytic process of examining the SNAP "Churn." She said that the agencies wanted to solve the problem of benefits lapsing and then individuals having to reapply. She said that over 10,000 people a month churn in this manner. The aim was to describe how much unintended churn takes place in SNAP. The goal was also to predict which individuals are high risk for unintended churn. Finally, it was important to prescribe and define what approaches are effective at reducing unintended churn.

Ms. Johnson pointed out interesting findings regarding SNAP Churn. She said that in particular there are a variety of different types of behaviors within that space. She identified three types of people: clients who knew to renew but engaged with DTA too close to the application deadline; clients who did not know they expired until they prompted by no access to benefit; and clients who let their benefit expire but returned due to life changes. She also noted that young people are much more likely to churn and suggested that this may be an avenue for digital intervention. In addition, there is also an isolated group who receives no benefits but qualify for SNAP for other reasons.

Ms. Johnson shared an example of a project from the TNC rideshare on the Mass.gov website. EOTSS partnered with Department of Public Utilities (DPU) to develop a data site for TNC rideshare (Lyft/Uber). The site helps the state and the public better understand ride flows between municipalities and over time. Ms. Johnson went on the website to show the interactive nature of the report.

Mr. Broderick asked how requests are prioritized and what happens when there are more requests that can be met. Ms. Johnson answered that they currently do not have more requests that they can meet and that they are currently involved in three projects: Learn to Earn, the juvenile justice data website, and the adult side of the criminal justice system. Mr. Broderick noted that as it becomes more well known, there may be a lot of competition for getting services. Ms. Johnson replied that it would typically prioritize the governor's office or where they have a particular advantage, such as data sharing/storage and higher level data science pieces. She said that typically engagements are shorter because they are doing a small piece of the project and handing it back. She also said that it helps to hone in on a part of the project where they could be most helpful, which also helps so that they don't become inundated with projects.

Ms. Threadgill said she is impressed with level of sophistication EOTSS brings to this work, and that it is helpful for state agencies to know that these particular services exist.

Closing comments:

Ms. Threadgill announced that there is no August meeting but that she will keep the group informed as the juvenile justice data site progresses, and send out a Doodle to schedule the fall meetings

Adjournment: 3:45 pm