European Platform Undeclared Work

Thematic Review Workshop on Data Mining for More Efficient Enforcement
1 & 2 June 2017, Helsinki, Finland

The aim of this Thematic Review Workshop was to provide an opportunity for Platform members to improve their knowledge and awareness of the importance of building efficient data mining systems in order to predict, prevent and detect undeclared work (UDW). The workshop provided an opportunity to exchange best practices, identify successful approaches to data mining at national and data sharing at cross-border level which have the potential to be transferred to other Member States. The workshop also explored the challenges that need to be overcome in order to develop efficient data mining systems.

Key findings:

- There are large variations in the **maturity levels** of EU countries in relation to data mining, with some countries in the very early stages of data gathering whilst others are using sophisticated techniques and technologies.
- A distinction needs to be made between **data sharing** (making data available to other users), **data matching** (large scale comparison of records or files collected for different purposes) and **data mining** (find patterns, correlations, outliers to be used for inspections). While the majority of Member States focus efforts on data matching, fewer undertake data mining activities.
- A key **problem that many Member States face is sharing data between ministries and other public bodies nationally**. The Grey Economy Information Unit (GEIU) was set up in Finland to address this problem. The unit is a central point for producing and sharing information on the grey economy and its control.
- In order to implement effective data mining systems, **a first key step is to remove legal barriers to the exchange of information** between agencies. Finland's GEIU is the central unit holding all data, so the challenges of sharing data between agencies was addressed early on. Authorities with the appropriate legal permissions can then become a client of the GEIU and access the different reports available. The Finnish Parliament has been responsive in removing legal barriers along the way.
- A key lesson is to **have good, accurate, and well-structured data.** Poor quality data can create more problems than it solves. It is also important to have a robust data referencing system with good descriptions of the data explaining what they are and the sources. It is crucial to devote time and resources to clean the data and to reference it clearly.
- **Detecting outliers** is the way in which some Member States (UK, Belgium) go about data mining. Detecting anomalies is primarily about defining normal behaviour then, companies deviating from that, may merit investigation. Convinced that fraudsters are often connected to each other (for example, via the same accountant, managing directors, clients, suppliers, etc.), Belgium has started to use machine learning and network analytics to rank and profile cases. The resulting reports are user fiendly for inspectors with colour coded results, flagging potential fraudsters and their networks. The first results show that this helps change the ways inspectors work by greater targeting of their work and undertaking inspections at multiple locations in parrallel. The UK also uses dynamic benchmarking to spot different patterns between similar businesses in close proximity.
- **Data protection and data security** are key issues which need to be build into any data gathering and mining system from the very start.
- **Return on investment should be more clearly measured** in order to help agencies understand the impact of more advanced data usage on the outcome of inspections. Such evaluations also increase internal and external awareness of the capacities of public administrations and the potential value of investments in data sharing/matching/mining in order to prevent and deter non-compliance with labour and tax rules.

KEY OUTCOMES

- Effective data sharing, matching and mining is **a step by step process** which requires political will and trust between the different parties involved.
- Data mining can support the ambitions of public administrations to **tackle undeclared work**. Currently, there is a plethora of initiatives across Europe, which are at different stages of development, and are sometimes fragmented.
- The focus needs to be on setting up effective **national systems** before then progressing to more ambitious cross-border cooperation. Look internally first as it is often surprising the data available within your own organisation and freely available on the internet.
- There appears to be little information on the **critical assessment of data mining initiatives**. It is important to define useful indicators to measure the performance and effectiveness of tools and to ascertain the return on investment.
- The **investment in technologies** ranged considerably from the use of free open source software to specially designed programmes and systems. There was a marked difference in the resources available for labour inspectorates and tax authorities.
- There is a need for **specialist staff** as this area of work becomes increasingly technical.
- It was clear from feedback that **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.
- Some of the technology being used is **user-friendly** and presents the data **visually** making it easy for inspectors to access and use. However, the data are still not open enough to all potential users. Sharing of data mining results can still be improved.
- The **quality of the data** being used is key and as more data becomes available a key question is how to get accurate data. Either initiatives have to start with good data or resources are needed to ensure that the data being used is clean, accurate and reliable.
- **Interoperability of the data shared by agencies is also crucial**. Clarifying terminology between agencies at national level is another important step. Belgium's Crossroads Bank started off with 120 definitions of 'wages', which were eventually narrowed down to 12.
- **The complexities of data sharing** at a national level are intensified when looking at a cross-border context. To be successful it needs a high degree of political will and trust between countries as well as a clear idea of what data needs to be shared.
- **Mutual learning is useful for those Platform Members who are at an early starting point, and those countries who are advanced in the process can also learn from each other.** Even for those countries who have the centralised data warehouse, there is more to do. The UDW Platform has the potential to bring Member States together, to facilitate exchanges between countries who have made advances in data sharing and data mining, and other countries which can step up their data mining tools and data sharing processes in order to implement and further efficient combatting of UDW.

Further information: A Learning Resource Paper and a Practitioner's Toolkit providing concrete guidance in planning, designing, implementing, monitoring, and improving data sharing and data mining tools will be produced later this year.

The workshop is part of a learning process which will be continued by a Follow-up Visit to one of the participating countries.

The input documents and presentations of the workshop will be uploaded to the collaborative workspace of the Platform.