European Platform Undeclared Work
Follow-up Visit

Data Mining for More Efficient Enforcement

27-28 September 2017, Brussels, Belgium

Summary Report

1. Introduction

This report summarises the information shared by the Belgian Federal Public Service Social Security with members of the European Platform Tackling Undeclared Work during a visit to their premises on 27 and 28 September 2017.

The visit was related to a Thematic Review Workshop on 'Data Mining for Efficient Enforcement' which took place in Helsinki on 1-2 June 2017. The workshop was hosted by the Ministry of Economic Affairs and Employment and brought together Platform members and observers from 11 EU Member States and Norway – including representatives of national ministries, labour inspectorates and tax and customs authorities. During the workshop, the participants shared working experiences and good practices on the use of data collection, data sharing, data matching and data mining to prevent and tackle undeclared work (UDW).

During the Helsinki workshop, participants were particularly interested in Belgian data mining experience, and requested a follow-up visit to discuss existing practice more in-depth. Belgium was selected through this process and agreed to host the visit. Representatives from seven European countries (Greece, Latvia, Lithuania, Spain, Slovenia, Finland and Norway) took part in this follow-up visit.

2. Background

There are four, intrinsically linked steps in the process of data exploitation, helping to improve efforts to tackle undeclared work, as follows:

- **Data collection** is the process of gathering data from internal and external sources.
- **Data sharing** is the process of making data available to other users.
- **Data matching** is the large scale comparison of records or files collected or held for different purposes, with a view to identifying matters of interest. With data matching, two or more sets of collected data are compared (comparison of records).
- **Data mining** can be defined as a set of automated techniques used to extract buried or previously unknown pieces of information from large databases. By the use of data mining, correlations or patterns among dozens of fields in large relational databases will be identified.

Data mining is a useful tool because the use of a data analysis system makes it possible to:

- Provide inspectors with an initial and immediate level of information on enterprises, employers and employees to enable them to verify this kind of information during an inspection;
- Guide inspection activity towards those actors for whom inconsistencies and/or high risk(s) for infringements are discovered through data analysis. At the same time, this data can be used to support the inspectors in their work;
- Estimate the size of undeclared work;

- Help policy makers to act on the results of data mining, and to raise awareness with the public of the wins in combatting undeclared;

- Increase the perceived risk of detection.

The follow-up visit took up the themes brought to light during the Thematic Review Workshop on Data Mining, which included that:

- Effective data sharing, matching and mining requires **political will and trust** between the different parties involved.

- The focus needs to be on setting up **effective national systems** before then progressing to more ambitious cross-border cooperation. The wealth of data available within an own organisation and freely available on the internet can be surprising so it is important to first find ways to make the most of what is already available.

- There appears to be little information throughout the EU on the **critical assessment of data mining** initiatives but Belgium showed how closely they monitor the performance and effectiveness of the tools being used, and to make adaptations as and when necessary.

- Data mining does **not replace the need for inspectors** but it can help target resources and lead to efficiency gains. Involving inspectors at all stages of the data gathering and mining process is important to gain their trust and to ensure that systems are accessible to them.

- It is important for the tools being used to be **user-friendly** and to present the data in an accessible and visually appealing way making it easy for inspectors to access and use.

On the following sections, Belgium's MiningWatch tool (which was a key focus of the follow-up visit) and systems (the Crossroads Bank for Social Security) as well as international initiatives (Benelux Cooperation) are summarised based on the discussions that took place during the follow-up visit.

3. Belgium's shared practices and systems

3.1 MiningWatch tool on data analytics for targeted inspections of social security fraud

MiningWatch is a data mining tool which uses predictive modelling to define fraud risks in three different sectors: construction, cleaning, and the hotel and catering sector. The analytical tool supports inspectors to choose and target their inspections based on the predictive risk modelling of fraud (including undeclared work, abuse of part-time working schemes, and bogus self-employment). The tool raises alarms for suspicion of UDW. Key questions related to this raising of alarms include:

- Whether the inspectors understand the alarm and whether the alarms are based on statistics, on previous results and how precise they are. A key input in the process is the involvement of experts in testing the alarms and validating them as it helps make MiningWatch more accessible to inspectors.

- Then, there are management aspects of acting on the alarms. These concern monitoring how many alarm-based investigations actually take place and whether they result in identifying infringements or not. Inspectors have some freedom to choose investigations based on their own initiative and in 2015, 25% of free initiative investigations resulted from the use of Mining Watch by inspectors. The aim is to increase this proportion to 30%.

- Another key question is whether the tool is accessible and integrated in the working environment of inspectors. MiningWatch has over 60 predictive automated models that run in the background. These are not all shown to inspectors – they only see the more accessible models. Developers check which alarms pass the tests of alarms being investigated resulting in infringements (with infringements) or not (without infringements).

Based on the MiningWatch predictive models, search results rank companies according to their risk level: red (high), orange (elevated), green (medium), and blue (low). MiningWatch can calculate a score for an employer based on data mining and prepares a score card with five variables, listed below in order of importance:

1. Strong personnel turnover
2. Few recent declarations in the DIMONA[1] system
3. Low business turnover
4. Tax variables such as VAT debts
5. Not declaring client listings.

The models are dynamic, they are monitored regularly and closely, so that adaptations are made as and when necessary. If a model drops below an acceptable level of predictions, then new models are re-designed. The models are complemented by Network analysis. If a company that has many links to companies with confirmed cases of infringements, then the ranking of the company increases, further prioritising it as a target for inspection.

Importance of the human factor (including IT team, inspectors, and networks of experts):

A key concern of the developers of Mining Watch is how to maximise the impact of data mining. The human factor and how engaged inspectors can be in the tool is very important in achieving this. A distinction is made between 'Power Users' who will make the effort to make the best use of the wealth of data available, the 'Casual Users' who will use it but not further it, and 'Anti-Users' who are 'Not Believers' in data mining. The number of Anti-Users has been declining over the past 15 years, thanks to general IT developments, but also thanks to concrete results that MiningWatch gives rise to. IT developers have to think how they can help users access information in one click.

In this context, the skills of staff are much more important than big vendor technologies. This includes experienced inspectors, data scientists that have to extensively align business and IT concerns and to translate needs into database features. These staff also need to be able to convince people of the usefulness of the tools. Data scientists also need to help formulate the problem, what is feasible, and to be creative at problem formulation. A key issue is how to attract and retain people with these profiles in government institutions and guaranteeing maximum independence from external consultants who are often not aware of the intricacies of the business itself.

In practice, for MiningWatch, the IT team is complemented by a network of 50 expert investigators. A key method of working of MiningWatch is that the characteristics of each predictive model are presented to expert inspectors and their input is asked for their feedback. This process is repeated until the characteristics from a risk profile are accepted and understood by inspectors. The model is then monitored to see if the profile stays predictive.

From the point of view of inspectors, MiningWatch is a tool to help select inspections as part of their free initiative in the sectors that it covers. MiningWatch is accessible to inspectors, they understand it and can learn how to use it, and they also feel it prioritises candidates for investigation well. Practice has shown that inspectors that did not use MinigWatch had worse results from their inspections so gradually a shared understanding has developed that MiningWatch helps them attain their targets in terms of numbers of investigations, infringements or income and thus use it more.

Summarising important enablers and success factors of MiningWatch, these include:

- **E-government:** Belgium has an extensive programme of e-government since the introduction of the Social Security Card (SIS) in 1997. This ranges from the Limosa system in 2007 which gives cross-border posting declarations and the DIMONA system of declaration of employment contracts, to Checkinatwork system for registering presence on construction working sites which started in 2014. Before these systems, all information was paper-based, which was resource intensive. Modern institutions have digitalised their information collection

[1] The DIMONA system is an electronic system all employers are required to use to register a new employee with the National Office for Social Security

and creation and having electronic forms for the declaration of information allows the calculation of all the social security contributions due by employers and also determines the social benefits for workers. This design allows data warehousing, data mining and advanced analytics.

- **Close collaboration with the business side of the labour inspectorate**: the data tools and processes are there to support the core business of fighting undeclared work. Input is needed and is given by inspectors as the users, in order to keep them relevant.

- **Feedback registration:** A system of rigorous, registered feedback is necessary for the success of the tool. Feedback can be collected in the short-term immediately after inspection, in the medium term such as the confirmation of fraud type X or Y when an inspection case concludes, as well as long term feedback after combining cases over time or across networks such as verdicts of legal cases, the proportion of the contributions actually collected etc.

- **Integrated working environment**: good, well embedded applications contribute to success. This involves single sign-on and one platform. In the past Belgium had many data mining models, while in the near future MiningWatch will become the single platform. An integrated approach refers to the fact that the process does not stop at detection of fraud, it continues through the case treatment and supporting the preparation of the legal case up to reporting of the results of each case and the return on investment.

- **Increased awareness, government endorsement and careful investment:** government endorsement, careful investment and long term maintenance, and attention to the legal aspects of privacy vs analytics are crucial enablers. Countries need to create legal frameworks allowing data mining to support investigation work, policing, and security.

3.2 The Crossroads Bank for Social Security (CBSS)

The Crossroads Bank for Social Security is a gateway for data from 14 social security institutions, and offers electronic services for citizens. The CBSS, despite its name, is not itself a databank, it is a network for data flows from different institutions. Each institution holds its own data, are the authentic source of the data and there are conventions about the treatment of the data, agreed through regular meetings and continuous collaboration. The CBSS initiative started in the early 1990s and has been developing ever since. The legislative changes needed for the CBSS to be created, included the legal translation of a common vision on information management and on information security and privacy protection and the obligation for each institution participating in the CBSS to use unique identification keys for their data.

In terms of success factors, the CBSS was able to gain support thanks to a clear long term vision which was also combined with some quick wins. A key success factor in gaining support included the fact that a small team consisting of experienced civil servants, scientific experts and political advisors worked closely with the Federal Minister of Social Affairs. Critical success factors included a common vision on electronic service delivery, support by policy makers at the highest level, the trust of all stakeholders, and respect for the legal allocation of competences. This top level political support and the gradual involvement of the general managers of all public social security institutions, the social partners, and the general managers of the private social security institutions was also significant. Last but not least, it was important to ensure that electronic service delivery included a multi-disciplinary approach including legal, ICT, communication, coaching, training and change management. Additional success factors included adaptability to an ever changing societal and legal environment, the availability of sufficient financial means and most importantly, a radical cultural change within government, from the hierarchy to the operators, who made the CBSS possible.

The advantages of creating this system included great efficiency gains. The Belgian Planning Bureau calculated that the information exchange processes implied an annual saving of EUR 1.7 billion per year. That was an enormous stimulation in continuing the process. There were also gains in speed and in effectiveness, as the CBSS made possible the provision of services of better quality as well as the provision of new types of services, such as personalised simulation environments and a push system of automated granting of subsidies.

3.3 Cross border cooperation in Data Mining

In the context of the Benelux Union, there is cooperation on preventing social fraud and social dumping. Top level political support and the importance of human resource capacity are equally crucial at the cross border level, as at the national level. The Benelux Steering Committee on Social Fraud/Social Dumping has recently focused on a pilot project on sham constructions. The idea was to test the need for and relevance for automated data sharing, matching and mining and exchanged 414 files between the Netherlands and Belgium. The pilot demonstrated the need for data matching and mining as many potential issues for investigation were identified. In the majority of cases, there was a medium to high risk of the employee not being insured or insured in the wrong Member State. Many mismatches of names and dates of birth were also identified, highlighting the importance of correct identification. The pilot project has led to improvements in daily operations but follow-up is difficult, in terms of what the Member States can/will do with the results of the pilot.

4. Summary of round table discussion on the draft Toolkit on Data Mining

Following the presentation of the experience of Belgium, the follow-up visit concluded with a discussion on the Practitioner's Toolkit on Data Mining which has been developed as a follow-up activity from the Helsinki workshop. The Practitioner's Toolkit provides concrete guidance in planning, designing, implementing, monitoring, and improving data sharing and data mining tools which have the potential to be implemented in EU Member States. The Toolkit also explores the challenges that need to be overcome in order to develop efficient data mining systems. The Toolkit also discusses return on investment thanks to a data analysis tool, and how this can be measured.

The Toolkit is aimed at national enforcement bodies who already implement or wish to implement practices of data collection, data sharing, data profiling, data matching, data mining as well as data protection and data quality management. The Toolkit is useful both for enforcement bodies who are at an early starting point and for those who are more advanced in the process.

The Toolkit outlines the key questions to be considered in each step of developing a data mining tool, including the questions a country should be asking*:*

- **During the initiation phase** including, what are the ambitions? What are the expected costs and benefits of an intervention? What are boundaries in terms of availability of data, technical feasibility, financial feasibility, political and legal boundaries etc.

- **During the definition and planning phase:** What data will be collected/shared? What do we have? What do we need? What are the partners involved? Are we allowed to? Are we willing to do it? What will be the 'privacy/data protection by design' strategy? What will be the data governance strategy (data protection and data security)? Etc.

- **During the implementation Phase** preparation of data for analysis, testing and implementation of the data analysis tool.

- **During monitoring and evaluation** input and feedback from inspectors, assessment of model outcomes etc.

- **While measuring efficiency and effectiveness** including coverage; accuracy in terms of the percentage of cases predicted correctly by the model etc.

5. In conclusion

When developing data mining systems, it is important to remember that not everything needs to be done at the same time: for example, Belgium focuses on four industries with the highest risk of UDW and so does not target all sectors. Since there is so much available data, it is important to prioritise.

Both at EU level and at national level the cooperation between tax administration and social security administration is crucial. Belgium has demonstrated that it is not necessary to merge the tax and social security administrations, they only need to be convinced to share their information.

It is important for the IT developers involved in data mining to feel they have a role in fighting fraud. Data miners can identify the risk, but can they also estimate the size of fraud, to give estimates of return on investment figures which are needed for policy makers to make the case for data mining - even poor statistics are better than no statistics. Poor statistics can be corrected, while a complete lack of statistics does not offer a basis for moving forward.The role of data mining and matching needs to be seen in a wider context. As preventing and tackling UDW cuts across multiple public sector organisational boundaries it increasingly requires a multi-agency approach. With this increased need for effective inter-agency cooperation is the essential requirement to be able to understand the data available and to share good quality data. This is the case at the national level and increasingly at a cross-border level.

With this requirement to work with other organisations nationally and cross-border is the need for protocols and bilateral agreements. Such agreements set common goals and facilitate work across different legal frameworks, importantly they establish a clear legal basis for cooperation and data sharing.

Ultimately the workshop and follow-up visit have shown that data mining is a useful tool for more efficient enforcement. However, tackling the complex problem of undeclared work requires a holistic approach. This means that to be successful, in addition to inspections and enforcement agencies need to take a preventing, deterring and combating approach as well as promoting the declaration of undeclared work.