Integrity and Internal Control in a Statistical Environment

Wouter J. Keller and Roel G.J. Rot

Statistics Netherlands (CBS)

Division Research and Development

Department of Information Technology Heerlen

Heerlen, The Netherlands

Keywords: Integrity, Security, Quality, Control

Abstract:

In a statistical environment like that of Statistics Netherlands, correctness of data and information is very important. A wide range of measures deals with the preservation of Integrity. In this paper we will show how Statistics Netherlands maintains integrity and security in general throughout the organisation and how internal controls are established.

1. INTRODUCTION

Statistics Netherlands is a data and information-processing factory. Its customers, businesses, individuals, National and European Government, expect the information to be correct or at least as good as possible. Furthermore they expect Statistics Netherlands to handle data and information in a correct way, according to national and international law and regulations, living up to the expectations of the society Statistics Netherlands is embedded in.

The care for quality has been important since the foundation of Statistics Netherlands in 1899. Particularly the statistical quality has been of great significance. Integrity is a major aspect of statistical quality. It concerns the logical match between the statistical results and the observed and surveyed reality, but also the security of the published data and tables towards the respondents. With the introduction of automated information systems in the early sixties technical integrity, became of more importance by and by. Technical integrity deals with the preservation of data and information in a well-defined and well-known state. Not that it had been neglected formerly but with the coming of automated systems that are merely black boxes wherein data reside, it became more obvious that care had to

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: [10.1007/978-0-387-35501-6_14](https://doi.org/10.1007/978-0-387-35501-6_14)

be taken of the integrity of data. From then on regulations and rules have emerged that aimed at security in general and integrity more in particular. However it is only since recent years that standardised security frameworks were published and made compulsory. The Regulation Information security for Government service (VIR) is one of them. A short introduction follows below.

This paper deals with various aspects of security and integrity that are recognised by Statistic Netherlands. The general framework regulation compulsory to Statistics Netherlands will be discussed in paragraph two. The next paragraph deals with the way Statistics Netherlands implements security and gives some examples of security rules and policies that are enforced within the organisation. In paragraph 4 the controlling mechanism is discussed. In the next paragraph a collection of outstanding problems and of possible future developments is shown.

2. SECURITY REGULATION FOR DUTCH GOVERNMENTAL SERVICES (VIR)

The Security regulation for Dutch Governmental Services (VIR¹) has become into force on January 1st in 1995. It governs information security policy for all Governmental Institutions. As automation strongly increased in the eighties there arose a need for a standardised framework for information security and control. Information security is seen as a vital part of the quality of information systems and the dependent processes. Standardisation would lead to a reduction of costs – no re-invention of the wheel – and more easily controllable security environment. Above that external contractors can be offered an equal approach for information security for all Dutch Governmental services.

The VIR tenets are controllability and responsibility. Controllability means that the compliance with the security measures should be controllable. The responsibility for VIR Security rests with the manager that is appointed by the Security Policy. A Security Policy is made compulsory by VIR. Each information system has to be analysed to determine a set of security measures that is sufficient and complete. The analysis process consists out of two parts. The first part brings the security demands for the information system, in terms of availability, confidentiality and integrity. The second part produces the security rules by setting security demands alongside with the menaces to the information system.

¹ VIR : Voorschrift Informatiebeveiliging Rijksdienst

3. IMPLEMENTATION OF INFORMATION SECURITY WITHIN STATISTICS NETHERLANDS

Security is implemented on various levels within Statistics Netherlands. On the highest level the Security Policy for Statistics Netherlands operates. All other security guidelines, measures and implementations are dependent of the Security Policy.

Security has a long history within the organisation. The policies on security have always followed consequently legislation and official Decrees. For all relevant fields security rules are implemented according to the intent of the laws and Decrees. With the introduction of VIR, information security had to be organised in a uniform way to obtain controllability, functionality, effectiveness, and completeness of the implemented and proposed rules. As we have seen in paragraph 2, the responsibilities for information security were assigned, especially for the organisation covering services:

- Personnel System
- Financial System
- Security of offices, buildings, staff and materials
- Transport
- Central Archive
- Statistical security
- Acquisition and exploitation of IT facilities and services
- Acquisition, development and exploitation of software programs.

Security baselines are specified for all the listed services. The baselines contain minimum requirements for maintaining security on availability, confidentiality and integrity. If there is more security needed or wanted this is possible; but it has to be demonstrated logically. Baselines are also specified for a well-defined number of processes that take place within Statistics Netherlands and that are essential for running the organisation. For example the baselines for preparation of micro data files and for decentralised – local at the statistical departments - development of software programs.

From the collection of security policies and rules that is far too elaborate to reproduce in its entirety, I have selected four items that are of special interest to Statistics Netherlands and that have direct links to integrity in one of its fields. In the next sub-paragraphs I will discuss “informed consent” that deals with general integrity of the organisation, “the network layer model”, technical integrity, “statistical security”, statistical or logical integrity, and finally “standardisation”, technical integrity.

3.1. Informed Consent

An organisation has to be sincere about its aims and the way it organises its production process. For statistical organisations like Statistics Netherlands this is of special interest. Statistics Netherlands depends on individuals (voluntary) and companies (obligatory) for its input of raw data. But if there is no good trust in the statistical organisation then it is not very likely that good and timely data will be provided. That might harm the statistical organisation and hamper the production of statistics.

The concept of informed consent has been known in the medical world since quite a long time. Patients are fully and wholeheartedly informed about treatment, medical acting and hazards, and give their consent by signing a form. This concept can also be made applicable to the field of statistical surveys: So what does that mean for the current practice of statistical data collecting and data handling? The process of data collecting has become more complex in the last decade. In the past Statistics Netherlands had a very direct relation with its data-subjects that were polled. A pollster visited a respondent at home or contacted him or her by telephone. Nowadays data may be also collected from administrations. Subsequently survey and administrative data may be matched on an individual level. The relation with the respondent then becomes more complicated. An example of such approach is given in the text frame. The Statistical Law for the Netherlands authorises Statistics Netherlands to access governmental administration databases on the condition that the purpose is uniquely statistical. The Privacy Law of the Netherlands makes also exemptions for statistical processing and research in general. The principle of informed consent implies that the respondent also should understand that his or her answers are supplemented with administrative data, exclusively for statistical purposes, before the questionnaire is administered. Such information is given in an introductory letter and a further explanatory brochure is available at request. Statistics Netherlands acts very carefully in these matters.

The Netherlands' statistics on **registered employment** are based on a combination of administrative unemployment data with the continuous labour force survey. Addresses drawn from the administrative files have a greater probability of being approached for the sample survey. Fieldwork results and administrative data are matched on the basis of sex, date of birth, and postal code. For purposes of statistical inference to national totals, a larger number of variables common to both data collections is used. The correlation between survey and administrative data on the question itself of being registered is far from perfect, by the way. A contract signed in 1987 between Statistics Netherlands and the Minister of Social Affairs is the foundation of this approach. It has been discussed by Parliament. Individual respondents can be, but probably are not, aware of the matching involved in the further use of their data.

Figure 1. An example of matching of data ("Informed Consent: Buzzword or Panacea?" [2])

For micro data files that contain detailed data and are provided to third parties for purposes of statistical and scientific research, the same considerations apply. Dutch legislation authorises Statistics Netherlands to provide micro-data for research purposes. But in the view of Statistics Netherlands respondents shall be made explicitly aware of the use of the data they provided. By proceeding with the interview they give their consent.

3.2. Statistical security

On of the aspects of the integrity of a set of statistical information such as statistical tables and micro-data is the disclosure risk. This is the risk that private and confidential data will be disclosed from the anonymous data in the set or will be inferred by relating it to other data sets published by Statistics Netherlands. In general the demand is that a set should contain a maximum of information and minimise the risk of disclosure at the same time.

Prevention of statistical disclosure is of major importance for Statistics Netherlands. National and International laws grant individuals and companies the right of privacy. There is also the consideration that a firm and sound security policy on statistical disclosure might enhance the willingness to take part in statistical surveys. Last but not least Statistics Netherlands aims at presenting data and information that complies with the highest standards of quality and security. Therefore the research on statistical security and in particular statistical disclosure always has had a high priority. As a result a theory on security and disclosure risks have been developed and published. From that a policy is adopted that standardises and rules the security for all kinds of data and information that is published or distributed. Moreover it is widely accepted within the organisation. And this is very important, because without the co-operation of the whole staff it is virtually impossible to maintain security and integrity in particular.

A brief introduction to statistical security

Two sets of data to secure can be distinguished: information to publish and so-called micro-data. The first is readily available to everyone at low costs. For example StatLine the statistical database that is accessible via the Statistic Netherlands site on the Internet at "<http://www.cbs.nl>". This database contains almost all the data that Statistics Netherlands publishes. On the other hand the micro data files hold low-level data for research purposes. These files are provided to research institutions under the most severe security stipulations, laid down in a contract. The information to publish is in itself more solid secured than the micro-data. Micro-data files have the contract as main part of the security although a set of strict security rules applies as well.

The general demand for the protection of micro data is that those files only have to be secured against spontaneous disclosure. Certain combinations of data may produce an aha-experience to the researcher while working with micro data. Without

the intention to look for a specified person or company the researcher recognises suddenly the survey respondent. For example one can think of combinations like a Chinese dentist in Maastricht or an astronaut in Amsterdam. These combinations give a good or even absolute likely disclosure chance. On the other hand if combinations occur quite often the researcher will not likely recognise respondents. For that purpose the rules demand that a minimal number of respondents from the population have the same identifying keys. If that number is not matched the lines or keys have to be withdrawn from the file. Combinations of rare keys a mentioned above will not longer occur in the micro data file. In contrast to micro data files it is a strict demand for publication files or tables that statistical disclosure has to be prevented at all costs. In this case prevention of disclosure counts for unintentional as well as for intentional disclosure. The intentional disclosure demand is the main difference to the security for micro data files and publication files. In publication files like statistical tables, sensitive data are not allowed. Moreover every combination of identifying keys has to be matched and examined on disclosure risks. This is not the case with micro data files where only some combinations have to be examined. Variables that can be publicly observed and that combine with other variables to very rare combinations have to be matched. Examples of such variables are sex, profession, ethnicity, nationality, and native country. Combinations of sex with profession or ethnicity with profession may lead to a high identification-risk.

Figure 2. Masking in publication tables

To facilitate the protection of statistical tables and micro-data, the research department has developed a tool called Argus that automatically applies correctly security rules.

3.3. 4-layer network model and the network gate-lock

To obtain a high level of security the inner networks on which reside sensitive data, are locked away behind three barriers of protection. We call that a 4-layer model in which the inner networks form the first layer and the outside world is the forth layer. In between there are two network layers that constitute an effective wall against hacking and other menaces.

Figure 3. The four-layer network model for Statistics Netherlands

Each layer has its own well-defined security policy. The first layer, the inner networks should be absolutely inaccessible for unauthorised access from the outside world, not to mention computer break-ins. On the inner networks reside data that have to be protected by all means. It is also extremely important that the integrity of the IT infrastructure keeps maintained as the whole staff and virtual all (statistical) processes are dependent of it. To obtain an adequate protection a so-called network gate-lock constitutes the connection between inner networks and the second network layer. As with a ship-lock where there is no direct connection between the downhill and uphill waterways, the network lock prevents a direct and continuous connection between the inner networks and the second layer network. Instead it gives controlled access to both of the networks and by its nature it allows only mail to pass. In the early nineties the concept of the network gate-lock was developed, as there arose a need for electronic connections with the outside world. It was obvious that the LAN's had to be secured from unauthorised access while at the same time a gateway for email should be in place. The network gate-lock developed then, was a mechanical lock that had electrical relays as its main components. Almost a decade later ever increasing volumes of mail and files, forced us to adopt the principle of shared disks. Nowadays the gate-lock consists of a master and client gate-lock server with one (conceptual) shared disk that is accessed by only one of the servers at the same time.

The access policy for the inner networks demands that only mail will be passed through the gate-lock. That mail should be clean and harmless. Therefore all mail is swept by a mail-sweeper that analyses mail into its finest details. Tests include a virus-scan of all components. If there is an indication of infection, the mail is put aside and put into quarantine. Programs of all kinds are stopped right away. Statistics Netherlands does not allow strange programs that not belong to the standard office automation, on its inner networks.

The level two and three networks can be accessed from the outside world. These networks carry dedicated services for data-communication, Internet access and security. Firewalls protect the connection between layers two and three and four (the outside world, Internet, x400, x25, isdn, and telephone). The configuration of the firewalls depends highly on the applications that function on layers two and three. The main goal is to prevent the outside world to get more information from or place information on the networks, than is needed or is allowed.

3.4. Standardisation

Standardisation of methods, software and hardware has always an objective of security policy of Statistics Netherlands. Regarding the methods we have looked into the area of Statistical Security in section 3.2. This section will mainly deal with the standardisation of software and hardware.

In the era of mainframes and mini-computers – not really mini, you had to reserve a room for it – standardisation was no big deal. It all was too centralised and too costly to allow alternative solutions. Standardisation came by itself. With the coming of the PC in the mid-eighties this was less obvious. Departments were likely to deviate from the rule especially as such rule did not exist. However with strict regulation that assigned selecting, buying, installing and maintaining of hardware and software to the central IT department, Statistics Netherlands managed to create a standardised IT infrastructure. And this is the case to this very day.

Standardisation goes very well with protection. By itself it is a mean of security and it certainly adds to the integrity of the IT infrastructure. Through standardisation in the early stage of office automation, Statistics Netherlands was able to introduce some drastic security measures. All servers are placed in a secured central environment. Workstations lack diskette-and or CD-ROM-drives; the space to use on the fixed-disk is very limited. Import from and export to diskettes is only possible on special stations attached to the second layer net, outside the inner networks. Imported or exported files are mandatory routed through a mail sweeper and the network-gate-lock; for these files the same rules as for mail apply. All applications and software that can be run from a workstation can be installed or made available from scratch via a standard installation procedure. Every member of the staff is able to run the installation procedure. Most applications are installed centrally on a program sever in the secured environment.

Data is stored on centralised dedicated servers; a back up is made every night. The statistical department bears themselves the responsibility for archiving data. For the aim of archiving two on-line StorageTek robots are available, each with a maximum capacity of 12 terabyte. One of the archive robots is placed in the office of Statistics Netherlands in Voorburg, The Hague, the other one in Heerlen. For high security needs the option exists to archive in both offices.

4. CONTROLLING SECURITY AND INTEGRITY

In principle the Security Policy governs the control of security and integrity in particular. The policy demands a yearly internal and external control on security schemes and implemented rules. Internal and external accountants are ordered to do so. These high-level controls do not directly deal with statistical security or the implementation of a sound virus-scan for example. To ensure control on these

matters, every department has its own security officer. The security officer has to take care of the implementation and proper functioning of the required measures. Control is facilitated as all measures are described in security documents. The security officer reports to the manager that has the responsibility for the processes and information systems.

From way back a lot of controls on integrity, plausibility and completeness have been included in software programs for statistical processing. We think that it is important that the data are as correct as possible straight from the beginning of the statistical collection process. Software programs for collecting data or designing electronic inquiries developed by Statistic Netherlands such as BLAISE, EDISENT and IRIS, do have the options to include various levels of controls on input.

Control on the statistical integrity begins with the design of the survey. Statistics Netherlands has a lot of experience in designing statistical surveys. Research on statistical methodology and non-response has to add to the statistical quality and integrity of the results and to the fact that controllable statements can be made.

Automatic controls and monitors are applied increasingly, especially within the IT department where many processes and machines that are running around the clock, need surveillance. These controls automatically report to staff just in time.

5. PROBLEMS IN MAINTAINING INTEGRITY AND FUTURE DEVELOPMENTS

Maintaining integrity is not a sinecure. Constantly new threats appear on the horizon. Mail for example tends to become more complicated. Already now a major part of HTML documents carry JAVA or ActiveX applets. Mail-sweeper sees such an applet correctly as a program and stops it at the front door. Some of these small programs have disastrous consequences once read and activated. At the same time applets may be absolutely necessary to show the mail to its full extent. Right now there is a need for sophisticated controls that can effectively discern between good and evil applets. But such controls do not exist today. The integrity of inner networks and its sensitive data is at stake. To cope with these problems research has been started that deals with finding a new network infrastructure and sophisticated ways to analyse mail and files. A totally secluded environment without even email possibilities is envisioned. This area should become the home of all statistical data. In an area other than that a more open network might exist. It would open the possibility to work at home; the electronic agenda's can be synchronised. At this moment there exists an increasing need for such open communications.

Not only the threats from the inside count. It is within the organisation that personnel have relatively easy access to the data. The risk of intentional and

unintentional misuse has to be minimised. The present division in logical separated networks where a user needs separate authorisation for any of the networks and data area's, works well and is considered secure enough for the moment. Nevertheless it may be necessary to turn to a Kerberos like security environment, where any access to a procedure, operation or data has to be requested and granted. In that respect the development of the currently operations systems in use, look hopefully. Last but not least there is also the personal integrity. Through information and security-oriented education this should be kept upright. To keep its staff well-educated and well-informed Statistics Netherlands goes to great lengths. Knowledge-management will be a aid to achieve that goal.

Statistical integrity is threatened by decreasing response figures. This non-response problem is dealt with by intensive research; and so far that has been successful. But there are some groups and phenomena in the present society it might need a different approach than the methods of sampling. Data-collection from external databases maybe a solution to the increasing non-response problem. Such access is granted by statistical and privacy laws. But that is not enough. Public approval is needed as without it politicians and lawmakers might review their position on this ticklish subject.

6. CONCLUDING REMARKS

The collection of security measures presented in this document shows there is a need for security and in particular integrity within the statistical organisation. Apart from the statistical integrity it is difficult to make a quantitative assessment about the size of this need. The theory and the rules of probability, combinatorics, statistical testing, and sampling govern the statistical integrity. The approach towards technical and physical security is more qualitative one; it always has been. It means that for example the need for integrity is assessed in terms as essential, important and unnecessary rather than in fixed numbers. It is a demand that the assessment has to be logical well founded. In this way there is no need for complex methods and difficult calculations. Logical reasoning and expertise in the process to secure is what counts. Security demands can be and have to be determined by the manager responsible for the process.

As almost all the technical measures that guarantee integrity have backdoors it is necessary that the whole staff will work co-operatively and faithfully in line with the security rules. This is the only way integrity throughout the organisation, throughout all processes and throughout all data can be ensured. Convincing staff and providing tools that make integrity self-acting is an everlasting need.

Statistics Netherlands considers security as extremely important. It will tighten the security to a degree that complies with the demands that come from inside and outside as well. We know our future depends on it.

7. REFERENCES

1. Voorschrift Informatiebeveiliging Rijksdienst 1994, Ministry of the Interior
2. Informed Consent: Buzzword or Panacea, Joris Nobel, Statistics Netherlands 1999
3. Handboek Statistische Beveiliging, Statistics Netherlands 1995, 1999
4. Security Policy, Statistics Netherlands 1997
5. Baseline Statistical Security, Statistics Netherlands 1998
6. Baseline IT security, Statistics Netherlands 1999
7. Standard Service Level IT services, Statistics Netherlands 1999
8. Using Principles of War and Knowledge-management, Lt. Col. Perry Luzwick, Information Security Bulletin, July 1999, p. 17 – 28