

**Document Title:** Data Quality as a Challenge to Modern Policing and Criminal Justice

**Author(s):** Miran Mitar

**Document No.:** 207984

**Date Received:** December 2004

This paper appears in *Policing in Central and Eastern Europe: Dilemmas of Contemporary Criminal Justice*, edited by Gorazd Mesko, Milan Pagon, and Bojan Dobovsek, and published by the Faculty of Criminal Justice, University of Maribor, Slovenia.

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this final report available electronically in addition to NCJRS Library hard-copy format.

Opinions and/or reference to any specific commercial products, processes, or services by trade name, trademark, manufacturer, or otherwise do not constitute or imply endorsement, recommendation, or favoring by the U.S. Government. Translation and editing were the responsibility of the source of the reports, and not of the U.S. Department of Justice, NCJRS, or any other affiliated bodies.

MIRAN MITAR

## DATA QUALITY AS A CHALLENGE TO MODERN POLICING AND CRIMINAL JUSTICE

*The contribution presents some reasons why data quality can be a challenge to contemporary policing and criminal justice, then shows the characteristics of ordinary available data and the framework of theoretically proposed questions for data collection. Main elements of data quality, which are necessary for achieving data bases and data of high quality are discussed. Last but not least the need for proactive approach to data management is emphasised, because it can promote the change from experience based to theoretically founded data management.*

*"A healthy democracy needs a healthy number system, and anything that erodes that number system undermines democracy" (Kenneth Prewitt, American Sociological Association, 1999).*

*Data is the Latin word for 'things given'; it might be more appropriate to call it (it collectively, or them, if you are thinking about individual items of information) *capta*; 'things taken', or, indeed, wrested with a great deal of effort from a recalcitrant social world.*

*Michael Procter, 1993*

### INTRODUCTION

The main purpose of my contribution is to present some reasons, why questions about data quality can be a challenge to contemporary policing and criminal justice, especially in societies in transition from authoritarian to democratic political systems. The second purpose is to show the characteristics of ordinary available data and the framework of theoretically proposed questions for data collection. The third purpose is to present main elements of data quality, which are necessary, if any administrative data producing organisation wants to have data bases and data of high quality. The fourth purpose is to show the need for proactive approach to data management, which can cause the change from experience based to theoretically founded data management. The contribution is one result of author's study and reflection about problems how to make theoretically sound and empirically based assessments of security of contemporary societies.

### WHY CAN BE DATA QUALITY A CHALLENGE

Data quality can be a challenge for police organisations in new democracies (in Central and East Europe and elsewhere), at first, because it is supposed, that the police and criminal justice organisations in the changed political climate and social environment have the duty to present public the real data (of high quality) about the contents and procedures of their work, and at second, because the scientific developments in the fields of information gathering, processing and analysing (besides the official crime statistics there are the social science data archives, which store research data from opinion polls, victim and self-report surveys), enable that social scientists, various researchers, journalists and members of critical public can detect various inconsistencies of official data published by national police and criminal justice organisations (Riedel, 2000). This situation can be a challenge for credibility of data producers, if they do not

improve their activities in the field of data gathering, processing and analysing and if they do not open their administrative statistics to scientific and public use and overview. Such situation has not only influences on the degree of openness, transparency and responsibility of police and criminal justice organisations, but can also be one of the most important steps in the changes of structures and functions of police and criminal justice organisations, from pyramidal hierarchies to more horizontal network systems and from reactive incident based style to more proactive evidence-based policing (Lawrence Sherman, 1998).

## **WHAT KINDS OF DATA DO POLICE AND CRIMINAL JUSTICE ORGANISATIONS USUALLY COLLECT**

Police and criminal justice organisations collect usually data about various events from the field of "internal security", which is usually divided in more sub-fields (traffic, public law and order, crime, border). There are various events (traffic offences and accidents, offences of public law and order, crime offences, border offences), which fall in the area of responsibility of various police and criminal justice organisations. They collect data about events and involved persons, but also the data about their own work. Systems of data collections are heterogeneous (theoretically and methodologically), usually decentralised, so it is difficult to have the holistic overview of the field, what is happening and what some organisations are doing at specific time points. Because there are no explicitly defined links in the process of work (and data) through the criminal justice system, from police, to courts and corrections, so there is difficult to assess the ongoing processes, outputs and outcomes through the time.

The main purposes of data gathering efforts of police and criminal justice organisations are determined broadly by constitution and law. But in practice the organisations collect mainly data, which they need to detect and investigate various offences and offenders from the point of view of criminal law, then they collect data about their own work. Both kinds of data are presented in their "case records", the amounts of items of information are usually great, but the quality is a variable, which depends on amount and quality of prior work of personnel involved in process of investigation and data collection in various cases. One of the most critical step is a step from "case record" to "statistical record", where many useful information may be omitted because of the characteristics of coding systems, which are not theoretically and methodologically explicitly defined. In spite of this, it is possible to obtain great amounts of useful information about events and work of organisations (for example, in Slovenia), but not in a comprehensive and systematic manner.

## **WHAT KIND OF DATA ARE NEEDED**

For comprehensive overview of the work in the field of criminal justice, we need firstly the theoretically and methodologically explicitly defined data about events, which form the content of work of police and criminal justice organisations. Also we need data about the various kinds of procedures, which were needed for investigation of events (so called cases). Both aims can be realised by the help of searching and finding answers on classic golden questions of criminal investigation (Table 1).

**Table 1: Overview of needed data about events and work of police (Mitar, 2000)**

	Characteristics of events		Work of police organisations
1	The first definition of event (victim, third person)	A	How, when and where had organisation get knowledge about event?
2	Who are the actors (offenders)?	B	Who (person, organisational unit) did got information, who managed the case and who collaborated?
3	Who are victims?		
4	Are the third persons present (or involved)?		
5	What was happened (an event or events) as result of combination of actions of one or more involved persons?	C	What kinds of actions did officers and organisational units take in solving the case?
6	With whom did the actors collaborate?	D	With whom did the officers collaborate? (without organisation)
7	How did the actors do?	E	How did the officers get information about events- directly from persons involved or indirectly in the process of criminal investigation?
8	What means did the actors use?	F	What means did the officers use to get information? Other means (cars, computers, ...)?
9	What are consequences of event (formal definition of event)?	G	What were effects of officers' work? (solution of case; use of various legal means for treatment of offence and offender)
10	Where did the actors do their acts?	H	Where did the officers do their work?
11	When did the actors act? When did consequences appear?	I	When did the officers do their work?
12	Why did they act?	J	What are (positive/negative) factors of success of work of police in the case?
13		K	Work with documents (opening, completion, end)?

Comprehensive overview of data of police and criminal justice organisations is (or can be) useful from various viewpoints:

- the needs of overview of quantitative and qualitative characteristics of events and involved persons;
- the needs of overview of overloads of organisations and assessments of quantity and quality of their work;
- the needs of searching and finding connections between various forms of activities and their results (what works and what does not work in some cases) and
- the needs of searching and finding connections with the characteristics of social and political contexts (characteristics of communities, characteristics of economic and social policies, ect.).

Such data can be used for various purposes, not only for the usually narrower purposes of police and criminal justice organisations (getting their job done), but also for broader purposes of assessment of conditions in the various fields of internal security, for assessments of activities of police and criminal justice organisations, and last not least the assessment of contextual conditions and their influences on various aggregations of events (criminality, for example) and work of police and criminal justice organisations.

## DATA QUALITY

There exists many definitions of data quality, but the starting points for discussion are usually the two basic methodological requirements for measurement in the scientific research:

- problem of validity of measurement (probability, that the measure truly represents the underlying concept) and
- problem of reliability of measurement (probability to getting similar or same results in repeated measurements).

Then some authors (Howard, Newman and Pridemore, 2000) define other components of data quality such as **objectivity** (Quality data are information that is accurate and unbiased and presented in a clear, complete, well-documented manner. Objectivity is achieved by using appropriate data sources and sound analytic techniques, by using proven methods, and by carefully reviewing the content of all information and reports.), **integrity** (Data with integrity is information that is not compromised through corruption or falsification.), transparency (Transparent data involve having a clear description of methods, data sources, outcomes, and related information that allows users to understand the data.), **reproducibility** (Quality data can be reproduced by others by using the documented methods, assumptions, and data sources to achieve comparable findings.) and **utility** (Quality data are information that is useful and available to its intended audience).

Similar dimensions are defined by Condelli et al. (2002) (Table 2).

*Table 2: Six dimensions of Data Quality (Data in statistical agency\*)*

<i>Relevance</i>	The <b>relevance</b> of statistical information reflects the degree to which it meets the real needs of clients. It is concerned with whether the available information sheds light on the issues of most importance to users. Assessing relevance is a subjective matter dependent upon the varying needs of user. The NSO's challenge is to weigh and balance the conflicting needs of different users to produce a program that goes as far as possible in satisfying the most important needs and users within given resource constraints.
<i>Accuracy</i>	The <b>accuracy</b> of statistical information is the degree to which the information correctly describes the phenomena it was designed to measure. It is usually characterised in terms of error in statistical estimates and is traditionally decomposed into bias (systematic error) and variance (random error) components. It may also be described in terms of the major sources of error that potentially cause inaccuracy (e.g. coverage, sampling, nonresponse, response).
<i>Timeliness</i>	The <b>timeliness</b> of statistical information refers to the delay between the reference point (or the end of the reference period) to which the information pertains, and the date on which the information becomes available. It is typically involved in a trade-off against <i>accuracy</i> .
<i>Accessibility</i>	The <b>accessibility</b> of statistical information refers to the ease with which it can be obtained from the NSO. This includes the ease with which the existence of information can be ascertained, as well as the suitability of the form or medium through which the information can be accessed. The cost of the information may also be an aspect of <i>accessibility</i> for some users.
<i>Interpretability</i>	The interpretability of statistical information reflects the availability of the supplementary information and metadata necessary to interpret and utilize it appropriately. This information normally covers the underlying concepts, variables and classifications used, the methodology of collection, and indications of the accuracy of the statistical information.

<i>Coherence</i>	The coherence of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic framework and over time. The use of standard concepts, classifications and target populations promotes coherence, as does the use of common methodology across surveys. <i>Coherence</i> does not necessarily imply full numerical consistency.
------------------	--

Gertz et al (2004) made survey of 179 data quality dimensions from the filed of management information systems and found four data quality categories (Table 3)

**Table 3: DQ categories and dimensions**

DQ Categories	DQ Dimensions
Intrinsic DQ	Accuracy, objectivity, believability, reputation
Accessibility DQ	Accessibility, access security
Contextual DQ	Relevancy, value-added, timeliness, completeness, amount of data
Representational DQ	Interpretability, ease of understanding, concise/consistent representation

The overview of DQ dimensions shows that data quality is not a simple concept, which can be easily defined and measured. But it also shows that there exists a some kind of minimal agreement of basic elements of data quality, which the people need for various purposes and on various level of work (operational, managerial).

These dimensions can be useful for assessment of data quality of existing databases of police and criminal justice organisations. It is easy to suppose that many inconsistencies can be find and that data quality can be improved.

## A PROACTIVE APPROACH TO DATA MANAGEMENT

Starting point for a proactive data management is that data systems must be developed from the comprehensive viewpoints, which includes on the one side the views of management and operational levels of police and criminal justice organisations, on the other side there must be included the data analysis and auditing needs.

In the field of criminal justice it is important to make first steps and find the answers on following four questions (Crime reduction toolkits,2004):

- What tasks do we need to carry out when auditing crime and disorder?
- What data will be required in order to do this?
- What data are currently available?
- What are limitations of current data?

The next step to develop a strategy for managing data is to find answers on set of questions (crime reduction toolkits)

- What action is required to realise data requirements?
- Who is going to take responsibility for improving the availability /quality of data?
- What targets have been set for improving data management?
- Does the Crime and Disorder Strategy contain a strategy for data management?
- Is the strategy for data management in the Crime and Disorder strategy reflected in agency's own strategies?

## CONCLUSIONS

Data quality (including its main dimensions) is a necessary, if the producers of data-bases want to achieve and maintain the credibility of source both from the viewpoint of information gathering, processing and analysing and from the viewpoint of political honesty in democratic society. The existence of data bases of poor quality can be interpreted as the result of inability (lack of resources and knowledge) to produce better data, but also as a result of management's neglect of the needs of contemporary democratic societies for data, which would enable a greater degree of transparency, oversight and responsibility of police and criminal justice organisations. Such neglect hinders also the development of new forms of policing (community oriented policing, problem oriented policing and evidence- based policing). At the end we can pose the question, who is responsible for data quality (Overview of Data Quality Assessment Tools in Connecticut, 2004).

## ABOUT THE AUTHOR

**Miran Mitar**, MA, Senior lecturer, Faculty of Criminal Justice, University of Maribor, Slovenia.

## REFERENCES

- Brackstone, G. (1999). Managing Data Quality in a Statistical Agency. In Statistics Canada, Survey Methodology, Catalogue No. 12-001-XPB, Vol.25, No2.
- Condelli, L., Castillo, L., Seburn, M., Deveux, J. (2002). Promoting the Quality and Use of National Reporting System (NRS) Data, American Institutes for research, Washington.
- Home Office (2004). Crime Reduction Toolkits, A Proactive Approach to Data management. (<http://www.crimereduction.gov.uk/toolkits/p031006.htm>) (20.06.04)
- Gertz, M., Ozsu, T., Saake, G., Sattler, K. (2004): Data Quality on the Web (<http://www.sirius.cs.uc.davis.edu/Dagstuhl03/dq.pdf>) (20.06.2004)
- Howard, G., Newman G., Pridemore, W.A. (2000). Theory, Method, and Data in Comparative Criminology, in Criminal Justice 2000, Vol.4 Measurement and Analysis of crime and justice.
- Mitar, M. (2000). Uvod v metodologijo znanstvenega raziskovanja varnostnih pojavov (splošni del). Visoka policijsko varnostna šola, Ljubljana.
- Overview of Data Quality Assessment Tools in Connecticut ([http://www.jrsa.org.ibrrc/data\\_quality/Conncticut/overview.html](http://www.jrsa.org.ibrrc/data_quality/Conncticut/overview.html)) (20.06.2004).
- Procter, M. (1993) Analysing other Researcher's data. In: Gilbert, N. (ed.). Researching Social Life. Sage, London, p.255-269.
- Riedel, M. (2000). Research strategies for Secondary data, A perspective for Criminology and Criminal Justice. Sage, London.
- Sherman, W.L. (1988). Evidence-Based Policing. (<http://www.policefoundation.org/pdf/Sherman.pdf>)