

Federal Justice Statistics Program Webinar

June 25, 2026

Ekokoi Gouahinga: Good afternoon, everyone, and welcome to today's webinar on the Federal Justice Statistics Program. My name is Koi Gouahinga. I'm a public affairs specialist here at the Bureau of Justice Statistics, and I'll be your host today. Please be advised that this event is being recorded.

First, a primer on BJS: the Bureau of Justice Statistics is the primary statistical agency of the U.S. Department of Justice. Our mission is to collect, analyze, publish, and disseminate information on crime, criminal offenders, victims of crime, and the operation of justice systems at all levels of government.

BJS also supports state, local, and tribal governments to improve both their statistical capabilities and the quality and utility of their criminal history records. I'd like now to take note of a few housekeeping items. First, for an optimal interactive experience, we recommend that you join from a laptop or desktop computer. If you do experience technical difficulties, please flag them by using the Q&A function on your screen. We have a member of our staff who is able to assist with that.

To minimize disruptions, all attendees are muted. And, again, you may use the Q&A feature at any time to submit your questions throughout the presentation. Our presenters will address your questions at the end of our session.

There's no need for you to record this webinar because we will make a copy of the recording available to everyone, along with a transcript. Should you need live captions, you can activate them by clicking or tapping the CC button or the three dots you see on your screen, depending on your device.

It is now my pleasure to hand it over to Kris Tennyson, our Deputy Director of Statistical Collections here at BJS, for her opening remarks. Kris?

Kristin Tennyson: Thank you so much, Koi. Welcome to the Bureau of Justice Statistics' second webinar for 2026 on our Federal Justice Statistics Program. We are delighted that you're able to join us today.

I'm Kristin Tennyson, a deputy director at the Bureau of Justice Statistics, and I'm really excited to introduce you to one of our flagship administrative data collections. The Federal Justice Statistics Program comprises data on the full federal criminal justice process, from U.S. Marshals' arrest data to prosecution, pretrial release, and adjudication data from the Administrative Office of U.S. Courts, sentencing data from the U.S. Sentencing Commission, and Bureau of Prisons for incarceration data. As you'll hear, federal offenses are distinct from what you might see at the state or local level. And BJS is the only entity that standardizes and links data across all six federal agencies that handle criminal offenses in this jurisdiction. We hope that you're as excited as I am for this presentation and that you'll be ready to engage in a lively discussion afterward.

This is just our second in a series of webinars that we'll be announcing in the near future. And if you want to stay up to date on what we're doing, I encourage you to go to bjs.ojp.gov/subscribe and we will post that link at the end of the presentation and there you can join our email list.

That, and if you want to check back regularly on our web page, those are the best ways to keep up with what we're doing. So, it's my pleasure now to introduce you to BJS's Chief of Justice Statistics, Suzanne Strong. She's going to talk to us a little bit more about the work of BJS and her unit in particular. She'll also introduce the other presenters for today's webinar. So, the floor is yours, Suzanne.

Suzanne Strong: Thank you, Kris. I'd also like to welcome you to our webinar on the Federal Justice Statistics Program, or the FJSP, but before we begin, I wanted to give a brief overview of the Bureau of Justice Statistics. As Koi mentioned, BJS is the primary statistical agency of the U.S. Department of Justice and is one of 13 independent federal statistical agencies.

Its primary mission is to collect, analyze, publish, and disseminate information on crime, criminal offenders, victims of crime, and the operation of justice systems at the federal, state, local, and tribal levels. BJS is organized by topic area and includes statistical collections focused on corrections, judicial statistics, law enforcement, reentry, recidivism, and special projects, and victimization. We also have units dedicated to statistical methods, criminal justice data improvement programs, statistical policy and standards, communications, and technology and data management. I am fortunate to be the chief of the Judicial Statistics Unit.

We collect data from both the state and federal level. At the state level, JSU focuses on criminal cases from receipt by prosecutors through case filing in court through the outcome of the case, whether that is a disposition or sentencing.

We collect administrative data from courts and survey data from prosecutor and public defender offices. Those surveys include staffing, caseloads, budget, responsibilities at the office, and special topics. For our federal data collection, we also have the FJSP, which is why you're here today. And I would now like to introduce you to three very talented JSU statisticians: Mark Motivans, Ryan Kling, and George Ebo Browne.

Mark will celebrate 26 years working for DOJ in July. He has been with BJS for 24 of the 26 years and has been working with the FJSP since his very first day. Under his direction, the FJSP has expanded to include more data and to focus on improving data quality and standardization across the agencies that contribute data.

Ryan has been with BJS for 2 years, but has worked with BJS data for the majority of his career. He is also focused on improving data quality, but especially focused on improving the quality of the data linkages between agencies that contribute data to the FJSP.

Ebo has been with BJS for 6 years and is focused on improving our data analysis tool, the Federal Criminal Case Processing Statistics tool, or the FCCPS. I'm excited for them to share the advances that we've been making with the FJSP, and I will now hand the webinar over to Mark.

Mark Motivans: Thank you, Suzanne. Today, I'll walk through the Federal Justice Statistics Program, or FJSP. We'll discuss BJS's approach to federal justice data collection, standardization and record linkage techniques. This will be followed by a demonstration of the Federal Criminal Case Processing Statistics web query tool. Our session will conclude with a Q&A, where we're happy to address any questions. Let me start by giving you some background on the FJSP itself and why it exists.

The FJSP was created by BJS over 40 years ago, in 1982. The program was created to serve as a national clearinghouse, essentially a central repository where all federal criminal case processing data comes together in one place. Before this, there really wasn't a systematic way to track what was happening across the entire federal criminal justice system. Congress passed the National Justice Improvement Act in 1979, which gave BJS the official mandate and legal authority to do this work.

The slide shows the key language from the statute because it really captures what we're trying to accomplish. The act authorized BJS to collect, analyze, and disseminate comprehensive federal justice transaction statistics and to provide technical assistance to, and work jointly with, other federal agencies to improve the availability and quality of federal justice data.

So, BJS has a dual mission. First, we're collecting and publishing statistics. That's the public-facing part you might see in our annual reports and data tools. But second, we're also working behind the scenes with our federal partners to help improve their data and ensure the quality of the information they're providing.

What makes the FJSP unique is that it represents 30 years of standardized administrative data and that word standardized is critical. We're not just gathering whatever data agencies may happen to have. We're working to make sure that data is comparable across agencies and over time.

So let's talk about what actually makes a crime federal versus a state or local crime. This is really important to understand because it determines which cases end up in our federal justice statistics. Federal crimes generally fall into a few main categories. First, there's the "where" factor. That is, crimes that happen on federal property. So, if someone commits an assault in Yellowstone National Park, that's federal jurisdiction because it's a national park. Same thing applies to military bases, federal courthouses, federal prisons, in Indian country, and even on airplanes during flight. The location itself makes it federal.

Second, there's the "who" factor. This includes crimes involving federal officials or committed by federal personnel. For example, if someone threatens a member of Congress or assaults a federal judge, that's a federal crime because of who the victim is.

Third, and this is a big category, there are certain areas where the federal government has exclusive power to regulate. The slide highlights four key examples: immigration violations, tax crimes, counterfeiting, and treason. States simply can't prosecute these. Only the federal government can. So, if someone is in the country illegally or doesn't pay their federal taxes, those are inherently federal matters.

Finally, there's interstate commerce, probably the broadest federal jurisdiction category. The U.S. Constitution gives the federal government authority over commerce

between states, which is why drug trafficking and weapons offenses frequently become federal cases, especially when they cross state lines. Now, here's where it gets interesting. States can also prosecute many of these same crimes, so you might have both federal and state prosecutors who could handle a drug case.

The decision about who prosecutes often depends on factors like the severity of the offense, available resources, and whether it's part of a larger federal investigation. Understanding these jurisdictional categories helps explain why our federal justice statistics capture the specific types of cases they do, and why federal caseloads look different from what you'd see at the state level.

Now, I want to show you the network diagram because it really illustrates the complexity of what we're tracking in the federal criminal justice system. The diagram is organized into seven stages, shown across the top: arrest and prosecution, pretrial release, adjudication, sentencing, appeals, and corrections. These represent the major decision points and processes that a case can move through in the federal system.

Let's walk through this, from left to right, starting with how cases enter the system.

A case can begin in several ways through a probable-cause arrest, a fugitive warrant, or summons. Then there's a criminal investigation. And after arrest, the case goes to the U.S. Attorney's Office, where they decide whether to prosecute. Now, here's where you start to see the complexity.

Not every case that enters the system makes it all the way through. Cases can be transferred to state or local jurisdiction. Maybe it's more appropriate for state prosecution. Cases can be resolved through other prosecution resolution, perhaps through a diversion program. Or prosecution can be declined altogether. The prosecutor may decide there isn't sufficient evidence or it's not in the interest of justice to move forward.

For cases that do move forward to prosecution, charges are filed in District Court and we get to the pretrial release stage. Here, defendants either get released, maybe on their own recognizance, or with conditions, or they're detained. If detained, hearings may be held to revisit whether the defendant remains detained while awaiting trial. This recognizes that factors justifying detention may change over time, particularly in cases with extended pretrial periods. If they are released and violate the conditions of release, there may be revocation hearings.

Then we move to adjudication. This is where the cases go to trial or result in guilty pleas. Notice there are two tracks here. Cases can go through the Federal District Court trials or they can be handled by magistrate judges for less-serious offenses.

After conviction, we get to sentencing. The court imposes sentences which could be federal prison time, probation, supervised release, fines and costs, or some combination. Cases can be appealed, and those appeals can be affirmed, dismissed, or reversed, sending the case back for further proceedings.

Finally, is the corrections stage. This includes time in federal prisons, followed potentially by supervised release in the community. And even here, there are decision points: violation hearings, potential reincarceration, continuation of supervision, and eventually discharge from the system.

What's important to understand is that cases can exit the system at almost any point along this pathway. You can see all those arrows leading out. Cases leaving federal jurisdiction, charges being dropped, acquittals, sentence completion. This is why we need data from all six agencies. Because to understand where cases are going, we need to track cases across each of these stages.

This diagram also shows why our standardization work is so critical. We're linking records across multiple stages, connecting an arrest record to a prosecution record to an adjudication and sentencing record and to a corrections record. That linkage allows us to follow individual cases and helps us understand the complete picture of processing.

The FJSP relies on six different federal agencies to provide data. The key data sources include the U.S. Marshals Service and DEA for arrests, the Executive Office for U.S. Attorneys for prosecutions, Administrative Office of the U.S. Courts for pretrial and post-conviction supervision, adjudication, and appeals, the U.S. Sentencing Commission for sentencing, and the Federal Bureau of Prisons for corrections data. Each agency contributes data from its area of responsibility. This multi-agency approach ensures that every stage of case processing is documented and can be tracked longitudinally.

So, now let's talk about the core challenge we face and why standardization is so critical to what we do. Here's the fundamental problem: we're collecting data from six federal agencies, and each one has its own way of organizing and managing

information. The agency data are, in effect, siloed. The U.S. Marshals Service tracks arrests one way, the U.S. Attorney's Offices track prosecutions differently, the courts have their own system, and the Bureau of Prisons has yet another approach. They're all doing their jobs effectively, but they're not necessarily thinking about how their data can best be connected with everyone else's. Another complicating factor is that these systems are constantly evolving. Agencies upgrade their technology, change their business processes and modify how they code certain information.

BJA has developed specific methods to standardize these annual data files, so we can actually compare information across agencies and track trends over time. So, what does standardization actually mean in practice? We've established four key elements.

First, we use a standard unit account, the person-case. This might sound simple, but it's crucial. We're counting each defendant in each case. So, if one person is charged in three separate cases, that's three person-cases. This gives us a consistent way to count across all agencies.

Second, we have a common reporting period: October through September. This aligns with the federal fiscal year, and it means that when the Marshals Service reports arrests, and the courts report adjudications, they're using the same time frame.

Third, we've developed a common offense classification. Different agencies might describe the same crime differently in their systems. We translate each of the agency codes into one standardized classification. So, we can accurately report on things like drug offenses, violent crimes, or fraud across the entire system.

And fourth, we use standard disposition outcomes. Whether a case is dismissed, results in a guilty plea, goes to trial, or ends some other way, we make sure these outcomes are coded consistently regardless of which agency is reporting them.

The result of all this work is comparability both across agencies at a single point in time and over time. Because of our standardization, it's what allows us to say things like, "In the last fiscal year, drug prosecutions increased by X percent. And the median sentence for drug cases in the last fiscal year was Y months." The data originate from two different agencies; however, we have added confidence that we're making a reasoned comparison.

Now, I want to talk briefly about how we ensure the data you see in our publications is accurate and reliable. Well, the first thing we do is compare the current year's data to

the previous year. We're looking for unexpected changes or anomalies. For example, if we suddenly see a 50% jump in drug trafficking cases from one year to the next, that raises a flag. It might be legitimate. Maybe there was a major policy change or enforcement initiative, but we need to investigate and understand what's driving that change, before we report it.

Second, we evaluate missingness. Basically, what data's missing and does it matter? If 2% of records are missing a defendant's age, that might not significantly impact our statistics. But if 30% are missing offense information, that's a serious problem that could skew our results. We assess whether missing data will affect the validity of what we're reporting.

Third, we compute summary statistics to identify trends in the distributions. We're looking at things like are the age ranges reasonable? Do the sentencing patterns make sense? Are there unusual spikes or gaps in the data? This helps us spot data quality issues that might not be obvious when looking at individual records.

Fourth, we validate identifiers, things like case numbers and defendant IDs, to ensure accuracy. This is especially important for when we link records across different stages. So, we can be confident we're connecting the right records together.

And, finally, we compare our generated statistics back to the agency's reports. The agencies themselves produce their own reports and have their own expectations about what the numbers should look like. So, before we publish, we're checking: do our statistics align with what the agencies are reporting? If there are discrepancies, we work with them to understand why and resolve any issues. This entire process is extensive and systematic. It takes time, but it's essential for maintaining the reliability of federal justice statistics.

I want to introduce you to what we call Standard Analysis Files, or SAFs. These are really the foundation of everything we do in the FJSP. We call them the building blocks because they're what make all of our data products and analyses possible. So what exactly are SAFs? They are annual cross-sectional data files that we create for each cohort, meaning each year's group of cases.

Each SAF is standardized by applying the four key elements discussed earlier. But here's another reason for what makes them powerful: SAFs include an identifier designed to link person-case records across all the different stages of the justice

system. Remember that network diagram we looked at? SAFs, together with a link identifier, allow us to follow a defendant from arrest through corrections, connecting data across the different agencies. My colleague Ryan will speak in more detail about how this link identifier is created and used by BJS, shortly. For now, let me walk you through how we create these files using this workflow diagram.

On the left, you see we start with the six federal agencies providing source data. We go through several steps in data processing before we end up with our final products, on the right.

First, we obtain the agency's raw data and documentation, and this is important. It helps us correctly interpret the data that each agency is sending us. Then we do two types of checking. Value checking, making sure the data values make sense, and data checking, verifying the data quality more broadly.

Next, we reconcile the data to FJSP standards. We're translating each agency's format into our common framework to create the actual SAFs. And, critically, we link records across SAFs, and this linkage is what allows us to conduct longitudinal analysis and track cases through the system.

All of this feeds into the FJSP database, which contains the source data files, the cross-sectional SAFs themselves, both incoming, outgoing and stock files, our linking system that connects records, and comprehensive documentation about everything. From that database, we produce a range of data products you can see on the right side of the diagram. We publish our annual Federal Justice Statistics Bulletin. That's our annual publication with key statistics and trends.

We also produce reports on specific topics. This includes recent reports titled *Methamphetamine, Cocaine and Other Psychostimulant Offenses in Federal Courts*, and *American Indian and Alaska Native Persons in the Federal Criminal Justice System*.

BJS maintains and develops the FCCPS data analysis tool. That's an online platform where users can create custom tables and do their own analysis of FJSP data files. My colleague Ebo will speak about this later.

Finally, sanitized data files are archived and made available for research and statistical purposes. This ensures that the data are preserved and accessible to the research community while protecting confidentiality.

I will close my segment by saying that BJS is continuously looking for ways to improve federal justice statistics. I would like to acknowledge the technical support we receive from staff at Abt Global LLC, the project's contractor. We hope you're excited about how this standardized data can inform policy, improve transparency and deepen our understanding of federal criminal justice. I would like now to pass it over to Ryan Kling, who will talk about the FJSP's Dyad Link System.

Ryan Kling: Thank you, Mark. So, Mark described a linking system that goes between the individual SAFs that, that standardized SAFs that get created, published every year. We generate reports, statistics, make it available to the public. One of the, one of the activities, though, that BJS has to take on wholly is linking across these, and I want to give you an overview of why we link, and then a taste of how we link.

I'm not going to give you all of the details. If you're interested, there'll be some contact information at the end of this presentation. Feel free to ask for more information, but I do want to give you a sense of how this works within the Federal Justice Statistics Program.

So why do we link? One big reason is to augment unique information available in one agency with data collected by other agencies. Each agency is responsible for its own work. So, for example, the Sentencing Commission will report a large number, a large amount of data about some of the evidence that's collected along, during a case.

Those data really aren't reported anywhere else, but one of the upshots, one example would be they'll report the amount and type of drugs that are involved with the drug trafficking case. That's really helpful if you want to look at other agencies who aren't necessarily reporting that. So, we want to, if we want to characterize drug traffickers in prison, we may want to sort them out by the types of drugs that they were trafficking. We can fill in missing data across the agencies.

Agencies do record some of the same data, particularly offense charges, conviction charges, personal demographics. Certain facts about the case are necessary for different agencies, but sometimes things don't always get filled in.

These are administrative data. There, very few things are actually required in administrative data, so anytime we can reinforce data from one section to, from one agency to another, we'd like to do so. Another reason is we can use substantive data from multiple agencies to improve reporting.

This is similar to the first bullet point, but it's not exactly the same. What we're trying to think of are things that only apply to a certain stage. We may, we may want to know time served for any particular conviction. The courts aren't really going to record that, we expect that from the prisons. We can use data from the courts and the prisons to reinforce each other and look at, look at the relationship between sentencing, conviction status, sentencing, and time served. So, that's just one example, but taking advantage of each agency's workload and substantive aim, we can improve reporting across the system.

And kind of most importantly is, and observing an individual's path from arrest to imprisonment. Mark showed you the complex diagram of an individual, a person-case's evolution through the system. We can describe that through using the data. We need to link, though, because no two agencies' data really talk to each other.

There are some challenges though to linking. No one identifier uniquely identifies individuals. People will have a Social Security number. If you're familiar with the criminal justice system, the FBI will introduce a number called the FBI number that is tied to an individual's fingerprints. This doesn't always get written down, and even in the places where it might get written down, it's not always written down correctly, or people don't always enter it.

It is the challenge of administrative data, but if we find enough identifiers we can fill in, we can reinforce links, we can fill in where we have missing individual identifiers. And that really goes to the second point, too. There's a variation in quality of data entered into administrative data systems.

And the reason, there's a practical reason for that. The administrative data systems are their case management systems. They're there to facilitate the workload of that agency. So, if you're in prison, it's to facilitate intake, prisoner movement, and then release from prison. That doesn't really need a lot, a lot of information about the prosecution charges or things that happened at the arrest stage. Every stage has that, and even if you put those types of measures into your case management system, not everyone's going to fill out measures that don't matter to them, and that makes sense. It's, it's hard work being a prison guard. It's hard work being a prosecutor. It's hard work being a court clerk or a judge. You really want to pay attention to the data that matters most to the job at hand. So there's going to be variation. We understand that and we can reinforce that. Even though we have challenges, we can overcome them and it is

the best way to follow individual experiences. You could imagine a survey where you're talking to people who have gone through the system. They're not going to know everything that happened at each stage, sometimes because a lawyer will handle it. Sometimes a prosecutor or, will know their business certainly better than the individual who's gone through the system.

So, it's, to really describe what's happening in the system, the case management, the case, case management software is the best one. Kind of the easiest way and the best record for that.

Next slide.

So, in service of this, BJS has developed methods to link records across stages and the first generation of this was a list-based system that grouped records from all of the agencies based upon their case numbers. And this makes sense. You have somebody who is arrested, they get a case number assigned by the U.S. Courts. You can follow them through to their disposition and what happens with corrections. There's an issue with it, though. Not everyone wants to write down the court numbers with fidelity. And, there's a reason for that. When you're in corrections, either prison or probation, you don't necessarily need to know the court case number. You just need to know this person's here, they're going to be here for this amount of time, and they, they need this sort, these sort of services.

The same thing at the arrest stage. You may not even know what the case number is at arrest. You may not, you won't know until a case is actually filed. You certainly won't have a case number during an investigation. So, the list-based system worked well for certain parts of the system, but it wasn't a holistic solution.

The second generation uses paired agency link files and kind of simply it is taking two stages of the system and linking them together. And we'll talk, we'll talk about the consequences of that in a second. It uses a variety of standardized identifiers, so unlike the first-based, the list-based system, it uses all of the personal identifiers: the FBI number we talked about, the name and date of birth, Social Security number, a lot of different identifiers to try and put, put things together, not just case number.

We report links through adjacent stages of the criminal justice system. We can still follow experiences if we chain those adjacent stages together. This lets us look at guided experiences of individuals, either within an agency — for example, from case

filing to case disposition — or across agencies, from case disposition to a sentencing recommendation. So, it, it provides a pretty flexible way to observe an individual, a person-case's experience over time.

Next slide.

This diagram is a simplification of the full diagram that Mark showed, and it's really here to motivate the links that we make in the FJSP. There are 10 links, paired agency links that we make available to the public that we routinely create. We create other ones, and we can talk about that if people have questions, but we can focus on these 10.

In each box you'll see a particular agency, the US Marshals, the EOUSA or the attorneys, AOUSC or the U.S. courts, the Sentencing Commission, the Bureau of Prisons, and each phrase under that are the SAFs that we create, are some of the SAFs we create from that. But I wanted to focus on the SAFs that actually get linked.

Within any one box, so focus on the EOUSA box, you'll see arrows between any two SAFs, so suspects and matters opened, and then matters concluded. We'll make links between those and we call those intra-agency links, and you can see there's five of those within the attorneys.

There are also interagency links between any two agencies, between the Marshals and investigative matters or investigations concluded for the EOUSA, we'll make a link. And you, you can see how those, how those work. And this generally follows the timeline that Mark showed — the U.S. Marshals are associated with arrest and bookings, the EOUSA is associated with the case process, but also the investigate, the investigation and the case process. But so is the, are the U.S. courts.

The Sentencing Commission is in the case process, it is toward the end when, after conviction, when somebody is sentenced, and then once, once things go through that, if somebody's sentenced to prison, we'll observe them in the Bureau of Prisons. One of the nice things about the Dyad Link file is that once we make the links, we remove all the personal identifiers. Mark had mentioned, there are pointers from one SAF to another from the links. That's really all that exists. There are some indications of link quality, but you can really think of it as this SAF, this SAF's record matches that SAF's record. And that's it. And that's really all you need to know. You don't real, if you're,

even if you're following an individual, you can follow them across the system. You don't need to know who, any of their personal identifiers, so we retain some confidentiality.

Next slide.

I want to talk a little bit about the identifiers we use and give you a sense of what we use and how they're used. So, the first thing we'll sort through are the person level identifiers, something that identifies the individual.

So we'll look at the name and date of birth. There are lots of people with the same name. You really need the date of birth to, to help distinguish an individual. So we'll put those together. The Social Security number, if it's available. Not everyone has an SSN, but most people do. The FBI number we mentioned before. The U.S. Marshals will assign their own unique identifier that sticks with an individual over time. It's called a register number. It's really only in use in the federal system. The FBI number gets used across federal and state, state and local justice agencies. But the register number is really for the federal system. But it's a convenient identifier for us because we're looking at federal data.

We'll also get the alien number. Frequently, we'll get the alien number assigned by the U.S. Customs and Immigration Service. Alien numbers get assigned to anybody who enters the country legally, and anybody who enters illegally that becomes known to USCIS. So if you're arrested, you'll get an alien, and you were not previously known by USCIS, you'll get an alien number. So, that, that also persists with an individual. So, it's, it's another way we can follow at least a certain group of people and reinforces individual identifiers when there's missing data.

And we'll also use the docket number, the case number assigned by the U.S. Court. That isn't exactly an individual identifier, because many cases have more than one defendant. You can think of a drug trafficking conspiracy that may have 20 or 30 defendants, all the way from, all the way from street dealing to international trafficking. But that gets you close. You, if you, they're in the same case, you know, it's probably one of those 30 people and then we can look at some of the other identifiers to figure out who that person is.

Next slide.

We also use case-level information for linking. We use this to sort out people's cases. There are people in the federal system that do recidivate. You may see something from

the 2000s, early 2000s, somebody from the early 2000s, and you might see them again, say, in the mid-2010s. This happens pretty frequently. So ,we want to be able to sort out their case level. We can't just rely on the individuals.

The case-level, level linking variables we use, we use the case numbers. The federal judicial district and filing year are part of that case number. Sometimes we will not really, I won't get into the details of why we may distinguish between different parts of that case number. Just think about, oh, there's a case number that's assigned when somebody goes into the courts, and we're using that.

We'll also use important dates. So, over that timeline that Mark showed, you can imagine there are dates associated with that. Their arrest and booking date, their investigation start and end date, their case filing and termination dates, the date they're convicted, sentenced, their prison admission and release dates. All of those follow a timeline, and they should be roughly, a logical timeline.

There may be gaps and we can, that are explain, well explainable, but these help us sort through both getting the case right when we're looking for person-case records, and if we have ambiguous matches between people. Say, they have the same name and we don't have a date of birth. If we can start looking at the case information trying to, the case date to try, to sort out, which, which one of these links is the correct link.

We'll use other shared variables that get reported across the system. The offenses get reported in most places, particularly the conviction offenses. All of the individual demographics, particularly sex. Sex is the best one. Race, race and ethnicity when it's available, but we will try and use other shared variables to particularly sort out ambiguous matches when we're not sure from some of the harder individual identifiers.

Next slide.

This is just a depiction of how any one program works. So, I showed you 10 Dyad links. Each one of those is associated with a separate linking exercise, because each agency reports only certain identifiers. The Sentencing Commission will report a large amount, a large number of identifiers but the U.S. Courts only report the name and the case number, and some, and some of the case facts.

So we need, we need, while we have 10 different programs, they all follow the same logic. We try and put together people, and once we put the people together, then we

sort through their cases. And some of the detail you see here is the tiering system that we use, and I'll get into that in a second.

But if you, the thing to remember is we're trying to look for people, trying to put them together across two places, whether within an agency or across agency, sort through their cases. Once we're satisfied with that, we call that a validated link and that'll appear in the Dyad Link file.

Next slide.

The easiest way to kind of explain this is through an example, and this is an example of linking records across two agencies. When, we've selected here, are the U.S. courts, their cases terminated and the U.S. Sentencing Commission, their guidelines recommendation for a sentence. This is a pretty good link. I'll just preview it now and despite, though, the AOUSC not really reporting a ton of personal identifiers, but we follow linking hierarchy. What we're trying to do is match as many identifiers between the two datasets, exactly, and call that a link.

So, to start with, we'll look at the case number, we'll look at the sentencing date, whatever we can find and say if we get all of the, the name, if we get all of these things to link exactly, we're going to call that a match. That's great. We can take that and set it aside and take the remainder and start looking for links where we don't have all the information. Maybe only 90% of the identifiers match, and that might be because it's missing in one dataset or the other, or it was mistyped in one dataset or the other. But if we find 90% that match, exactly, from the remainders, we're pretty sure, we're really sure that matches, we're going to play the same game. We're going to set those aside and then do another round of matching. And we're going to keep doing that, taking away identifiers that match because they either don't match because they aren't exact or they're missing.

We employ some fuzzy matching logic and I'm happy to get into that later if, if anyone's interested. But if you have two things where one character is off, two strings where one character is off, but everything else is linking pretty well, you're probably going to say that's the right person, if you were doing this by hand. Well, we're mimicking that in code.

We also have some logic when people write nicknames down. So, for example William, maybe somebody named William, maybe William in one dataset, but Will or Bill in

another dataset. We try and take that into account. So, there are some fuzzy linking matches that we do further down the hierarchy.

We get, we'll get to a point, though, where we start making matches and we don't believe them. There, and, there's a quantitative way of expressing this, but if you were to look at it, you'd say those two records don't match. Once we see that, we stop and we go back and say, well, what is the minimum amount of matches? Fuzzy and exact matches that we think are links, and we stop it there, we call that the Dyad Link file. So, there are some details here, but you're better off to think through the overview and then the details are customized to, to the link.

Next slide.

And we can characterize how well we're doing. So, this graph is showing link rates between the two, these two agencies, the U.S. Courts terminations and the guidelines calculations.

Take the top line, the green line. We're looking at it from the guidelines calculation side and we're saying did we find a termination in the U.S. Courts? And if you're looking at it, you're, we're, we're showing the link rates from 2000 to 2023 here, we're finding between 95 and 97% of guidelines calculations we can find their U.S. Courts termination.

That's pretty good. We're really happy with that. There's some technical reasons why they may not link. We know why they don't link. We can explain that. We do have extensive checking of the ones that don't link to make sure, of the ones that do and don't, to make sure there's nothing systematic, that we're not systematically excluding certain cases that we're making good links. This is a really good representation, and if we were going to do any statistical reporting from this, we can, the missingness doesn't really worry us.

The bottom line is playing the game from the other side, looking at U.S. Courts terminations and looking for their Sentencing Commission record. This improves over time from roughly 82 to 85% to 93, 94%. There's some reasons for this. In most datasets, the, the U.S. Sentencing Commission has been very good for a very long time, but in most federal data, in most federal datasets, but most case management systems, data fidelity has increased over time both through UI/UX improvements, but through, but also through the recognition that better identifiers yield better data.

So, we expect that to increase over time. Not everything links, and we know why there are certain, there are certain cases that just don't get a guidelines calculation. Either the severity is too low, the case is adjudicated so quickly that there's no opportunity for the Sentencing Commission to issue a guideline. That happens rarely, but those are the explanations we have.

And, again, we do a lot of checking for, looking for bias in these links. We don't see any. We're really happy, we're really comfortable with saying this represents true links and we can represent what's going on in the federal justice system between here.

Next slide.

There are also within-agency links. I'm going to only touch upon this because this is a much more straightforward link than the tiering system and the fuzzy matching that we described before. And the reason why this is, this is much easier is because we retain all the administrative identifiers within any one agency. Since we do that, everything should link perfectly. And spoiler alert: they do. But we want to, we want to use the, the administrative data identifiers where we can because that's the best record of how this case works.

And it makes sense if you, and I like to use hospital admission as kind of a something similar. If you go to a hospital, you'll go in, they'll write your admission date, they'll write down what, what you're in for, they'll write your exit date. It'll be stored in one record in a database. It may be a relational database, but it will be all tied to one record. You won't have an end file, an out file like we have in FJSP. Well, we can re-create that using the administrative identifiers, which is what, what we do.

Next slide.

This is a representation of how well we do linking, linking the cases at start and cases at the end. The top line is looking from the case termination, looking back to their filing. There are some reasons why we won't have 100%. It's 98, 99%, 99.9% toward the end of series, there are some cases that just last decades. One case you might think about is Whitey Bulger. Since I live in Boston, Whitey Bulger, when he was found in the Caribbean, his case was resurrected from the 80s.

Well, that case lasted decades. It stays on the AO's caseload, so not everything, you won't, since our data don't go before 1994, there's some very old cases that just never

get, we never, that are terminated, we never see their, their beginning. From the other side, from case filing, you'll see a significant drop-off, starting around 2017, 2018.

Normally, cases in the federal system will be disposed of within 180 days or 210 days, but not all of them. Some of them will last years, particularly if they're, if they're particularly complex cases or defendants can't be found. In those cases, they'll be pending for years, and that's the representation here. You're not find, for cases that are filed, filed, you're not finding their termination.

And that's why the link rate decreases significantly in 2023. Cases filed in 2023, if we're looking for cases terminated in 2023, many of them will still be pending, by the end of 2023.

Next slide.

I'm going to wrap up with just a handful of uses from linking files.

This is scratching the surface and it's really here to help you think about research questions you might want to ask. We, they, there are a vast number of questions that link files can answer that no one agency can. But it is helpful to see some examples that either BJS has done or has facilitated with other researchers.

So, for the interagency file, what we're looking for are system-level type of research questions. Think about something that's court terminations linked with the Sentencing Commission. Maybe we want to look at case outcomes in the Sentencing Commission, their ultimate sentence, their enhancements that they get from sentencing guidelines. But we want to see if their filing charges affected that.

Well, the Sentencing Commission doesn't report the filing charge, but the U.S. Courts do. So, by linking these together, we can, we can drag in the filing charges into our sentencing database. That link also lets us study case progression. So, if we look at filing, then we can look at the progression to conviction charges because they're not always the same.

Somebody may get charged with an assault and some evidence is further found, and it all of a sudden turns into a drug trafficking crime because that is a much, that was a much more salient crime for the prosecution. But we can also go into the Sentencing Commission and look at some of the evidence that was found for that particularly for, for enhancements or for discounts in a sentence and that, that's an important thing.

And the other thing we can do is drag in the individual correlates of those outcomes. The Sentencing Commission does a really good job reporting, kind of ascriptive individual identifiers like race and sex, and even marital status, where the U.S. Courts really doesn't bother themselves with that. So, we can put these together and start sorting through case progression and with respect to individual characteristics or case facts.

Within agency files, really think about those agencies' workload. So, think about court filings linked with terminations. We can start looking at things like court back, backlog. When you get to the end of the series like I, like I showed you, there's, there is a backlog. We can start seeing, look, observing the backlog, seeing what's coming up for the courts. We can also look at historic adjudication time. This is important when we're thinking about, there is a constitutional guarantee for, for a speedy trial. If we're looking at adjudication time, though, if we're looking at backlog, terminations can help address some of that historical data.

But if we're thinking about new cases and we're making projections, we really need to know what's coming onto the docket, what's being filed. We may not observe their termination, but if we can look at data from historic data, historical data from terminations, we can make projections about adjudication time, but also case dockets at the end of the year. And that's, doing projections is one of the, one of the really nice things about the intra-agency linked files.

It also lets us pick through certain types of cases. So, say we want to just examine a subset of cases, ones that are adjudicated quickly. We don't know why. Why, is it a fact about the district that they, the location there, that it's adjudicated in? Is it a fact about the case type itself? Is it a fact about something about the...? Really, those two, two things that we're looking inside, say the courts, we can, we can start examining those cases and picking them out for further, further review. That kind of ends how we're thinking about linking. What I want to do, do now is pass this to my colleague Ebo, who's going to talk about the online data tool.

George Brown: Thanks, Ryan. Very, very interesting stuff. So, now we'll be moving on to talk about the Federal Criminal Case Processing Statistics tool. I'll refer to it as FCCPS for short. FCCPS is an interactive public-facing data analyst tool that allows interested parties to examine various aspects of the federal criminal justice statistics system or federal justice statistics system.

I'll probably make the mistake of referring to this as a new program or new data analysis query, but in all reality it's actually been around for a while. I think we've had a query tool at BJS since the 1990s, but in 2022 one of the first projects I did was kind of revamp it. So, the new tool is powered by Tableau. It allows a lot more data visualization, different ways for the user, approved-user experience. So, users are able to create different dashboards, quick queries, slice and dice, and look at the data a little bit different than it was before. The underlying data is coming from the FJSP.

Next slide, please.

So, the FCCPS tool enables querying of data on persons processed in the federal criminal justice statistics system in the three major programs, or case processing stages of the criminal justice system.

So, if you think about it, the three stages are law enforcement, prosecution or courts, and incarceration. So, each one of these are individual dashboards. Within these dashboards are different cohorts, different ways to examine the data. I'll talk about it in a little bit. And we also have a dashboard that looks at the federal statutes. So, it's statistics on specific criminal statutes or the United States Criminal Code can be queried. So, again, I'll go into that a little bit detail here in a second and it'll make a lot more sense.

Next slide, please.

OK. So this is a landing page. Again, this is a public facing page. You go to this by going to our website, hitting the FCCPS tool. This will be the landing page.

The first thing you see here we're on the Home tab. You'll see at the top: Law Enforcement, Prosecution, Incarceration. Each one of those are individual dashboards. If you were to select one of those, that would take you to the dashboard and you'd be able to examine the data that's particular to that cohort that you're interested in examining.

Next, the United States Criminal Code. So, that's a dashboard that allows you to query the federal statutes.

And next to that is Supporting Materials. Supporting Materials is one of those things that, one of the main visions with updating this tool was to improve the user experience. So, we wanted to make sure if you're not a criminal justice expert or you're

not familiar with the federal system, we're going to give you all the tools to make use of this and actually understand what's going on. And that Supporting Materials is kind of your best friend understanding how to use this.

So, first is a quick, or we call it, I think we call it a Quick User Guide, and that is just like a high-level overview of you know the data variables that are being analyzed, some of the data sources, and, again, just a high overview of FCCPS and the data that's linked in there or available in there. And then we also have a more detailed query or more detailed list of what to examine. It's, we have it, I think it's called the User Guide, the Detailed User Guide, and that gives you more of the meats and potatoes on what's going on. So the first is, you know, a really high overview and the other one is more of, you know, everything you kind of need.

There's also a supplemental Terms and Definitions dictionary, and that, and that will allow, again, I know I'll probably try to do my best and not use all the acronyms, with a lot of acronyms we end up using in the federal system, and those will give you definitions of those acronyms, different variables being analyzed, and some of the data sources. So, again, that's probably your best friend if you're not too familiar with the federal system.

And I'll talk about this in a little bit later, but there's also a download results. So, if you're querying something, you see something really cool and you want to pull the underlying data that's protected and aggregated or you want to do a screenshot, that Download Results will be how you do that.

And then, also, we see By Stage there. So, By Stage on this right here, page, we have that Law Enforcement, Prosecution and Incarceration. And with, within each dashboard there are three cohorts. So, within the Law Enforcement dashboard, we can examine the dashboard on a cohort of persons in investigations initiated, persons arrested and booked, persons in investigations concluded. For Prosecution and Courts, you have persons charged, persons in cases closed, and persons sentenced.

And last but certainly not least, incarceration. You have persons entering federal prison, persons in year end, and persons exiting federal prison. Then again, we try to make this as much as user-friendly as possible. So, again, this is, this is the landing page that should be going to. If you were to hit plus on that, that would give you the, that would expand the query that's being, that would expand the, the note and give you kind of the underlying information.

Next slide please.

So, if you hit plus like on Law Enforcement, that would expand and you see the underlying data is coming from the Executive Office of U.S. Courts and U.S. Marshals. So, give you some background on that. And if you were looking at, you know, specifically, persons arrested and booked, that shows you there by expanding that tab, that shows you that the underlying data is from the U.S. Marshals. And that, that availability, that functionality is for each of the cohorts in each of the stages.

So, again, hit plus, expand, it'll give you all that information. If you hit minus, it'll reduce it and you won't be able to see it, and you can do that for each one of the cohorts in each one of the stages.

Next slide, please.

So, let's say now that you've you're in the dashboard, you want to, so you've gone to our landing page, you've hit home. Now you go into law enforcement. You want to examine the law enforcement tab. So, you just again toggle that by selecting law enforcement at the top. And then when you do that, this is a landing page for law enforcement for persons in investigations initiated.

So, here you see you have an option to select a variable. The select variable will give you a dropdown menu of all the variables that are applicable to that dashboard that's being analyzed. And then under that you're able to filter by individual variables. OK, so you're able to select a variable, then also condense the variable or slice and dice by putting on a filter for what exactly you're interested in examining.

I'll also mention here: next to that is that top, is that first dashboard. It was basically just a table and this table is just analyzing what you have selected for being analyzed. So, for this one I simply have selected, I think I just have the year selected on this one.

Next slide, please.

OK. So, here is a dropdown of the variable.

So, again, this is for persons, this is for Law Enforcement persons in investigations initiated. The options that we have is year, for variables, it's year, age, sex, race, citizen status, arresting agency as a summarized list, arresting agency as a detailed list, offense category, and offense category as detailed list. So, each one of those are

individual variables that you can select that will allow you to be used to examine the data.

Each variable is used as a filter, so you can select a variable as a main filter as, as a main variable you're examining. Then you can select individual variables to serve as filters.

The year range for this dashboard is from 1998 to 2023. So, let's say you were wanting to do a subset of that data. So, let's, let's say you want, you want to look at data from 2015 to 2023. You would select that and then the data will be filtered in that.

So, all filters are multiselect and can be applied at the same time, and data updates as the filters are applied. So, as you put on a filter, all the dashboards will be updated with the filter that you have on there.

Next slide please.

So then I'll try to give you kind of a word picture of how this looks, how the dashboard looks. So first we had a landing page. We selected a variable. The first would be our options of what we selected to filter by and then next to that would be the, the table. And when you're selecting table, above that you have an option to sort.

And you can sort by variable, and sorting by variable I should note that it's not by alphabetical name of the variable, instead by BJS priority list. So, this is the most serious offense to least serious offense. So, if you hit select variable for persons initiated – excuse me, for persons arrested and booked – you see that violent offenses are the first because they're the most serious, followed by property offenses, drug offenses, and, and thereon. And then, probably what's probably most useful would be sorting by highest to lowest count. So, if you sorted by highest count, that would make the highest count at the top. So, you see there that immigration offenses are the highest, followed by drug offenses and so on. You can also do that highest, highest to lowest or lowest to highest.

Next slide, please.

So, then that is the first dashboard. At the top, we have that table, then under that you just simply scroll down on your screen. The first data visualization that you would see, the first kind of the first dashboard is just a table, the next under that would be more of

a data visualization. This is how you can quickly see trends. The first data visualization would be this line graph.

So, the line graph is, I like it because it helps you see trend over time, right? So, quickly see right here: persons arrested and booked from 1998 to 2023. We see that there's a steady increase from, from 2000s, from 1998 until, looks like about 2013, 2014. Then the drop. It looks like from 2016 to about 2020 it increased. And then from 2020 and thereon a decrease with a slight uptick in the last year or two. So, again, helps you see, quickly, trends over time.

The next is a bar graph, so this is looking at — oh, sorry. One slide back. Sorry. So, so next is the bar graph. The bar graph is, looks at persons arrested and booked, and this very similar to the, the table before by sorting from least to, to greatest. This just shows you quickly which variable has a highest count. So, on here you see that immigration offenses has the highest count, followed by drug and supervision. So, again, super useful way to quickly see trends first over time, then also the highest count.

Next slide, please.

This is one of my favorite graphs. This is a multi-line graph, so scrolling under that, first we have the table, then we have the line and bar graph, if you were to keep scrolling down on that dashboard this would be the next thing you would see. On this, it's the multi-line graph and this basically slices up one variable with, you know, multiple, multiple entities over time. So, for example, looking at the offense category, we have offenses of violent property, drug, public order, weapons, immigration offense, supervision, material witness, and unknown.

You see that, for a majority of offenses, they kind of do the same thing. They're all consistent, for all of them, other than immigration offense. You see immigration offense has a much higher count and ebbs and flows much more than the others with, it looks like, you know, pretty steady increase from the year 2003, 2004, all the way to about 2014, 2015 a decrease, and in 2017 increases, and then decreases again. So, again, allows you to see those trends by variable and looks at any, you can see any discrepancies within the variable with this dashboard right here.

Next slide, please.

And this would probably be — I'm a little bit biased — this is probably my favorite dashboard. This is a heat map. So, the heat map allows you to examine intersectionality, so it helps you to see how two variables are related to one another. So, here I have age and race selected as my variables that I'm querying and the legend is your friend here. That, that legend shows you the heat map and kind of, kind of the color trend.

So, the, the range is going to be from 70 to, what is that, 553,000? And with the darkest color being the one with the highest count. So, if I was interested in understanding, for persons arrested and booked, what was the age and race with the highest count? Looking at my legend, I see it's going to be 553, 553,251. I'm looking over on the heat map. I see that for defendants between the ages of 25 to 29 who are white were the highest persons arrested and booked from the year 1998 to 2023.

If I wanted to know the lowest count, I could see there, that's going to be Asian and Pacific Islander who were defendants who were less than 18 years of age. So, again, intersectionality lets you see the relationship between more than one variable, and that legend will help you kind of understand how to drive the table and understand what's going on there.

Next slide, please.

So, one of the goals of BJS is always to protect the data, by not only protect the data as in making sure it's, you know, used properly, but also protect the data of the individuals with the information where, that with the information that it, that it's being collected on. So, one of the ways of doing that is by suppression. So, when you're querying the data, you'll notice a, the symbol will arrive, and this, this is suppression symbol for if the count is less than 10. So, if you're querying something and you know you put some, in, in, in the field that you've selected that you're querying, the filter that you have on, get a less than 10 count, this symbol will be there.

And this symbol's essentially saying that it's a low count, so you won't be able to get that exact number. And this suppression rule applies for the next-lowest count in that row, as well as the number that's being examined, and also there's secondary suppression for the row or column also suppresses the next-lowest count. So, again, this, this symbol will arrive if you're querying something, put some filters on. It's a low count. It's not saying you've done anything, you're wrong. It's instead just showing you

the symbol to show you that we're protecting the data for anything of a count of less, 10 or fewer.

Next slide, please.

And then next is the United States Criminal Code. So, the United States Criminal Code allows you to query the U.S. federal statutes to examine the data, so, data from persons charged and persons in cases closed using title and section for U.S. Criminal Code. So, this can, U.S. Criminal Code can be examined by a, a multi-year or single year.

So, using multi-year you're examining the number of persons in cases filed, the number of persons in cases closed, and for single year you're able to examine the outcome for persons in cases closed and the type of sentence for persons convicted. And also for mean persons, mean, mean prison sentence or probation sentence or fine amounts for persons and convicted for, for persons being convicted.

Next slide, please.

So, here, this is how this looks. So, you would select that United States Criminal Code and then you have the dropdown menu option so you can select the statistics that you're interested in examining. I have here, looks like number of persons in cases filed and I have my option to select that Criminal Code. Again, the goal with this was to make it as user-friendly as possible. So, rather than, so, I know the U.S. statutes for many offenses, many offenses I do not.

We have a link here, if you hear, if you see that "click here" feature, that "click here" feature will take you to an external website that allows you to examine, to look up a statute, an offense, and figure out what the statute is. So, for example, if you were interested in specific white-collar offense or a specific offense related to treason or human trafficking or what not, you'd be able to examine that, you'd be able to look up what the Criminal Code is by using that feature, then coming back to this and then specifically selecting that Criminal Code you're interested in, and the statute would be populated.

The year range for this is from 1994 to 2023, but you're also able to, just like all the other ones, you're able to apply a filter. So, if you wanted to select the years, you know, 2005 to 2015, you could then select that by that filter tab, right there on that right-hand side.

Next slide, please.

And then last but not least is you're able to download the data. So, you're able to download the data in the aggregate form. So, if you're querying something, you got some really cool trends, you wanted to pull that data, you can download that as a CSV file and then you'd be, again, the same suppression rules and all the things that apply would also be applied to that data. But, also, if you were — I've done this a few times where I've been querying something I really liked, really liked the image — rather than being forced to like screenshot it or, you know, take a picture of it, we actually have functionality through the dashboard that allow you to do that. So, you're able to grab a screenshot of it as a full-screen image as a PDF or PNG.

So, this is the website that you'd go to access the dashboard. Definitely pay attention to it because we're having some updates. Right now, the most recent year of data we have is 2023. We should be having 2024 data coming available pretty soon, and also we'll be having new data agencies as well. So, we'll be hopefully having new cohorts coming in the near future. That's all I have. I'll pass it over to Ryan to talk about data archiving.

Ryan Kling: Great. Thanks, Ebo.

Yeah, the online tool is an impressive tool and if you haven't used it, just check it out. It will answer a ton of your research questions. And if your research question can be answered by the tool, definitely use it. The data are public. They are available now. They are updated relatively quickly. And as Ebo said, we're we're making, some, really exciting expansions to it.

If you need the individual data, there is a way to do it. They are in restricted access that the National Archive of Criminal Justice Data at the University of Michigan. They're restricted from general dissemination, but you can download the codebooks freely available and there's a resource guide that will describe the FJSP for you. The process is fairly involved for restricted access, so if you can use the FCCPS, you should. But you'll need a restricted data use agreement. You'll need to basically upload your proposal and then obtain IRB approval or an exemption from an IRB to, for your research.

Next slide.

We just want to give you an overview of the steps to request FJSP data. Some of the steps may be a little difficult to parse. NACJD is there to help if you have questions. And if you have questions, you can even go to BJS, at the contact information at the end of this presentation.

So, the steps you're request FJSP data. Prepare your research questions. You probably already have this through a proposal that you've either done if you're a graduate student with your advisor, or if you're at a research institution at, for a grant funding cycle, you want to review the FJSP Resource Guide at NACJD to identify the data you need to request. Since you can get all of the codebooks, you'll know what variables are available, what years are available, what stages of the FJSP are available and you want to go through that and pick out the things that you think are going to be related to your research questions.

Next slide.

Then you'll have to go through the standard application process. What you're seeing here is a depiction of the, the standard application process for one of the many files from FJSP. This is the linked file that we talked about before. What you'll do to do this, you'll search for something like Federal Justice Statistics Program link, which you see here, and what will pop up is the latest paired agency link files.

If that's what you want. If you want to look across agencies or within agencies, you'll want to select that and add it to your basket. Then you'll want to go back through for anything else you're adding. Say you want something from incarceration, or you want something from arrest. You'll add those, you'll have to search for those. Search for the years that you want and add those to your basket.

Once you've done that, you can create an account and follow the prompts to complete your application. This is just the data request because the next slide will tell you the other things you'll need to know.

Next slide.

You'll need an affiliation with an accredited academic or research institution. This isn't as big a hurdle as you might think. If you were getting funding, say, from the Arnold Foundation or something, but you are working very small, kind of research shop. That satisfies as a research institution, but something where you're doing research and you can show that to NACJD.

You'll need your research proposal. You'll already have that idea, so you should have that. You'll need to sign a data use agreement. The biggest hurdle is really IRB approval and that is certainly depending on, dependent on your agency and your situation, how you obtain that and you'll need to give NACJD a data security plan. Once you submit all that, they, it goes under review from NACJD, from BJS. Once it's approved, then you'll need to follow your data security plan, and the data will be provided.

Next slide.

That's the end of our presentation. I'm going to pass it back to Suzanne to address questions and complete the webinar.

Suzanne Strong: Thank you, Ryan, Ebo and Mark for a fantastic overview of the FJSP. And I want to apologize to all of our participants about the various technological difficulties that we encountered. We do have an abbreviated question and answer session available, so I wanted to start with the first question in our Q&A. Is there an FJSP to NIBRS offense crosswalk? And I will pass that over to Mark.

Mark Motivans: Thanks, Suzanne.

At present we're, we're, we're working on this topic of translating NCIC codes to title and section of the U.S. Criminal Code. It's not a particularly straightforward task, as you have codes that are recorded by the U.S. Marshals Service for arrest type activity, and then attempting to, to, to bridge those two titles in section of the U.S. Criminal Code.

There are discrepancies in how those two classifications were designed that make it very challenging to crosswalk. But we are interested, very interested in attempting to do so. So, work's underway, I guess. Stay tuned. Thanks.

Suzanne Strong: Thank you. And our second question. Thank you all at BJS for your hard work processing these data.

Does BJS have plans or methods for reporting specific incarceration or pretrial detention statistics on people held by the U.S. Marshals? It appears that the current FCCPS data do not include all people convicted of crimes and sentenced to incarceration in federal courts; only those that are sentenced to a BOP-operated facility. Is that a correct interpretation? Ryan, I'll turn that question to you.

Ryan Kling: So, I think that's a couple of different questions. The, because you want to sort out pretrial detention from post-conviction incarceration. So, if we start from post-conviction incarceration, anyone convicted and sentenced to prison is going to be incarcerated by the Bureau of Prisons. So, we do, the FCCPS will report on all of those.

That's not everyone who goes through the justice system. On the front end, when you, when you're arrested, awaiting trial, you, you may or may not be detained. Currently, we don't report anything about to, pretrial detention. That's going to change. We're, currently, we're currently working on, adding statistics about pretrial detention, and release, and kind of some of the experiences in, under pretrial services. I can't promise when it's going to happen, but, we're actively, working on it right now. So, we do, but if you're convicted, you're, sentenced to prison. you will see their prison record through the FCCPS, and eventually we'll see pretrial detention. Just not yet.

Suzanne Strong: Thank you. And I will allow one last question before I end the presentation. For the entire panel. For someone not familiar with the federal justice system, where should I start learning about the data? Ebo, I'll start with you and then Mark and then Ryan.

George Brown: Sure. That's a really good question. I know for me, again, I try not to be biased, but I definitely am. FCCPS is, we made it for users. So, it's supposed to be user-friendly.

I remember when we were updating it, I reached out to my sister, who does background on public health. I was like, go on there, play with it. Let me know what you think. I was like, go on there, play with it. Let me know what you think. And she was able to pick up what I was putting down. And so that's one of the biggest things. And, like I said earlier, there's a bunch of user-friendly stuff, so we have some supporting materials that should help you make sense of it. But really, if you just want to see what data is there, what data is available, the tool allows you to do that.

If you really wanted to see what options do I have for law enforcement? What agencies are there? What things can I query for the, the court system? For persons incarcerated, what prison trends have I seen in the federal system? That would probably be the best way of doing it. Also utilizing our, you know we produce a bulletin.

There's a lot of BJS products that we've, that we print out that that are out there for, for the public to disseminate. If you really just want to get your hands on the data, again

showing my bias, I would say the tool would be your one shop, your first stop doing that.

Mark Motivans: OK. I'll take a stab at this and you know I would say replicate, replicate, replicate. So, you know, obtain the data and some of the data that's in the FJSP are also publicly available from the agencies themselves. So, the Sentencing Commission posts similar data. Not exactly the same data that BJS receives. We receive identifiers and the IDB, the integrated database with the Federal Judicial Center.

Go ahead and try to replicate the numbers from the courts, from the Sentencing Commission, from what you see in BJS reports and use the tool or if you can get your hands on some of the individual-level data from these public sites, it can help you craft a proposal to submit to the University of Michigan for receipt of the restricted data from the FJSP.

But look, look, look for how others have done it, including in academic articles and so forth where they've used federal justice data. And, yeah, follow your curiosity. That's what I would say. Thank you.

Ryan Kling: Yeah, it's really hard to recommend too much else. I, I would start with the FCCPS and some of the, some of the data that the agencies make available. I would look at NACJD, at the FJSP Resource Guide because that, that does lay out the stages for you, and it does in a pretty concise way.

And you can start to see, start seeing how, kind of crosswalks work and how offenses are charged. I think, kind of, all those little details that, that you won't know until you start looking at the data. So, yeah, the NACJD, the NACJD kind of, page is kind of the first place.

The other place are the old, are the, the recent, but also the older FJS compendium reports that, Mark has written quite a few of them, mainly because it lays out the system for you and it lays out some of the, kind of important milestones. Not everything, but it will help you think about, well, what does the system actually do?

And then, definitely follow Mark's advice. Go, as you get interested in one stage or the other, follow the rabbit hole down as far as it goes with those, because they go pretty far.

Suzanne Strong: Thank you, Ryan, Mark, and Ebo. I really appreciate it. And then I want to remind everybody that all of the links in our last slide are clickable. So, you are welcome to go ahead and look at some of our recent reports and some of the other resources that we have available. And now I'd like to turn it back to Koi for some closing remarks.

Ekokoi Gouahinga: Thank you to our moderator, Suzanne Strong. I also want to thank our Deputy Director of Statistical Collections, Kris, and I want to thank our great presenters, Mark, Ryan, and Ebo. On behalf of all of us here at BJS, I want to thank you, all of our attendees, for being with us today and for your thoughtful questions.

Before we close, I wanted to mention a couple of useful notes. The recording of today's webinar, along with the transcript, will be made available soon and it will be posted on the BJS website. We will also post PDF copies of our presentation slides. If you have not done so, please sign up for our JUSTSTATS e-mail notifications that you are informed of the next event.

And, finally, I want to mention that you will soon receive an email invitation to take a short, voluntary and anonymous survey. The point of that survey is for you to tell us how we did on this webinar and to help us make future webinars better. That being said, this concludes our webinar on the Federal Justice Statistics Program. I want to thank you, and I hope that you enjoy the rest of your day. Goodbye.