Workshop on access to administrative data sources

Brussels, 13-14 September 2016

Workshop technical report
The information and views set out in the report are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Authors:

Willem De VRIES (Agilis SA)
Photis STAVROPOULOS (Agilis SA)
Anna AMPATZOGLOU (Agilis SA)
# Table of contents

1. Introduction .................................................................................................................. 2
   Objectives of the workshop .......................................................................................... 3
   Conclusions of the report ............................................................................................... 3
   Report structure ............................................................................................................ 6

2. Workshop sessions ........................................................................................................ 7
   2.1. Plenary session – 1st day ....................................................................................... 7
   2.2. Parallel Session 1 - Legal framework for statistics and general policy on data access and use ................................................................................................. 9
       2.2.1. Objectives of the session ............................................................................... 9
       2.2.2. Summary of the main issues of discussed .................................................... 10
   2.3. Parallel Session 2 - Permanent mechanisms of cooperation with data owners ......................................................................................................................... 11
       2.3.1. Objectives of the session .............................................................................. 11
       2.3.2. Summary of the main issues of discussed .................................................... 11
   2.4. Parallel Session 3 – Cooperation with data owners on quality issues ............... 13
       2.4.1. Objectives of the session .............................................................................. 13
       2.4.2. Summary of the main issues of discussed .................................................... 13
   2.5. Plenary session – 2nd day ...................................................................................... 15
   2.6. Parallel Session 4 – General institutional environment and communication policies ........................................................................................................... 17
       2.6.1. Objectives of the session .............................................................................. 17
       2.6.2. Summary of the main issues of discussed .................................................... 17
   2.7. Parallel Session 5 – NSI best practices related to internal management; financial aspects of accessing administrative data ...................................................... 19
       2.7.1. Objectives of the session .............................................................................. 19
       2.7.2. Summary of the main issues of discussed .................................................... 19
   2.8. Parallel Session 6 – Cooperation with data owners on technical issues; dealing with privacy issues ................................................................. 20
       2.8.1. Objectives of the session .............................................................................. 20
       2.8.2. Summary of the main issues of discussed .................................................... 21

3. Conclusions of the workshop ......................................................................................... 23

Annex I. Workshop programme ......................................................................................... 24
1. Introduction

The promotion of the use of new sources, including administrative data, for the compilation of official statistics constitutes one of the five key areas of the ESS Vision 2020. The European Statistical System (ESS) Vision Implementation Project on Administrative Data Sources (ESS.VIP ADMIN) (2015-2019) was launched with the view to provide assistance to the ESS Member States on the wider use of administrative data.

The presentation of Mr Baldacci of behalf of Eurostat in the opening session of the Workshop on access to administrative data sources, dealt with main priority areas and goals of the ESS Vision 2020 as well as the current state of the art of the ESS.VIP ADMIN project, its scope, objectives and challenges.

Recent developments stemming from the ESS Vision 2020 imply a move towards a multisource environment, where the potential of new sources should be exploited and investments in new IT tools and methodology should be made so as to continue improving the existing data collection methods. The three pillars in order to achieve this target are (a) reviewed and reinforced legal frameworks, (b) common and verifiable IT security standards, and (c) tailored and timely communication among all ESS partners and the data owners.

The ESS.VIP ADMIN project is organised in seven work packages (WP), addressing typical challenges confronted in accessing and using administrative data. It addresses issues with regard to access to administrative data sources (WP1), statistical methods (WP2), quality (WP3), use of Commission data (WP4), sampling frames (WP5), pilot studies (WP6) and methodological support to Member States (WP7). Under the framework of the ESS.VIP ADMIN project, Eurostat offers the National Statistical Institutes (NSIs) grants for improving the use of administrative data for statistical production.

The Workshop on access to administrative data sources, held in Brussels on 13-14 September 2016, is linked to WP1, which aims at the gathering information on the current access situation in the EU Member States and EFTA countries, the exchange of best practices on access to administrative data, dealing with cooperation issues with the data owners and the analysis of further actions to improve access to administrative data.
Objectives of the workshop

The Workshop focused on the legal, institutional and organisational issues related to the access to and use of administrative data for the production of official statistics. The aim of the Workshop was to facilitate information exchange and sharing of experience concerning the legal and institutional environment, mechanisms of collaboration with administrative data owners, cooperation with administrative data owners on quality and technical issues, National Statistical Institutes’ (NSI) efficient internal management procedures and practices, and actual implementation obstacles.

The expected outcome of the Workshop was to provide insight on the difficulties that still persist at national level in accessing administrative data sources and on the practices adopted by the NSIs to overcome such difficulties.

The workshop has been organised to allow the NSIs’ staff in the ESS and candidate countries to meet each other and exchange experience on accessing administrative data sources. Both problems and solutions were discussed among participants.

To achieve these objectives, the Workshop included:

- An opening presentation on the ESS.VIP ADMIN project and the main challenges confronted;
- Eight presentations given by experts from different NSIs, covering a wide variety of topics;
- Six parallel sessions, led by moderators, devoted to the identification of typical challenges faced in accessing and using administrative data and exchange of relevant experiences.

Conclusions of the report

The workshop included discussion sessions on six topics related to access to administrative data sources. The main conclusions are listed below.

Legal framework for statistics and general policy on data access and use

- It is important to cultivate good bilateral relations with the owners and the data protection authorities and to gradually persuade them that the risks of disclosure are minimal.
- The employment of statisticians by administrative data owners and data protection authorities helps, since they could understand the probabilistic arguments of the NSIs.
- A way to legally enforce the NSI’s access to administrative data would be to legislate that the NSI is forbidden from carrying out a survey if the desired data are already available in an administrative source.
- The NSIs should try to develop good relations with the data protection authorities in order to get their support, instead of trying to make use of the legal provisions.
- Trying to change the statistical law has the risk of ending up with a less powerful law.
- The General Data Protection Regulation\(^1\) (GDPR) recently voted by the European Parliament,

\(^1\) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.
may lead to changes in national laws. This creates opportunities for increasing and threats of reducing the NSIs’ powers, depending on the direction of the revision of the national legislation.

• In most cases the privatisation of public authorities poses a threat of getting certain data out of the NSIs’ reach.

• It is feasible to influence the administrative sources when good relations are developed and when the data owners understand the needs of the NSI. The NSIs advice is generally not taken into account by the data owners when it entails additional work/costs for them. Sometimes the NSI offers the owner some manpower to implement the requested changes.

Permanent mechanisms of cooperation with data owners

• Formal agreements (MoU) are necessary to establish cooperation between the NSIs and the administrative data owners, but most of all; common understanding is necessary so that MoUs are respected.

• Provision of data to the NSI may result in additional human and financial burden for the data owners. Additional sources of financing this work might be necessary at least for the establishment of the cooperation.

• In many cases, the NSI lacks the power to influence the content of administrative databases. Higher political pressure could be necessary to enforce this right.

Cooperation with data owners on quality issues

• Better insight into the quality of administrative data sets can be achieved by close contacts between the statisticians and the administrative data owners.

• Variables that are of secondary importance for the owners of administrative data tend to be of weaker quality than primary variables.

• The owners of administrative data should be encouraged to inform the statisticians about changes in definitions, coverage, specific (changes in) rules applied to certain sub-groups, etc.

• Linking of data from different sets of administrative data may be problematic for various reasons; the analysis of possible biases is very important.

General institutional environment and communication policies

• The public already share a lot of data about them in social media and moreover, are perhaps resigned to the fact that a lot of data are collected about them without their knowledge. This may make them less resistant to NSIs’ access to administrative sources.

• On the other hand the growing distrust of some parts of the population towards mainstream politicians and governments is also reducing the same parts’ trust towards NSIs.

• Some parts of the public may confuse official statistics with other so-called statistics that abound nowadays on the Internet and the low quality of some of the latter may reduce trust in the former too.
• Trust in the NSI’s integrity and expertise is paramount and a lot more important than any communication activity. However, communication and education with the public and other stakeholders is considered as very important too. The NSI should be involved in the public dialogue about statistics.

• The public should be educated, even from a very young age, about the role and use of official statistics.

• The messages that should be disseminated in these communication activities are the following:
  o What administrative data the NSI uses, for which purpose and what processing and analysis it does (full transparency).
  o Micro-data only come into the NSI; they do not go back out under any form or for whichever reason (building trust). In many countries NSIs are not allowed to report back to the administrative source owners errors at micro-data level (Sweden, Netherlands). This however makes difficult the provision of useful feedback to them, which would create ‘win-win’ situations.
  o The NSI performs its duties with competence and integrity (building trust).
  o Official statistics are of much higher quality than other statistics (building trust).
  o What are the benefits of administrative data use for the public, e.g. reduction of respondent burden.

NSI best practices related to internal management; financial aspects of accessing administrative data

• Administrative data exploitation within the NSI should be centrally coordinated by a separate unit dedicated to this purpose.

• Apart from the permanent staff of this unit, which should consist of IT experts and methodologists, consultation with the field experts on an ad hoc basis is considered necessary.

• Further training of the NSI’s staff, i.e. organisation of seminars or workshops for informing them about the necessity of using administrative data is considered important.

• Coordinated investments on IT developments in order to eliminate problems arising due to incompatibility of the IT systems may require interventions at a higher hierarchical level.

Cooperation with data owners on technical issues; dealing with privacy issues

• Good and up-to-date metadata are a requirement for the use of administrative data by statisticians.

• Changes in data sets must be reported timely and precisely. Lack of metadata must also be addressed as a priority issue.

• While many administrative data systems were not designed for statistical purposes, the statisticians must consistently and continuously strive for their interests to be taken into account. The same applies for the fact that some data owners do not attach enough importance to statistical requirements.
• Problems with providers of administrative data could often be solved by more intense communication between the data owners and the statisticians. Training courses and personal assistance were also useful instruments.

• Privacy considerations must be addressed systematically. Statisticians and users of microdata must be thoroughly trained in data disclosure and privacy protection issues.

Report structure

The structure of the report follows the order of the workshop sessions. It firstly presents a summary of the main points of interest from the presentations of the first day of the workshop and continues with the presentation of the key conclusions from the parallel sessions. The same approach of presentation is similarly followed for the second day of the workshop.

The workshop programme is included in Annex I of the report, while all presentations are available on the CROS portal.\(^2\)

\(^2\) https://ec.europa.eu/eurostat/cros/content/workshop-presentations-now-online-0_en
2. Workshop sessions

2.1. Plenary session – 1st day

*Moderator: Photis Stavropoulos*

The session contained four presentations about the internal organisation of four NSIs for matters related to administrative data. The representatives of the Central Statistics Office (CSO) of Ireland and of the Statistical Office of the Slovak Republic (SOSR) spoke about all administrative data treated by their organisations. Those of the Central Statistical Office (CSO) of Poland and of the Federal Statistical Office (FSO) of Germany spoke about specific domains: agricultural statistics and justice crime statistics respectively.

‘Data Governance in the Administrative Data Centre’, Richard Murphy (Central Statistics Office, Ireland)

The presentation was about the Administrative Data Centre (ADC) established by the CSO in order to liaise with and compile administrative data from the Irish Public Sector Organisations. The main concern and care of the ADC is to protect the privacy of personal information contained in the administrative data so as to build and retain the trust of their providers and the public. The main actions taken are the following:

- An Memorandum of Understanding is signed with each administrative data source owner describing in detail and in a formal way all aspects of the data provided to the ADC, e.g. variable specifications, data file structure, data file format, transmission frequency, transmission deadlines, etc.
- A High Level Liaison Group is set up with each administrative data source owner and is convened at least twice per year to discuss the cooperation with the CSO.
- The CSO has statisticians in many different public sector organisations in order to compile and disseminate official statistics in these organisations, undertake formal communications between the CSO and Statistical Units in the public service, monitor the development and implementation of their data strategies, provide advice and promote adherence to international principles.
- All members of staff of the CSO have to submit a separate Privacy Impact Assessment (PIA), signed by the sanctioning Data Office and a Management Board member for each individual sensitive dataset they need access to. The PIA has to be renewed every six months and whenever the member of staff moves sections in the CSO.

‘Best Practices in Establishing Permanent Cooperation with Administrative Data Owners in Slovakia’, Peter Mižík (Statistical Office, Slovak Republic)

The SOSR has established a specific department, the Department of Administrative Data Sources, in order to liaise with and compile administrative data from the public authorities of the Slovak Republic. The SOSR has the right to access data from any administrative source and use them for any statistical purpose with the exception of the data of the Ministry of Health, which is provided only in aggregated form. The department:
Workshop sessions

- has in place a standardised set of procedures for identifying administrative data sources, interacting with their owners, documenting and using the data;
- despite its legal powers, puts emphasis on its informal relations with source owners;
- tries to give things in return to the source owners, e.g. data from its statistical registers so that they improve their registers or aggregated data useful for their operations;
- tries to be involved in the design of new administrative data sources.

The SOSR tried to insert additional clauses in its favour in the recent revision of the national statistical law, e.g. the obligation of administrative data source owners to inform the SOSR one year before discontinuing any source. It was not successful in this respect. The revised law obliges the owners to inform before discontinuing their sources, but not as early as one year in advance.

The attention of the participants was mainly drawn to what the SOSR gives back to the owners, since the creation of ‘win-win’ situations is of great interest. Another question was about the extent to which the data source owners follow the SOSR’s recommendations about improvements in their data. The presenter underlined the importance of being in close contact with data owners at the stage when they start to prepare the changes.

It should also be pointed out that the SOSR’s work on administrative data was partly funded by a Eurostat grant in the framework of the ESS.VIP ADMIN project.

‘Practice in use of administrative data in agricultural statistics in Poland’, Artur Łączyński (Central Statistics Office, Poland)

The CSO has a specialised unit which is responsible for treating all administrative data received by it. The presentation was about the production of agricultural statistics which relies on a variety of administrative data sources.

A recent change in the statistical law explicitly mentions the administrative sources to be used by the CSO. This was considered necessary because the legal acts it is meant to amend were not considered sufficiently important by the administrative source owners. The CSO also tries to initiate revisions, to its favour, of the legal acts regulating specific administrative data sources.

In the discussions, the importance of having a proper legal basis was underlined. It was also mentioned that the NSI should screen the legal basis of the data owners and try to change that.

‘Permanent coordination structures in decentralised administrations’, Thomas Baumann (Federal Statistical Office, Germany)

The presentation by the FSO was about statistics on criminal prosecution agencies and civil and criminal courts. They differ from the previous cases in that administrative data are not a new and alternative data source; they are in place already for a long time and are the only available data source regarding the total number of criminal prosecution proceedings and civil and criminal courts proceedings.

The German case is also particular in that the German court, police and statistical systems are decentralised, with authority devolved to the 16 Länder. In addition, statistical reporting on police is not a task of the Statistical Offices but of the Criminal Police Offices at the 16 Länder and at federal level. So, the Länder’s statistical offices get administrative data (micro-data) from the Länder’s
criminal prosecution agencies, civil and criminal courts. The Länder's statistical offices then produce statistical tables for their Länd and provide anonymised data at Länd-level to the Federal Statistical Office. The latter produces totals at federal level based on data at Länd level.

The attention of the participants was drawn to how the FSO manages to coordinate the data collection in all Länder within the system of statistical offices in Germany as a part of the European Statistical System. This is achieved and facilitated by the following actions and conditions:

- Coordination between statistical offices at national level: Once a year the Federal Statistical Office invites the Länder statistical offices to discuss and decide on statistical and technical aspects of compiling different statistics on justice at subnational and national level. The Federal Ministry of Justice attends this meeting as well and informs on new political issues.

- Coordination between statistical offices at EU level: Once a year, Eurostat invites the statistical offices of the member states to discuss and decide on statistical and technical aspects of compiling statistics on crime and criminal justice at national and EU level. DG Home and DG Justice attend this meeting as well and inform on new political issues.

- Coordination between ministries and statistical offices: Once a year the Länder Ministries of Justice meet to update data requirements in the set of statistics on justice. The Federal Ministry of Justice, the Federal Statistical Office and the Länd Statistical Offices attend this meeting as well. This meeting discusses and decides on harmonised data collection as a main task of the ministries of the Länder and on harmonised data preparation and validation as a main task of the statistical offices.

- Though there are different types of statistics on justice (criminal courts, civil courts, labour courts etc.), for each type of justice statistics the same data collection form used in all Länder. This unique form is implemented in the software tools used by the courts to carry out their everyday business. Within these tools there is a special tool to transmit statistical data to the Länd statistical offices.

- It is also fortunate that the same laws apply in all Länder, rendering the administrative data comparable between Länder.

2.2. Parallel Session 1 - Legal framework for statistics and general policy on data access and use

*Moderator: Photis Stavropoulos*

2.2.1. Objectives of the session

This session covered the legal framework in which the NSIs function, including the rights given to the NSIs by the national statistical laws, the issues created by potential conflicting interpretations of the legal provision on the NSIs rights, as well as the difficulties created by the legislation governing the owners of the administrative data sources and the data protection legislation.
2.2.2. Summary of the main issues of discussed

12 persons participated in the session. The discussion was quite lively and led to the following conclusions:

The powers granted by law to NSIs are quite extended but the latter do not always enforce them. They prefer to maintain good relations with the administrative data source owners rather than appearing as too forceful.

- In Malta the statistical law is very powerful but the NSI does not make full use of it. The NSI has the right to ask for changes in administrative data sources but does not use this power.

- The situation is similar in Ireland, where although the law gives to the NSI the right ‘to demand’ access to administrative data, the NSI prefers to ask for them without invoking their legal rights.

- In the UK the statistical law is under revision. The proposed new version would give the NSI much stronger rights.

The legislation protecting the privacy of personal or business data is the main real or invoked legal obstacle to the NSIs’ efforts to access administrative data. Administrative data source owners or data protection authorities raise concerns about data privacy. These concerns come often from not understanding the NSIs’ argumentation, based on probability theory and statistics, that the probability of disclosure is minimal. An example came from Malta, where the tax authority stopped providing data to the NSI when it was made public that it does so, due to potential concerns that the NSI may contribute to identifying the tax evasion. Netherlands on the other hand is free of such obstacles because its statistical law forbids other authorities to refuse data provision on claims of protecting data privacy.

Solutions discussed were:

- Cultivation of good bilateral relations with the owners and the data protection authorities and gradually persuading them that the risks are minimal.

- The employment of statisticians by administrative data owners and data protection authorities since they could understand the probabilistic arguments of the NSIs.

- A suggestion for another way to legally enforce the NSIs’ access to administrative data would be to legislate that the NSI is forbidden from carrying out a survey if the desired data are already available in an administrative source.

- It was pointed out that the NSIs should try to develop good relations with the data protection authorities in order to get their support, instead of trying to make use of the legal provisions.

- Trying to change the statistical law has the risk of ending up with a less powerful law.

In a similar vein, it was pointed out that the General Data Protection Regulation (GDPR) recently voted by the European Parliament, may lead to changes in national laws. This creates opportunities for increasing and threats of reducing the NSIs’ powers, depending on the direction of the revision of the national legislation. The NSIs could be actively involved in the revision of the legislation at national level in order to seize the opportunities. Another point of discussion concerned the right to access data held by private owners; in many countries this is a grey area subject to interpretation. In Malta, the law does not specify clearly whether private owners of administrative
data are also obliged to provide them to the NSI. A private telephone company is refusing to provide them and challenges the law on these grounds. In Ireland, the data protection authority also stopped the NSI’s effort to obtain data from mobile phone operators. On the other hand, in Bulgaria the law does not make any distinction between private and public sources. It was concluded that in most cases the privatisation of public authorities poses a threat of getting certain data out of the NSIs’ reach.

As regards the right of the NSIs to influence the administrative source, the general feeling was that this is feasible when good relations are developed and when the data owners understand the needs of the NSI. The NSI advice is generally not taken into account by the data owners when it entails additional work/costs for them. Sometimes the NSI offers the owner some manpower to implement the requested changes.

**2.3. Parallel Session 2 - Permanent mechanisms of cooperation with data owners**

*Moderator: Anna Ampatzoglou*

**2.3.1. Objectives of the session**

This session dealt with the formal cooperation mechanisms with the data owners, for example the creation of permanent coordination bodies, signing of memoranda of understanding, establishing single contact points, the involvement of a third party (such as an authority for the protection of personal data) in the agreements between the NSI and data owners, etc. The purpose was to identify cooperation procedures and mechanisms that work well and ways to overcome possible problems.

**2.3.2. Summary of the main issues of discussed**

Participants were invited to discuss about the cooperation mechanisms established with the administrative data owners. Some participants presented the problems confronted, while others briefly presented the cooperation mechanisms established in their country.

In Statistics Netherlands, the cooperation between data owners and the NSI is not always effective, since the administrative data owners do not always provide the requested data. The main reason for this is the lack of a common understanding of the NSI’s needs. An example mentioned was the cooperation of the NSI with the Tax Authorities.

To overcome such issues, in the UK, employees of ministries, working on a secondment basis in the NSIs act as mediators between NSIs and ministries (data owners).

In Latvia, domain experts are assigned to discussion with the administrative data owners for each and every specific issue that may arise.
Confidentiality issues have been also been raised by the participants emphasizing the fact that in some cases, access to microdata is limited due to overlapping or contradicting regulations.

In Poland, the provision of data to the NSI results in additional burden for the data owner in terms of human and financial resources. This creates reluctance of data owners to cooperate.

In this respect, most participants agreed that alternative sources for financing this work - at least at the stage of the establishment of the procedure of cooperation - have to be investigated.

Another issue raised was that NSIs lack the power to influence structure of administrative databases. The role of the NSI on this issue is advisory. It was suggested that NSIs insist constantly so that this situation is reversed. In some cases higher political pressure could be necessary.

It was stated that communication with data owners based on a single contact person is at risk to stop functioning when the relevant person moves to another position.

Most of the participants agreed that more “formal” means are required for establishing cooperation with administrative data owners.

For example, in Montenegro, the department responsible for the coordination of activities with data owners organises several meetings with them at various hierarchical levels. Meetings are generally formal and aim to establish a close collaboration with the data owners. So far, certain MoUs have been signed with some administrative data owners.

In Spain, cooperation with data owners is – in most cases – coordinated by the Department of Methodology of the NSI. For main providers (such as Social Services and tax offices, and some institutions of the autonomous communities) collaboration agreements have been signed with the INE. However, sometimes there are also direct contacts between domain experts and data owners and some production units still keep their own collaboration mechanisms with administrative data owners. These are to be integrated gradually in the indicated common coordinated system. The main issue preventing the NSI from a more extended and effective use of administrative data is the lack of resources and relevant IT infrastructure.

In Estonia, the legal framework is considered strong. Administrative data are received from the data owners, but the lack of financial resources constitutes sometimes an obstacle (e.g. data are stored in paper archives and the relevant authority requests additional resources for the data entry procedure).

Through the whole discussion a common consensus has been reached: MoUs although signed are not always respected. For example, in Slovakia, although MoUs have been signed with data owners, these are not considered adequate. In this respect, any formal agreements are usually accompanied with informal bilateral ones between experts on the field.

**The discussion led to the following key points:**

- Formal agreements (MoU) are necessary to establish cooperation between the NSIs and the administrative data owners, but most of all; common understanding is necessary so that MoUs are respected.
- Provision of data to the NSI may result to additional human and financial burden for the data owners. In this respect, additional sources of financing this work might be necessary at least for the establishment of the cooperation.
• In many cases, the NSI lacks the power to influence the content of administrative databases. Higher political pressure could be necessary to enforce this right.

2.4. Parallel Session 3 – Cooperation with data owners on quality issues

Moderator: Willem de Vries

2.4.1. Objectives of the session

The session was devoted to the quality of administrative data and possible ways to improve quality by cooperating with data owners, for example by introducing common validation procedures but also by the involvement of statistical authorities in the design and changes in administrative sources. The purpose of the session was to see how some of these issues could be solved by improved cooperation.

2.4.2. Summary of the main issues of discussed

It was mentioned that in Hungary the statistical office often did not have sufficient information about the quality of administrative data. One persistent problem is the lack of reports on the quality of data.

It was agreed that in general the variables which are of lesser importance for the owners of administrative data are weaker in quality than the core variables.

In Croatia the plan for the next Census is to conduct a traditional Census using electronic questionnaires and possibly use some administrative registers. In Croatia it is sometimes not entirely clear why certain variables are weak or missing. By comparing information from different administrative sources it was sometimes possible to derive the correct value. As an example of a problematic variable, 'disability' was mentioned (as earlier explained in the presentation by Croatia to the plenary session).

Estonia strives for a register-based Population Census. However, it seemed to be very difficult to get valid information about the variable 'occupation'.

A problem in Belgium is metadata changes, i.e. changes in records happened without the NSI being informed. It was also noted that coordination of the statistical system of Belgium was particularly difficult because some statistics and administrative registrations are a federal responsibility, while others are the responsibility of the (language-based) communities and (territorial) regions.

Italy reported that the administrative datasets acquired are generally of a high quality in terms of their statistical usability. With regard to the relationship with administrative data holders, a ubiquitous problem is timeliness. A generalized quality assessment system that uses the framework defined in the international project Blue-Enterprise and Trade Statistics (ETS) is implemented.

In Greece, ELSTAT received access to administrative data for Structural Business Statistics (SBS) 4 months ago. Errors were detected by means of various forms of statistical analysis. It was noted that the administrative data had a bias in the sense that companies had higher turnover in administrative registers than in statistical surveys. Linking of different data sources appeared to be
a problem. Another issue was that VAT was only registered for companies with a turnover of 1 Million Euro and more.

In Germany, wine statistics are entirely based on registers and this is very satisfactory. Every wine producer must be registered in a register of vineyard owners. Some over-coverage occurs when owners own vineyards in more than one Land (double ownership). Registers are also used for land-use statistics. According to the law, these registers should take statistical interests into account, but this is sometimes problematic. Remote sensing information is used as a complementary source.

In Latvia, the idea is to base the Population Census on administrative data. However, the variable ‘educational attainment’ is problematic. Another problem is that the time reference of different administrative data sets may vary.

For international trade statistics of Germany, some problems were mentioned concerning the use of the V.I.E.S. (VAT Information Exchange System). Data on recapitulative statements collected in each EU country regarding exports to other EU countries are shared through the V.I.E.S. in order to control for the respective imports (mirror data). V.I.E.S. could serve as an instrument for controlling the partner country level data for imports and exports within Intrastat (trade with EU countries). However, there is no possibility of using it as missing months and various inconsistencies were identified on the imports’ side. It had been tried to solve this problem by consultation with the respective Tax authority, but the results were so far very modest. For crafts statistics, one problem was that VAT statements were sometimes made for groups of enterprises. So VAT information was not available for all enterprises at individual level. Another issue was that NACE codes were not supplied by the data owner. It is difficult to address these issues with the owners of administrative data.

The representative from South Africa briefly explained that the NSI is trying to develop a geo-reference system (GIS) for household surveys. One problem was that identical names of different locations often occurred.

It was suggested that win-win situations might be created by explaining to administrative data owners that it was in their own interest as well to cooperate with the statisticians. In Croatia, this had been tried repeatedly. An example in the Netherlands was mentioned, where the CBS had produced brochures for starting businesses, in which basic statistics for certain industries were presented. These brochures were distributed by the Chambers of Commerce, which is the most important basis for the CBS statistical business register and where all businesses must register.

The discussion led to the following key messages:

- Better insight into the quality of administrative data sets can be achieved by close contacts between the statisticians and the administrative data owners.
- Variables that are of secondary importance for the owners of administrative data tend to be of weaker quality than primary variables.
- The owners of administrative data should be encouraged to inform the statisticians about changes in definitions, coverage, specific (changes in) rules applied to certain sub-groups etc.
- Linking of data from different sets of administrative data may be problematic for various reasons; analysis of possible biases is very important.
• Differences in time reference may create problems that are often difficult to solve.

• It is important to convince administrative data owners that there may be benefits in working closely with the statisticians; various incentives may be offered, e.g. training in classification etc.

2.5. Plenary session – 2nd day

Moderator: Willem de Vries

The plenary session of the 2nd day contained four presentations about the exploitation of administrative data. The presentations given by the representatives of Statistics Finland and the National Institute of Statistics of Italy spoke about all administrative data treated by their organisations, while those of the Central Bureau of Statistics of Croatia and the Federal Statistical Office (FSO) of Germany focused on specific domains, namely the population census and the craft statistics, respectively.

‘Administrative data collection at Statistics Finland’, Nico Maunula (Statistics Finland)

Statistics Finland’s Havas project started in April 2013 and ended in March 2015. The main goals of the project were:

• To find and record all the administrative data that are collected and used at Statistics Finland

• To define and build a centralised IT system for receiving administrative data by line transfer

• To define and build a SAS system for pre-controlling the data received

• To describe all the data and variables in a centralised metadata system

• To organise a team to whom the centralised administrative data collecting was handed over after the project

Statistics Finland uses 150 different registers of administrative data to collect and compile statistics. Now a system is in place that utilises process metadata to control and inspect the received administrative data.

All data are initially received on a Transfer Data Server, which is outside Statistics Finland’s own data network. The group responsible for this data collection system handles 65% of all administrative data files suitable for the centralised system. The remaining 35% of data files are collected by other means (e.g. encrypted emails).


In the 2001 and 2011 Population Censuses, measuring the incidence of disability was based on the statements of respondents. In the government Registry of disabilities, on the other hand, registration of disabilities is based on medical records. In addition, the definition of disability differed between the 2001 and the 2011 Census.

While the 2001 Census questionnaire contained a general question on disabilities of any kind, the
2011 questionnaire specifically asked for difficulties in performing activities of daily life, due to long-term illness, disability or old age. The answer to this question was given independently from the certificate issued by the competent government institutions confirming a person’s difficulties (disability, impairment and handicap). These forms of disability are submitted to the Registry electronically by e-card or in writing on the prescribed form, by the 10th of each month for the previous month. Electronic records or copies of other necessary documents are submitted to the Registry no later than 30 days from the date of their issuance.

The Registry of persons with disabilities is one of the best registers in Croatia in which the 2001 Act on the Croatian Register of Persons with Disabilities regulates the method of collecting data about the cause, type, degree and severity of health impairment, the rights of persons with disabilities, methods of processing and use of data, as well as the data protection. It covers all persons with disabilities in the Republic of Croatia, with geographical detail to the level of settlement.

Register data are updated on a daily basis which allows current monitoring of the number of disabled persons, the type and extent of damage as well as their territorial distribution of the persons. Because of all mentioned above, the CBS for the next Census intends to consider the use of the Registry of persons with disabilities to obtain data about the disabled persons and their characteristics, omit the questions on disability from the Census Questionnaire, and thus shorten the time of data collection, reduce workload of the respondents and simplify the enumeration.

It should also be pointed out that the CBS’s improvements on the use of administrative data were funded by a Eurostat grant in the framework of the ESS.VIP ADMIN project.

‘New data collection organization in the Italian Statistical Institute and the IT tools to collect and manage administrative data’, Grazia Di Bella (National Institute of Statistics, Italy)

In recent years in ISTAT there has been a sharp increase in administrative data acquisition, further enhanced by the ISTAT modernization project. This is centred on a Registers System and a reorganisation, during which a Directorate was created, responsible for data collection, both primary data (survey) and secondary data (data produced for other purposes, in particular administrative data). This new Data Collection Directorate is also in charge of data integration with the aim of supporting statistics production from Registers.

In 2016 more than 300 acquisitions are planned, involving 150 separate administrative datasets and 40 data holders - more than a terabyte of data mainly containing personal data. Therefore it was necessary to design IT processes for the acquisition of administrative data. The AD Acquisition System, ARCAM, has been created with this objective. It uses different technologies and ISTAT standards and, through a special web site, it allows to acquire data in accordance with current legislation on data confidentiality and to preserve data integrity. This web interface can be a useful tool for communication with the administrative data suppliers in order to share the commitment to the use of data for statistical purposes and to improve usability/quality levels.

The acquired data are stored in a single repository and are made available to internal users in a timely manner through the ARCAM management of permissions for data access.

Administrative datasets that include microdata enter into the standardized process of the System of Integrated Microdata (SIM) and each administrative object, identified as a statistical unit (individuals, economic units and places), is integrated. The data integration is performed through record
linkage procedures ending with the assignment of unique codes constant over time. Access to integrated data for internal users is also regulated in compliance with the law on data protection.

‘Using Machine Learning Algorithms to reduce clerical effort and improve the quality of administrative data sources’, Jörg Feuerhake (Federal Statistical Office, Germany)

In Germany statistics about the enterprises of the crafts sector are a separate part of the programme of official statistics. To define the statistical population for the crafts statistics, Germany’s Chambers of Crafts provide their membership lists to the statistical offices on a yearly basis. These lists are an important administrative source for the crafts statistics.

The statistical offices link the lists to the business register. However, not every unit in the lists, which can be linked to a business register unit, is relevant for the crafts statistics. Hence identifying irrelevant units is essential to maintain a meaningful statistical population for the crafts statistics. The chambers of crafts cannot identify the relevant cases beforehand because they do not have the necessary information. Consequently the statistical offices have to invest a lot of clerical effort to identify the irrelevant units.

Using correctly classified data of previous years as ‘training data’, Support Vector Machines make it possible to define the statistical population for the crafts statistics with greatly reduced clerical intervention. A Support Vector Machine (SVM) is a supervised machine learning algorithm that can be employed for both classification and regression purposes. SVMs are more commonly used in classification problems.

2.6. Parallel Session 4 – General institutional environment and communication policies

Moderator: Photis Stavropoulos

2.6.1. Objectives of the session

The purpose of this session was to discuss the general institutional, political and administrative environment in which the statistical institutes function. This covers established or emerging national initiatives, policies or strategies for open government data, data sharing in the public administration, interoperability of the public administration information systems and other similar issues and the way that the NSIs can influence and exploit them. The discussions also aimed to cover public opinion on the access and reuse of the data by the NSI and the communication and promotion strategies that the NSIs have to create public trust. Difficulties created by public perception and ways to overcome them, as well as the exploitation of the opportunities created by broader political initiatives on data were also in the scope of this session.

2.6.2. Summary of the main issues of discussed

The session had two topics: (a) the public’s concerns about data privacy, (b) the existence of government-wide policies, strategies or initiatives for data sharing between public authorities, and the impact of both on the NSIs’ ability to access administrative data.
The session had 11 participants. Most of the discussion concentrated on the first topic.

(a) Public concerns about data privacy.

In the Netherlands, Poland, Germany, Sweden and Spain the public is positive or at least indifferent to the use of administrative data about them by the NSI. The public is positive because it wants to ‘provide data only once’. There are cases when the Spanish NSI learns about potential administrative data sources from respondents’ responses in surveys of the following kind: “Why do you ask for this information? We already provide it in ...”. Sweden mentioned the case of Norway which has a central hub where all administrative data of public authorities are deposited and which forbids any public authority to collect directly from citizens, households or enterprises data that are already available in the hub.

Moreover, concerning Poland and the UK it was mentioned that usually is not the public but certain politicians or journalists that stir public opinion against the sharing of administrative data between authorities.

The participants also agreed on the following points:

• The public already share a lot of data about them in social media and moreover, are perhaps resigned to the fact that a lot of data are collected about them without their knowledge. This may make them less resistant to NSIs’ access to administrative sources. The Netherlands in fact mentioned that the NSI is trusted by the public more than, for example, the tax authority.

• On the other hand the growing distrust of some parts of the population towards mainstream politicians and governments is also reducing the same parts’ trust towards NSIs.

• Some parts of the public may confuse official statistics with other so-called statistics that abound nowadays in the Internet and the low quality of some of the latter may reduce trust in the former too.

• Trust in the NSIs’ integrity and expertise is paramount and a lot more important than any communication activity.

However, communication and education with the public and other stakeholders is considered as very important too. The following recommendations were put forward about actions:

• The NSI should be involved in the public dialogue about statistics. The UK’s Office for National Statistics has created a special media team in order to communicate what the ONS uses administrative data for. Moreover, during the two-year public consultation of a proposed new statistical law that would give it stronger rights to access administrative data the ONS was extensively involved in dialogue with politicians, data privacy lobbies, business associations, citizen groups, etc.

• The public should be educated, even from a very young age, about the role and use of official statistics.

The messages that should be disseminated in these communication activities are the following:

• What administrative data the NSI uses, for which purpose and what processing and analysis it does (full transparency).
• Micro-data only come into the NSI; they do not go back out under any form or for whichever reason (building trust). In many countries NSIs are not allowed to report back to the administrative source owners errors at micro-data level (Sweden, Netherlands). This however makes difficult the provision of useful feedback to them, which would create ‘win-win’ situations.

• The NSI performs its duties with competence and integrity (building trust).

• Official statistics are of much higher quality than other statistics (building trust).

• What are the benefits of administrative data use for the public, e.g. reduction of respondent burden.

(b) Government-wide policies, strategies or initiatives.

No specific such policies were mentioned, with the exception of Poland where, there are ongoing efforts for interoperability of government databases in technical terms. It is not clear yet which government authorities will be allowed to provide to the CSO access to their data.

There are also policies for publishing government data in open formats (e.g. in the UK and France). This however does not mean that all data should be shared between authorities.

Spain mentioned that the autonomy granted to the country’s regions complicates coordination between regions and access to their administrative data.

2.7. Parallel Session 5 – NSI best practices related to internal management; financial aspects of accessing administrative data

*Moderator: Anna Ampatzoglou*

2.7.1. Objectives of the session

This discussion session was devoted on issues such as: creation of an administrative data centre, coordinated policies between internal NSI departments to identify new administrative sources and promote their exploitation, organisational and management procedures for the acquisition and treatment of administrative data, training, support and information activities in NSIs, NSI strategy to build relevant staff skills. The session also covered financial aspects related to the exploitation of administrative sources, like payment arrangements with the administrative data owners for data collection / processing / transmission, coordinated investment in IT between the NSI and the data owners etc. The purpose of the session was to identify the challenges linked to management, staff and financing activities and to identify ways to overcome these difficulties.

2.7.2. Summary of the main issues of discussed

In the UK, administrative data collection is coordinated centrally in the NSI. A dedicated unit deals with all issues regarding administrative data exploitation. Plans for the future include the organization and storing of all data centrally, regardless of the source they stem from (administrative or survey data).

Additionally, participants discussed which type of staff should work in such a unit within the NSI.
It was commonly agreed that IT experts are necessary. Methodologists must also be part of such a unit since they are the ones that can assess the quality of data received. The necessity of the participation, in an ad hoc basis, of domain experts has also been raised. Such unit exists for example in Slovakia, although with limited resources.

In Italy, survey and administrative data collection is centralised. A specific unit responsible for the administrative data management coordinates data needs, the data collection process and the integration procedures. Domain experts do not belong to this centralised unit but are coordinated centrally. The IT services are provided by the IT Directorate. There is an emerging need for training of the staff on the procedures related to the use of administrative data. In this respect, it was suggested that handbooks dedicated to such activities are necessary.

Also, it was highlighted that the staff needs to be persuaded that use of administrative data will not hamper their work. Seminars or workshops are an important means for the exchange of opinions, possibly also with staff from other NSIs, so as to promote the usage of the administrative data.

In Malta, administrative data coordination is decentralised for the moment. However, plans for a more centralised way of data collection are in place as this is considered to be a more efficient practice.

Legal frameworks in all countries grant free access to administrative data but in some cases, like in Malta, further data processing takes place by the data owner and in this respect the financial support from the NSI is required.

Finally, as discussed, coordinated IT investments require interventions at a higher hierarchical level. Such investments could have impact on the whole IT organisation of the public sector of the country.

The discussion led to the following conclusions:

- Administrative data exploitation within the NSI should be centrally coordinated by a separate unit dedicated to this purpose.
- Apart from the permanent staff of this unit, which should consist of IT experts and methodologists, consultation with the field experts on an ad hoc basis is considered necessary.
- Further training of the NSI’s staff, i.e. organisation of seminars or workshops for informing them about the necessity of using administrative data is considered important.
- Coordinated investments on IT developments in order to eliminate problems arising due to incompatibility of the IT systems may require interventions at a higher hierarchical level.

2.8. Parallel Session 6 – Cooperation with data owners on technical issues; dealing with privacy issues

*Moderator: Willem de Vries*

2.8.1. Objectives of the session

This aim of this session was to address typical technical difficulties encountered concerning the
practical access to administrative data sources: problems due, for example, to incompatible data formats or incompatibilities of the respective information systems. The discussions also concerned the additional challenges raised by the need to protect data: different data flows and technical solutions that have to be adopted in order to ensure data protection. The purpose of the session was to discuss difficulties created by the technical aspects and the need to protect data and identify solutions to overcome these difficulties.

2.8.2. Summary of the main issues of discussed

It was pointed out that incompleteness of administrative data sets and differences (e.g. in the definition of variables etc.) sometimes present great problems. It was also mentioned that the ESSnet Admin Data made an important contribution in this respect. Another issue concerns the International Accounting Standards which were sometimes changed, without regard for statistical interests.

Germany pointed out that there are technical problems due to the different way Länder handled data. A solution that had been applied was direct contact between the Statistical Federal Office and specific data holders.

Romania remarked that administrative data sources were often not designed to meet statistical requirements as well. In addition, some data owners are reluctant to give access to their data, e.g. because of privacy considerations. In general, there are fewer problems with data about businesses than with data about persons and households.

In Ireland, the CSO has – by law – the right of access to all administrative data, with the exception of health and criminal justice data. However, the owners of administrative data are not always sufficiently cooperative. One of the reasons was that they gave priority to their core function over supplying data for statistics (their job is to collect taxes, not to produce information for statistics). Consistent reach-out was necessary to obtain cooperation.

Germany mentioned some technical problems with data for the Business Register, obtained from Finance and Labour authorities.

Estonia is moving to a system of machine-to-machine interchange of data, on the basis of the XML format. The target is to achieve this by 2020. Data owners can apply for government funds to change their systems.

In the Czech Republic the aim is to move towards a ‘virtual census’. However, there are many obstacles to overcome: lack of metadata, difficult-to-use data formats etc. While there is a population register, this seems to be yet unsuitable for statistical purposes.

It was remarked that problems with providers of administrative data could often be solved by more intense communication between the data owners and the statisticians. Training courses and personal assistance were also useful instruments.

In Latvia, improvements were reached by agreements between data owners and statisticians, e.g. about the way quality issues were handled.

In Kosovo, the NSI does not yet have access to administrative microdata. The representative also mentioned problems with the Ministry of Housing, the Tax Authorities (lack of timeliness) and in general with incomplete data.
France has some problems with data obtained from the Ministry of Finance, regarding information for updating the business register. Corrections made by INSEE were fed back to the Ministry.

About privacy issues, it was mentioned that this was a major concern for many NSIs. E.g. in Ireland very strict rules applied to the Administrative Data Centre. Staff members had to follow a course in data protection. In Belgium there was a separation between data collection units and subject matter departments. The dissemination department had no access to microdata. In France, everyone dealing with (administrative) microdata must sign a confidentiality declaration. A Secrets Committee oversees the protection of confidentiality and privacy.

**The key findings to address are summarised below:**

- Good and up-to-date metadata are a requirement for the use of administrative data by statisticians. Changes in data sets must be reported timely and precisely. Lack of metadata must also be addressed as a priority issue.

- While many administrative data systems were not designed for statistical purposes, the statisticians must consistently and continuously strive for their interests to be taken into account. The same applies for the fact that some data owners do not attach enough importance to statistical requirements.

- Problems with providers of administrative data could often be solved by more intense communication between the data owners and the statisticians. Training courses and personal assistance were also useful instruments.

- Privacy considerations must be addressed systematically. Statisticians and users of microdata must be thoroughly trained in data disclosure and privacy protection issues.
3. Conclusions of the workshop

It was concluded that in the last few years the statistical authorities have made important progress in the use of administrative data sources. The Workshop was a good opportunity to exchange best practices in order to further improve access to administrative data. The conclusion of the discussions sessions are detailed in the corresponding sections of the report.

Eurostat reminded the NSIs that the ESS.VIP ADMIN offers grants for improving the use of administrative data for statistical production.
ANNEX I. WORKSHOP PROGRAMME

Programme of the Workshop on access to administrative data sources
Brussels 13-14 September 2016

<table>
<thead>
<tr>
<th>Day 1: 13 September</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>12:00 - 14:00</strong></td>
</tr>
<tr>
<td><strong>14:00 - 14:30</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>14:30 - 15:50</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>15:50 - 16:20</strong></td>
</tr>
<tr>
<td><strong>16:20 - 17:30</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>17:30 - 18:00</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Day 2: 14 September

9:00 - 10:20
Presentations:
- ‘Administrative data collection at Statistics Finland’, Nico Maunula (Statistics Finland)
- ‘New data collection organization in the Italian Statistical Institute and the IT tools to collect and manage administrative data’, Marina Venturi, Grazia Di Bella (National Institute of Statistics, Italy)
- ‘Using Machine Learning Algorithms to reduce clerical effort and improve the quality of administrative data sources’, Jörg Feuerhake (Federal Statistical Office, Germany)

Chair: Willem De Vries (Agilis SA)

10:20 - 10:50
Coffee break

10:50 - 12:00 Parallel sessions (Group discussions)

<table>
<thead>
<tr>
<th>Parallel session 4</th>
<th>Parallel session 5</th>
<th>Parallel session 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>General institutional environment and communication policies</td>
<td>NSI best practices related to internal management; financial aspects of accessing administrative data</td>
<td>Cooperation with data owners on technical issues; dealing with privacy issues</td>
</tr>
</tbody>
</table>

Moderator: Photis Stavropoulos (Agilis SA)

Moderator: Anna Ampatzoglou (Agilis SA)

Moderator: Willem De Vries (Agilis SA)

12:00 - 13:00
Presentation of the conclusions of the parallel sessions
Conclusions of the workshop and closing

Chair: Adam Wronski (Eurostat)