

# Managing Data for Crime Analysis

---

In police departments, data management is the red wheel barrow upon which so much depends. It constitutes its own discipline of police science—occupying far more depth and detail than we can hope to cover in the scope of this article. Data management—a process in which all police departments engage, in some form or another—is the most crucial step in the crime analysis process, because everything that follows relies upon it. Bad data systems pollute the entire analysis process, from pattern identification to strategy evaluation. Good ones make an analyst’s job much easier and allow him to spend more time conducting crime analysis and less time tracking his own information. (In many cases, the most effective and valued analysts are “data masters”—experts at querying, manipulating, and displaying data.)

The means by which various police agencies approach this critical process are as varied as anything in police science. In many agencies, data management remains in its infancy, characterized by the collection of very little data in paper-based systems; in others, it is extremely advanced, characterized by elaborate computerized records management systems, with aggressive quality control and strong data querying capabilities. Whether an analyst finds his department’s data management processes in fine order or in mind-breaking chaos, his goal is the same:

To have **easy access** to a **quality set of timely, complete data** on which he can conduct **flexible queries** in order to find and analyze different crime phenomena.

The extent to which the department’s data management systems allow or do not allow for each of these four crucial factors determines how much time the analyst must spend entering and managing his own data. In departments with utterly intolerable systems, the analyst may find that he needs to spend half of his day simply entering and tracking reports in a database or matrix of his own design. In departments with superior systems, the analyst will be able to jump right into the process of data scanning and pattern identification first thing in the morning.

We look at each factor in more detail below. A chart at the end of this section sums up the various methods that analysts can access and manage data, and when to use them.

## Factors Critical to Data Management

### Access to Data/Querying Capability

The analyst’s first chore is getting to the data in the first place. Issues regarding data access may be political, procedural, and technological.

**Political Factors.** Some departments have made a policy decision to restrict the analyst’s access to certain data. For instance, in many agencies, information on sexual assaults is carefully restricted to sex crimes investigators only. In other agencies, the analyst may find that he cannot access investigators’ narratives or supervisors’ reports. Sometimes the restrictions are due less to confidentiality than to simple politics: the analyst’s boss is at war with the lieutenant in charge of Records, or the Administration captain opposes giving a civilian the permissions to read police reports. In such agencies, the analyst should give the agency two options: allow him access to the data he needs, or find another crime analyst.

Analysis depends on data. Restricting data from the analyst makes very little sense. Departments should ask why they feel the analyst should not have access to the data, and deal with these issues on the other side. If the department feels that the analyst might be careless with the data and release it to the wrong people, then they should develop a series of policies and procedures on information dissemination, not restrict the analyst’s access in the first place. Good analysis simply cannot be done without open access to

as much data as the department has. Restricting access to sex crimes is slightly more understandable—but patterns of rape, indecent assault, and other sex crimes do occur, and unless the police agency is confident that the sexual crimes investigators will be as vigilant as the analyst in finding patterns, as skilled as the analyst in analyzing patterns, and as diligent as the analyst in disseminating information about patterns, then they should take whatever legal steps necessary to allow the analyst access to this information.

**Procedural Factors.** Procedural problems with data access usually have something to do with a failure on the part of the agency to ensure that the Crime Analysis Unit is in the “information chain.” Information flows in various ways throughout police departments, and some departments simply neglect to give the analyst a pipeline to that flow. The Crime Analysis Unit, of course, ought to be in the center of the information flow, but it takes some agencies a great deal of time to put this principle into practice.

Ensuring procedural access to information often simply involves making sure that the analyst is copied on all incident reports, arrest reports, supplemental reports, and other documents that he requires; ensuring that he has a user account and password in the department’s Records Management System—with all necessary permissions—and so on. Departments need to remember that the analyst is often the only one who is going to do anything with most of the information that they collect, so it’s in their best interests to ensure that he has appropriate access to that information.

**Technological Factors.** Of the three factors that make an analyst’s access to information difficult, technological factors are generally the only forgivable ones. Many departments are using records management systems that are either very old or very poorly designed. Many such systems simply do not allow easy access to the records in it, or are not designed to flexibly query the records.

Technological factors are really threefold:

- 1) Ability to **export or link** to the data for use in desktop publishing, presentations, and mapping. If the records management system is separate from the agency’s desktop computer network, then the analyst may have trouble getting the data in a format that he needs. Some agencies resolve this issue with various communications tools that allow desktop computer users to TELNET to the RMS, but this is still only a workable solution if the analyst can export the data to a network server or to a disk that can be read from the same computer on which his mapping and word processing programs reside. Other departments may try to set up ODBC (Open Database Connectivity) links to the data tables used by the RMS. Good, modern RMS systems, however, are fully Windows compliant and run on a Windows NT, Novell, or other Windows-based network. These require the least creativity in accessing the data.
- 2) **Flexible querying** of the data. Even modern Windows-based systems may not have been designed very thoughtfully when it comes to querying. Some systems offer the ability to run searches only with a few fields (Incident Number, Date of Report, and Crime, for instance), which is almost useless to crime analysts. An analyst ought to be able to query every field in which data is entered. He ought to be able to ask the system for all of the house burglaries that happened in the last month between midnight and 04:00 on Sundays in Patrol Area 3 in which the suspect was a white male with a tattoo. Otherwise, what is the purpose of collecting all this data? Unfortunately, few systems are designed for such flexible querying, requiring instead that users master third-party software such as Seagate’s Crystal Reports to run “advanced” searches.
- 3) Ability to display data in a **matrix form**, which is most conducive to analysis. An analyst who runs a search like the one above doesn’t want to see the results in “report” form—he wants to see it in tabular form, so he can quickly scan the data for commonalities and pick out patterns.

Almost no commercial records management system satisfies all of these criteria, so analysts frequently export data and use ODBC links even when their records management systems are fully Windows compliant.

## Quality of Data

An analyst may have extraordinary data access capabilities, but they all become moot if the quality of the data entered in the system is bad. Data quality encompasses two areas:

- 1) **Validity**. Data entered into the system should simply be correct, particularly in regard to crime classification and crime location. The location should be the location of the offense, not the location from which the call for service came.
- 2) **Completeness**. Individuals who enter records into the system should take care to fill in all applicable fields.

Bad data quality is a fiendish problem, because it is often impossible to know when the data is bad. If a records clerk entered the wrong street address, or an officer neglected to list a certain piece of stolen property, how will the analyst ever know? He won't, and yet his analysis will still be poisoned by this problem. It is vital that the agency adopt strict quality control standards that provide sanctions for continually sloppy work.

Since the analyst is the primary connoisseur of the information in the records management system, he probably cares more than anyone else if it is correct and complete. For that reason, if no other, he should have the ability to correct entries into the system. If he cannot, and if data quality is bad enough, he may have to abandon the agency's RMS and spend a great deal of his time managing his own data in his own system, or at least export the RMS data and import it into his own database, where he may change it at will.

## Timeliness of Data

Timeliness is a chronic problem in many agencies, where records units are always three weeks behind in entering incident reports. Such a situation is intolerable and inexcusable. If the records unit is *consistently* three weeks behind, then they are entering the same *volume* every day. If the agency could afford to pay for some overtime, or hire extra help, until the records unit is caught up, then they should be *consistently* timely from that day onward.

Good crime analysis requires that no more than 48 hours pass between the receipt of a crime report and its appearance in the records management system. This doesn't mean that the analyst shouldn't have the capability to analyze crimes that occurred yesterday—but since we expect analysts to read the physical crime report every day, then it isn't absolutely necessary that those reports be in the records management system.

We applaud some agencies that require officers to enter their own reports into the system—whether through a mobile computer or at the police station—thus ensuring that data isn't only timely; it's immediate. This type of system works well as long as the agency has worked out the consequent data quality issues.

If the analyst does not have access to data 48 hours old, then his ability to find and analyze current crime patterns and series is severely stunted. He may have to compensate by tracking his own information in his own system.

## Completeness of Data

When we speak of "completeness" here, we're not worried about individuals ensuring that every applicable field is entered—that's covered under "data quality" above. We're concerned about whether enough fields *exist* within the system to capture all of the data that an analyst needs.

For instance, an analyst is rarely concerned with the exact statute that the offender violated when committing a crime—he's concerned with broader crime categorizations. At the same time, he may want to sub-divide these broader categorizations into sub-categories that have nothing to do with statute. If the

records management system doesn't allow for such custom fields, how will the analyst track this information?

There are hundreds of fields that various records management systems may or may not track. If an analyst doesn't have access to at least these types of fields, then he may have to track his own information in his own system (we've excluded fields common to every records management system, such as incident number, date of report, victim's name, and so on);

- Crime Analysis, UCR, or NIBRS Categorization
- Crime-Analysis-Defined Sub-Category
- Date of Occurrence (Should be two fields to account for ranges; some RMSs only track the date of report or offer only one date of occurrence field)
- Time of Occurrence (Should be two fields to account for ranges; some RMSs only track the time of report or offer only one time of occurrence field)
- Midpoint time (for ranges)
- Type of Location
- Police Area (may be more than one field depending on how the agency sub-divides its jurisdiction)
- Whether a suspect was described
- Whether an arrest was made
- Relationship of Offender to Victim
- Point and Method of Entry (for Burglaries)
- Weapon Used (for Violent Crime)
- Injury Inflicted (for Violent Crime)
- Type of Property Stolen
- Recovery Date & Location (for Auto Theft)
- Special codes (hate crime, domestic crime, juvenile crime, etc.)

Most records management systems will fail to track many of these important fields.

Modern database technology allows analysts to create data tables that will track the special information they want, and then link to the original tables (either imported into the database with the custom tables, or linked through ODBC protocols). That way, the analyst only needs to enter the information lacking in the RMS, rather than duplicating all fields in his own system.

## Approaches to Data Management

Depending on the nature and seriousness of the various access, quality, timeliness, and completeness issues that the analyst encounters, he may use any of the following systems to manage the data he needs:

### Using the Records Management System

Most analysts will rightfully bellow laughter at the thought of simply using their records management system to track, query, and analyze crime data. Police records management technology has lagged far behind other technologies in the past decade. Police departments today find that there are hundreds of competing vendors offering a variety of confusing packages, many of which lack the most basic functionality of modern mouse-and-menu-driven operating systems, omit important data collection fields, and fail to provide a flexible, robust querying ability.

The plethora of companies offering public safety records management products means that very few have a large market share, and thus few of them make enough money to encourage innovation and aggressive product development. Police departments are likely to remain with the same system, however insufficient, for many years, both because of the high cost of a new system and the sharp learning curve (caused in part by the lack of standards from one vendor to the next) associated with the implementation of a new system. Thus, almost a decade after the introduction of user-friendly operating systems, stable client/server networking, and powerful database technologies into the average police department, many

agencies are running records management systems on VMS or VAX servers or, even worse, on 20-year-old mainframes. Analysts skilled with desktop database programs such as Microsoft Access could theoretically develop a far better records management system on their own than the one currently used by their agencies. Some have.

Despite this disturbing situation, a handful of analysts will find themselves working for agencies with modern, intuitive, complete, robust records management systems that handle all of their analysis needs. These analysts should consider themselves very fortunate. Analysts in this situation will find little need to pursue any of the other approaches to managing data, and will be able to focus the bulk of their efforts on information scanning and analysis.

### **Creating an ODBC Link to the RMS**

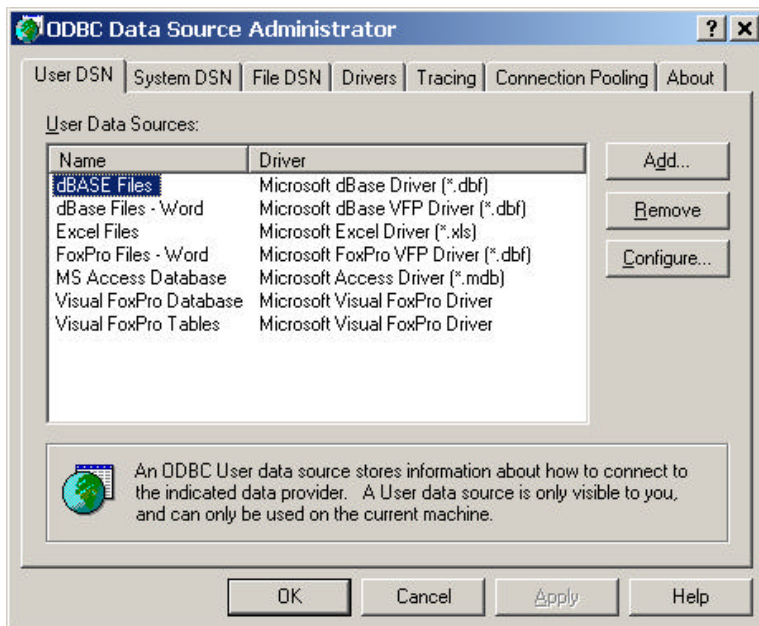
Most records management systems offer a small selection of pre-defined reports, or perhaps allow a user to query data by a few fields—and, even then, offer the results in a format unacceptable for good, quick analysis. Only an uninspired analyst will allow himself to be constrained by these pre-packaged, short-sighted tools. A good analyst will circumvent the limits of his RMS by linking to its data tables with a modern database program, such as Microsoft SQL Server, Oracle, or Microsoft Access, or with a reporting tool, such as Seagate Crystal Reports. The analyst may then run as complex and as flexible queries as his analytical mind can imagine.

We refer to our previous hypothetical query: all residential burglaries that happened in the last month between midnight and 4:00 a.m. on Sundays in which the suspect, a white male with a tattoo of an anchor on his arm, pried a cellar door. If your RMS doesn't allow for such a flexible query, then you will often need to simply ignore it and link to the data it stores.

ODBC (Object Database Connectivity) is probably the best thing to happen to databases *ever*. Developed by Microsoft, ODBC uses the Select Query Language (SQL) to connect to almost *any* database source from almost *any* Windows-based application.

ODBC allows you to open information from an Informix database in a Microsoft Excel file. It allows you to connect to a SQL Server database from a GIS program like MapInfo or ArcView. And—best of all—it allows you to bypass your old, clumsy, inadequate Records Management System and link directly to the data with a user-friendly application like Microsoft Access.

Sounds great, right? The problem is setting up the connection. First, you have to obtain the ODBC drivers for whatever database your RMS uses. Then you have to set up a connection to the database in the ODBC Data Source Administrator (this is found in Windows 2000 in Control Panel | Administrative Tools | Data Sources (ODBC)). This requires entering a number of parameters exactly right.



**Figure 1. Screen shot of ODBC control panel from Microsoft Windows 2000.**

Setting up a data source in ODBC Data Source Administrator is beyond the scope of this article. It's really beyond the scope of the average user entirely—your best bet is to get a technical representative from your RMS vendor together with your Information Systems Manager and have them set it up for you.

Once you've set up the data source, though, you can link to it from almost any application.

Truthfully, the difficulty doesn't quite end there. You'll be given a list of data tables that your RMS uses, and once you open them you'll have to figure out what the fields mean. Probably they were named by someone who started using computers when file and field names were limited to eight characters. What does POLINCDT mean? If you're lucky, your RMS vendor gives you a data dictionary to help sort it out.

## **Exporting & Importing Data**

In rare cases in which ODBC drivers are not available for your records management system's database, or when technological problems preclude the creation of an ODBC link, you may option to export data from your RMS and import it into another database format, like Microsoft Access. Once imported, you can perform the same queries that you can with linked data—you just won't have "live" information.

There is another reason for exporting and importing: when you want access to your RMS data, but you also want to make changes, and your agency's administrator's won't let you change the original data. In such circumstances, you can draw the data from your RMS, and then alter it however you like without affecting the original.

## **Linked/Imported Data Joined with Analyst's Data**

Assume that you can create an ODBC link or a successful export/import routine—however, your records management system doesn't have fields for all of the types of information you want to record. For instance, your RMS may have a field for "offense," but not a categorization field. The section on "completeness" above shows other fields your RMS may lack.

The solution is to create your own data tables that capture this additional information, and link them to the original tables through some common field, like the incident number or file number. Then you can query them together.

## Going it Alone

Finally, you may be in the worst of possible situations. Maybe your RMS is 20 years old and unusable (or you don't have one at all). Perhaps your Records Unit has a 75% error rate and is consistently five weeks behind. Maybe your IT director has never heard of an ODBC link and, anyway, your agency isn't on a network. This is when you take a deep breath, make peace with the burden that this places upon you, and start tracking your own data.

Doing your own data entry will take an enormous amount of time, and you may feel that it's not really your job. As true as this may be, in many cases it's the only way to ensure that your data is accurate, complete, and timely. The amount of time and effort it entails is worth the peace of mind that comes with a strong foundation for good analysis.

You have several options:

### *Paper Matrixes and Logs*

Analysts in the pre-desktop computer era tracked crimes in paper matrixes and crime series logs.<sup>1</sup> Each matrix or log is based on a single crime type, and it lists information specific to that type of crime. For instance, a robbery matrix might show the type of weapon used, the description of the suspect and victim, the type of property stolen, and any words spoken. A commercial burglary matrix would show the type of business, the point of entry, the means of entry, and the type of property taken.

Obviously, paper matrixes and logs are very time consuming to maintain, impossible to query (other than by scanning them with your eyes), and impractical for *all* crimes that occur in your jurisdiction. Almost any computer-based solution is better.

### *Spreadsheets*

Tracking data in a spreadsheet format, such as Microsoft Excel or Lotus 1-2-3 provides great benefits over tracking it in paper-based format. Entry is simpler, you're not limited by the physical size of a page, and you can sort, search, and filter the data.

Case #	Crime Type	Date	Time	Location	Premises Type	Victim	Offender	Property	Weapon	PIR	MUR
2002-0005	Vandalism	1/1/02	3:50	59 Pine St	Parking Lot	N/A	LNA	1999 Chev Corvett	N/A	Window	Smashed
2002-0004	Burglary	1/2/02	17:00	101 Hobart St	House	N/A	W420-20	Antique rug	N/A	Rear Door	Unlocked
2002-0051	Robbery	1/4/02	21:30	15 Main St	Conv. Store	RN06	3 W415-20	\$1000 Cash	Personal	Fr. Door	Unlocked
2002-0061	Auto Theft	1/5/02	12:30	225 Main Ave	Parking lot	N/A	Link	2001 Honda 620	N/A	N/A	N/A
2002-0137	Burglary	1/6/02	12:45	150 Elm St	House	N/A	BR425-30	Silverware, Coins	N/A	Fl. Window	Prtd
2002-0139	Robbery	1/7/02	23:35	Pine & Maple	Street	2 W417-16	3 W418-25	\$25 cash	Handgun	N/A	N/A
2002-0153	Burglary	1/9/02	9:00	200 Oak St	House	N/A	Link	VCR, TV	N/A	Rr. Window	Smashed
2002-0251	Larceny	1/9/02	12:00	10 Pine St	Health Club	N/A	Link	Walle	N/A	Locker	Prtd
2002-0261	Larceny	1/10/02	14:00	20 Second St	Jewelry Store	BP-35	AP136-45	Purse	N/A	N/A	N/A
2002-0238	Robbery	1/11/02	22:44	Main & Spruce	Park	W4195	BR115-18	Cell Phone	Knife	N/A	N/A
2002-0245	Burglary	1/12/02	10:45	30 Third St	Paint Shop	N/A	Link	Laptop Computer	N/A	Fl. Door	Smashed
2002-0223	Auto Theft	1/12/02	17:19	40 Fourth St	Driveway	N/A	Link	1985 Toyota Camry	N/A	N/A	N/A

**Figure 2. Crime data tracked in a Microsoft Excel spreadsheet.**

The drawback of a spreadsheet is that all information is stored in the same table. This is fine if all you want to track is basic crime data—date, time, offense, location, premises type, and so on—but matters become complicated if you want to track information about each person involved in each crime, or if you want to track property stolen in each crime. For instance, if you have information about two suspects and two offenders in a robbery, do you then enter four rows of information, entering the same information about the type of crime, the time the date, and the location for every row?

<sup>1</sup> See, for example, Steven Gottlieb et. al. *Crime Analysis: from First Report to Final Arrest* Montclair, CA: Alpha Publishing, 1994, pp. 143-145.

To track multiple *types* of information without duplicating effort, we need a more advanced solution.

### *Relational Databases*

The most effective solution to tracking your own data is to take the time to learn a desktop database program like Microsoft Access or Lotus Approach. Then, create your own relational database with multiple linked tables—create, in effect, your own records management system.

It takes time to learn the database application, but once you know it, you'll find that data entry is much easier than in spreadsheets and paper matrixes, and you'll learn that querying is much more flexible and robust. Later, when your agency gets a good records management system, you can still use your database knowledge to link to the RMS and query the data (see above).

## **Summary**

Data collection and management is the first step in the crime analysis process. Since good analysis requires good data, it's worth investing the time and energy to make sure your data is accurate, complete, and timely. Use your RMS if it's a good one and if it allows flexible querying; otherwise, link to your RMS from a desktop database program and use the program's more advanced querying capabilities. In desperate situations, enter and track your own information—preferably with a custom database, but a spreadsheet or paper matrix may suffice until you learn how to create a database.