



Improving cross-government data and information exchange on national level and identifying good practices

**Ruslan Stefanov,
Center for the Study of
Democracy
in collaboration with ICF**

Report of the peer learning dialogue

December 2021

Contents

1.0 INTRODUCTION	3
2.0 DATA AVAILABILITY, FRAMEWORK CONDITIONS, AND INTERNAL PROCEDURES.....	3
2.1 Availability of data and trust among authorities. Political and legal challenges	4
2.2 Overcoming political, trust and legal challenges	6
2.3 Pilot testing of actions: key takeaways	8
3.0 IT INFRASTRUCTURE AND TECHNICAL REQUIREMENTS.....	13
3.1 IT infrastructure and technical requirement challenges	14
3.2 Overcoming IT infrastructure and technical requirement challenges	16
3.3 Pilot testing of actions: key takeaways	17
4.0 REFLECTIONS AND SUGGESTIONS.....	19
4.1 Suggestions for countries	19
4.2 Suggestions for ELA	20
5.0 BIBLIOGRAPHY	22

1.0 INTRODUCTION

Dialogues for peer learning and cross-border action (hereafter PLD) are a demand driven activity of the European Platform tackling undeclared work (hereafter 'the Platform'). Their aim is to explore an issue in greater depth with a small group of peers and/or to 'test' new approaches for tackling undeclared work.

This report documents the outcomes of the PLD on improving cross-government data and information exchange on national level and identifying good practices. The PLD aimed to produce recommendations and suggested actions (1) which Platform members can implement in their national practices and (2) where mutual learning at ELA and Platform level could occur to prevent all authorities having to "reinvent the wheel".

Discussions focused on the improvement of the national-level data exchange practices, as a first step towards the establishment of efficient cross-border information sharing procedures.

Representatives of five countries (Belgium, France, Greece, Latvia, Portugal) discussed this topic in three meetings held from October to November 2021:

1. Identifying and overcoming challenges related to data availability, political will, trust, national legal framework, and internal procedures.
2. Identifying and overcoming challenges related to IT infrastructure and technical requirements.
3. Sharing feedback on feasibility of actions/recommendations/ proposals in terms of challenges and tips on overcoming them, and information/evidence on specific proposals. Agreement on the key actions / proposals / recommendations to ELA and Platform members.

The first two meetings concluded with a set of actions, which participants took forward and tried to implement in their home countries. The lessons learned from these actions were then reported and reflected upon at the last meeting. This outcome paper summarises the discussions during the three meetings.

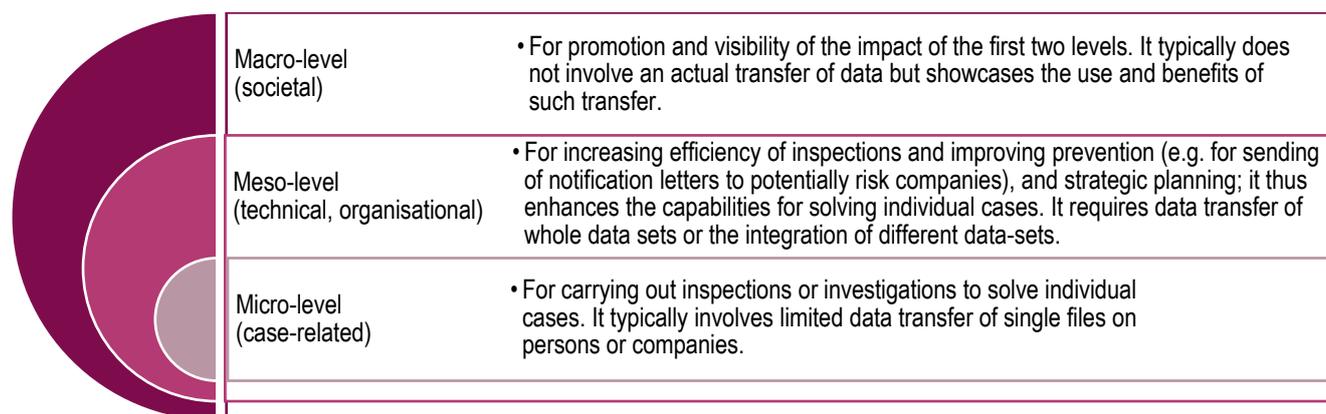
Sections 2 and 3 report the reflections on data availability, framework conditions and internal procedures and on IT infrastructure and technical requirements respectively. Both sections include the key takeaways from the pilot testing of actions. Section 4 draws conclusions, reflections and recommendations for further action by the Platform and its members.

2.0 DATA AVAILABILITY, FRAMEWORK CONDITIONS, AND INTERNAL PROCEDURES

Despite the development of databases across the enforcement authorities responsible for tax, social security and labour law compliance, a fully coordinated approach to data sharing is often missing.¹ In [the 2019 Annual Platform Survey](#), authorities reported challenges with the slow and inefficient data sharing in Member States. The [2017 Annual Platform Survey](#) highlighted, that only 57% of enforcement authorities involved in tackling undeclared work have access to a database that would allow them to detect potential instances of undeclared work.

¹ European Platform tackling undeclared work. (2021). Dialogues for peer learning and cross border action: Improving cross-government data and information exchange and identifying good practices. Background report.

Figure 1. Data exchange is essential for tackling undeclared work on three different levels



Source: European Platform tackling undeclared work. (2020). Seventh plenary meeting. Report on data protection and data exchange.

Hence, the PLD focused on identifying the key challenges and possible solutions in:

- ▶ Changing the way labour inspectorates and other authorities exchange data and create trust with counterparts. Ensuring political support and improved internal procedures.
- ▶ Changing the way labour inspectorates and other authorities manage data privacy issues.

2.1 Availability of data and trust among authorities. Political and legal challenges

Key challenges identified by participants

- ▶ Lack of political will/culture of exchange
- ▶ Lack of understanding of data protection regulations
- ▶ Missing operational processes
- ▶ Exchange with tax authorities seems to be more limited, although relevant
- ▶ Some countries, lack a detailed legal framework, and rely on informal cooperation
- ▶ Different systems and targets among authorities
- ▶ New data sources call for new cooperation solutions (e.g. social media and online data scraping)
- ▶ National vs local level coordination
- ▶ The paradox of too much information and ability to use it

Data sharing requires **political will** and **trust** between the different parties involved, and a clear idea of what data needs to be shared. Willingness to exchange data requires an understanding of its added value for all. National legislation can encourage or discourage data sharing between authorities depending on the clarity of regulations, established definitions, procedures, etc. All Member States have legislation in place that protects **personal data** and safeguards **privacy**, which could be the key prerequisites or stumbling blocks for effective data sharing.² The EU's General Data Protection Regulation (GDPR) requirements are an issue for more than half of Southern and West European countries, according to the 2019 Annual Platform Survey, while the lack of resources to respond

² European Platform tackling undeclared work. (2021). Dialogues for peer learning and cross border action: Improving cross-government data and information exchange and identifying good practices, Background report.

to GDPR provisions poses a barrier for Northern and Southern European countries.³ In addition, the public authorities should create their own **internal procedures** for data gathering, storage, processing and sharing, according to the national-level legal base.

The participants reached the following conclusions:

All authorities are concerned about data and information sharing. There is a common understanding that information is power, and not all political leaders are willing to share that power. Hence, convincing the political leadership of the need and value of data and information sharing is critical. Inspectors working in the field are critical to defining data needs, emerging trends or gaps, around which internal procedures and national legislation for data and information exchange could be built. Sometimes there is more political will to find a solution on an international level than on a national level, prompted by social dumping and cross-border problems.

- ▶ **Political will** is the single most important prerequisite for successful investments in data and information exchange on a national level. To reach this point, policy-makers need to understand the outcome and benefits of the data exchange. Labour inspectorates in most Member States seem to get information on an individual level without any problems. But they usually cannot access the systems without relying on informal contacts. To exchange data effectively, the authorities need to have the same goals.
- ▶ The main drive for data exchange comes from the political level. For example, in Belgium, the highest-level policy-makers have regular meetings between labour and social security services. However, such top-down solutions fail to appreciate and work on **bottom-up solutions**. Inspectors from different government services are best at defining the needs and picking up early warning signals.
- ▶ There are varying **cultures** for the exchange of data and information, which sometimes depend on the amount and value of data held by the institutions. This is visible in France, which has a legal basis for data and information sharing between the inspection services within the public administration, yet still lack full interoperability of databases. **Joint actions** between authorities allow for better understanding and improve data and information exchange culture.
- ▶ **Cases of great public concern** typically provide a boost for data sharing solutions. For example, check-in at work data, requiring everyone at the construction site to register, emerged from cases of accidents on a construction site. The **emergency** of the COVID-19 crisis has pushed governments across the EU to start delivering more services in an electronic environment. E-solutions have proven very effective to provide services, which could be a good factor in convincing policy-makers to do more data and information sharing.
- ▶ **Data protection** laws and their update to reflect the EU's GDPR Directive added barriers to the data exchange in the short term, as public authorities scrambled to allocate resources to respond to the introduction of the new requirements. Yet, GDPR provides a solid basis for legal handling of personal data and information and should allow for better data and information exchange, but it also slows down the speed of data exchange. Labour inspectorates had to put in additional efforts to comply with GDPR, which in many cases resulted in less willingness for data and information exchange. A well-functioning data-protection system is critical to developing compliance with GDPR and ensuring the legality of data exchange. In countries with well-established data sharing legislation and practices like Belgium, there are very strict practices for **justifying data enquiries** (Data Protection Officers, DPO) on a case-by-case basis, in regular intervals, and before establishing a data sharing routine. GDPR also provided a lot of good publicity regarding the issue of data collection and data-sharing.

³ European Platform tackling undeclared work. (2019). 2019 Annual Platform Survey: [Tackling undeclared work in the collaborative economy and bogus self-employment, data exchange and data protection, and cross-border sanctions](#).

- ▶ Data and information are **more readily available and exchanged between departments within the same institution or policy domain**. Typically, labour inspectorates have easier access to labour relations (incl. labour inspection and OSH data) and social security data. However, getting data from tax authorities or other less related government services is typically more difficult. Some information can be shared electronically but on a very limited scale and case-by-case. Sometimes the lack of exchange causes data collection overlaps between institutions.
- ▶ The lack of a **solid legal base** and **culture** for data exchange reduces the capacity of enforcement authorities. Data sharing is also important for **prevention capacity**. For example, on undeclared work, the labour inspectorate data can be compared and matched with social security, tax, migration or road transport data. This way, inspectors are better prepared and have a better historical overview of people and companies.
- ▶ In addition, there is a need for **internal procedures** to be well aligned with the legal base and between the different authorities. Also trust is needed to find quick solutions in emergencies, like COVID-19. Trust often bridges the gap between the desire for swift action and the necessary bureaucracy created by legal requirements.
- ▶ Data-matching and mining require **long-term interoperability connections** between the systems. This technical side also calls for a common vision and goals between the different authorities over the long run, as it requires alignment between the needs and benefits of data exchange and the need for underlying investments in technical infrastructure and capacity building.
- ▶ **New data sources**, such as social media and web-scraping for information from websites, are promising for tackling undeclared work, but pose further questions and data privacy issues. For example, can cameras from vehicles and homes be used, in what cases, and for what purposes? This also calls for cooperation, as some government departments, such as police, have higher levels of authority to access such data.
- ▶ Lack of good data policy coordination between national and local levels can disrupt effective data sharing in tackling undeclared work. For example, in tackling undeclared work at the local level, information from the municipality on the number and places of container requests for renovation services can be matched with statements about hiring for renovation services and/or requests for construction permits. Administrative decentralisation requires better **coordination capacity** for data sharing.
- ▶ Fragmentation and lack of trust can be overcome by specialised **teams of data investigators** or **centralised agencies** that provide access to the data of different authorities without moving them to one central location.
- ▶ Sometimes, there could also be **the phenomenon of 'too much information'** if it is not sufficiently structured. This can lead to additional problems, requiring data matching and mining. Just providing access to databases is not enough. There is a need for **training on data handling**, understanding and interpreting the data. People need to know what each piece of data means and what it is part of, how it was collected, and for what purposes it can be legitimately used.

2.2 Overcoming political, trust and legal challenges

Key solutions proposed by participants

- ▶ Need for strategic vision (then legal framework could follow); need to convince political level, agree on common outcomes, but also via societal, economic and political benefits, showcase return on investment
- ▶ On the operational level, need to develop new data-sharing tools and procedures bottom up to some extent and communicate them throughout the hierarchy (e.g. Belgium's quick response team)

- ▶ Nurture informal contact and joint activities important to understand partners mandate and to establish cooperation
- ▶ Address capacity and skills needs: train the trainers, exchange of staff with other organisations
- ▶ Introduce centralised e-agency or data investigation teams, which possess specialised data knowledge or IT capabilities, which they can then 'lend' to other national or local authorities and drive data exchange
- ▶ New data sources call for new cooperation solutions (e.g., social media and online data scraping) – beware it might not always be legal or allowed in court

The participants reached the following conclusions:

- ▶ If full access to the databases of other organisations is not feasible for some reason (often the case with tax authorities), then there are alternative solutions for **partial or limited data access**. For example, following preliminary discussions, it can be agreed that data be made available to a selection of labour inspectors, for a specific period, for working on a specific undeclared work problem. Still, a formal process should be in place and described in detail.
- ▶ There are many **ways to build trust among authorities**, such as (i) short internships among different institutions to create better understanding between the different organisations; (ii) communication and experience sharing; (iii) detailed data protection safeguards and established internal procedures. For example, the institution that needs the information has to explain why it needs it, what measures it will take to protect this information, how is the requested information's size and detail proportionate to the needs, what else can be expected to be done with the acquired information, etc. This way, the asking institution builds trust and understanding in the receiving institution.
- ▶ The discussion on data sharing with other authorities needs to happen on **many levels**, e.g., management, inspection, etc. For example, when joint inspections / common experiences are organised, both sides can better understand each other's constraints and possibilities. Sometimes the other side might not see data exchange as a win-win situation, and it needs to be defined in terms of a public service or duty to demonstrate its common added value. Sometimes it is the way each side communicates those matters, including using the same language and definitions of concepts, that eases or constrains data exchange.
- ▶ Often, solutions can be found by **organising data sharing through intermediary institutions**. This could be a central inspection intelligence centre or service, that manages data exchange between the different inspection authorities in the country (labour, social security, tax, police, local, etc.). Intermediary institutions can bring together different inspection services to develop a specific, common solution. They can give access to the needed data, allowing everyone to see the information in the same format, as in a central database, a warehouse or an exchange.
- ▶ Funding is often an issue about who pays for such joint initiatives. From an individual budget perspective, institutions might be wary of cooperation because it could lower their budget for other activities. Therefore, when seeking policy-makers' understanding and buy-in, it is important to provide a **return-on-investment point of view**: when a specific problem is tackled, what would it bring back to the institution. The management might want to hear what the benefits would be, e.g.:
 - ▷ public service;
 - ▷ social responsibility;
 - ▷ preventing or addressing a big socially sensitive accident;

- ▷ associate the social partners and win their support; they often can also help with the solution;
- ▷ solving social dumping issues, which then attracts the big business, which could be a powerful ally;
- ▶ Having specific legal arrangements or conventions (e.g., in France) about tackling undeclared work, including at sector level, helps streamline and regulate data sharing. **Partnership agreements** and **memoranda of understanding** can also help exchange data and bring additional flexibility, e.g., to include social partners.
- ▶ Data sharing requires different **capacities**, developed through continuous and sophisticated **training**, e.g. data filling, handling, processing, using the data for building cases and knowledge (e.g. how you move from one step to the next, how you combine the data for meaningful information). More experienced people in using the data and deriving knowledge from it can train others about: what data is available where, and what knowledge can be drawn from the available data. Training is needed on how to treat and use the data, as well as on what not to do with it, so that evidence holds in court. There is a need for better **step-by-step instructions** with 'do-s and don't-s'.
- ▶ There is a **huge new data domain**, which remains unexploited although it is the driving force of today's economies (Internet, search engines data, social media, etc.). There is a need for inspection authorities to develop knowledge and protocols for using it. Some countries, like France, have already started legally defining the possibilities for collecting ('scraping') such Internet data. But judicial practice is still patchy, and there is a need to collect more information about its practical implementation. Not everything that is on the web can always be used for evidence. And Internet companies warn explicitly their data cannot be used for other purposes without their consent.

One way to use publicly available Internet data more efficiently and data available in public administration is by **bringing in specialised expertise** (e.g., the data scientist teams in Belgium). They are usually seconded from a contracted ICT company to build systems that the inspection services need. A good practice is to involve data scientists alongside inspectors to find solutions for undeclared work. For example, Belgium has created a **rapid response** team, that includes inspectors who can swiftly test new ideas on the ground. The team consists of the most experienced inspectors who can work across the country. There is a constant exchange between the data scientists and the field team. The rapid feedback is used to improve the data models later mainstreamed across all inspectorates to detect networks of fraudsters active in different domains.

2.3 Pilot testing of actions: key takeaways

During the first meeting of the peer-learning dialogue participants agreed on the actions to test the knowledge and recommendations. This section highlights insights gathered through them.

Presentation and discussions of the following actions:

- ▶ Assess / list capacity needs / gaps for data sharing, and in what areas (training, skills, etc. for staff). If there is already training available, what else is needed? (**Action: BE, EL, FR, LV, PT**)
- ▶ Research in own organisation on data needs and what are the most important authorities to get data from (**Action: BE, EL, FR, LV, PT**)

In **Greece**, labour inspectors would be very interested into having access to different databases for inspections analysis and planning purposes:

- ▷ of the Social Security Institution for social security contribution details, findings and fines of their inspections, companies risk assessment level (if any), etc.

- ▷ of the Employment Agency, for unemployed persons;
- ▷ of the Tax Administration, for companies' income level and companies risk assessment level (if any).

There **are two successful foundations** on which such better data exchange can be built. One is the existence of the ERGANI labour registration system. And the other is the Interoperability Centre at the Ministry of Digital Governance (KE.D), created to ensure IT systems' compatibility and to ease data exchange. The latter though could create delays as data requests increase. One of the key concerns regarding data exchange remains the lack of detailed knowledge of GDPR requirements. **Training and specified guidelines** are necessary regarding GDPR and data use. One possible way of going around individual GDPR concerns is to provide labour inspectors with access to risk assessment tools for undeclared work, rather than specific data pieces. In addition, there is a need for a better understanding of what new knowledge can the data provide and how it can be used to improve the work of inspectors in the field. Often, people say they would like to have access to data, but they lack a detailed understanding of different purposes they can use data.

Although ahead of most other Member States in data sharing with its DOLSIS system, **Belgium** is missing access to **tax data**. The following tax system information adds value to tackle undeclared work: (a) outstanding VAT obligations, or specific company information; and (b) particular data on self-employed persons. Access to tax data would save resources and reduce the burden on the employees and the employers. In addition, it would allow to coordinate inspections of social security, labour and tax authorities. However, access to tax records in Belgium is granted only on a case-by-case basis. There is still hesitancy, rooted in privacy concerns, to bridge social security and tax data systems for inspection purposes. The Belgian social security authority is currently working on a proof of concept exploring technical possibilities to exchange information with the tax office automatically.

The process of granting access to data in Belgium starts with a clear vision and willingness at the top political level for the need of data exchange. Then, the authorities follow **a three-stage process** to arrive at a solution:

1. what are the technical needs/prerequisites to allow such access; is there access to an existing technical solution, and do authorities have enough technical capabilities to realise it;
2. the authorities prepare another proof of concept and file it to the data protection officers; the concept states that the authorities will only test (dry run) the system, and won't do any inspections based on the data exchange; this allows for an easier approval;
3. permanent system set up; this involves a legal team proposing the necessary legal changes to make the data access and exchange possible.

To **avoid privacy concerns**, data could stay within the domain of each department that collects it rather than being stored in a new location. The system thus creates a temporary information bridge from one organisation to another, which authorised officers can use to cross and get the specific data they need, for a limited time.

The Belgian social security authority is using data in **predictive modelling**. This requires sufficient historical records to put in a training set and use in machine learning. For example, based on 2016-2017 information, the model can be "fed" with data for 2017-2018 to predict undeclared work behaviour in 2019. Also, models' predictive power depends on the number of data sources underpinning them. Models would become better if tax data would be added to labour and social security information. Such models can predict the willingness or the ability to pay with goods, which aids inspections.

Another type of data useful for social security inspections is **residence and work permit information** of people from the EU and third countries. These permits are managed at the regional level in Belgium, making access more difficult. Currently, legal framework and technical tool are underway to allow social security inspectors to access

residence and work permit data. Acquiring permit data on a structured basis can be very useful for data mining purposes.

An important element of effective data sharing systems is the **traceability and accountability of use**. In Belgium, the DOLSIS system contains logs showing who accessed what data, when, and for how long. In addition, there is a compulsory field in the system, asking people who request access to briefly explain why they need the data. Finally, a sample of random system users is selected monthly to justify why they have logged in.

Training needs in data sharing and data matching are permanent to ensure that all the data is used according to the best privacy standards. In the first place, training is needed for inspectors to understand how they can get useful information from the available data. Second, it is needed to raise awareness of GDPR. Inspectors have the most access to the databases from the whole organisation. Hence, they need to know well how to use information ethically and in compliance with GDPR.

Portugal has identified different capacity building needs for access, equipment, and human resources. Labour inspectorates in Portugal only have experience with data sharing with social security authorities. They would also like to gain access to tax, migration, and transport authority data (e.g., goods delivery driver's licenses to uncover undeclared work). They do not yet have experience in data mining. There is certainly an understanding and desire to share data with other authorities and gradually reach the levels of data proficiency of more advanced countries. This would require similar efforts as in Belgium, including assigning different profiles and levels of access to data; precluding data transfer but only consulting it in the other organisation; ethics training; etc. In terms of **equipment**, Portuguese labour inspectors only have personal computers to access data from their office or home (through VPN). They do not yet have access to the information on the go via tablets or phones. Such on-site access is a much-needed next step in data gathering and sharing. There is a need for **training** in data management, soft skills on ethics, GDPR, etc.

In **Latvia**, the labour inspectorate would also like to have access to tax and migration data. In terms of equipment, they would want to have access to tablets and phones to check data on the go.

In **Belgium, Greece, Latvia and Portugal**, the tax authorities are seen as better than labour inspectorates in national data gathering and internal sharing, and information system. One of the reasons for such understanding is that tax authorities have bilateral agreements with most countries. Sometimes justice and police departments have more opportunities to gather information because of their special authority. Typically, police officers can get more access to information than labour inspectors.

Tax authorities in Belgium would not like to disclose what type of data they have access to or how they use available data. Even upon request by phone from their social security colleagues, they would not want to share such information, which precludes any understanding of how tax-related data can be combined with social security and labour data to detect undeclared work.

In **Greece**, tax authorities are also unwilling to share this information with labour inspectorates and other authorities. **Latvian** tax authorities only share information on the start of the employment status and salaries with labour inspectors.

Participants noted that an important prerequisite for effective data-sharing is to **know the information landscape** of the other authorities in the country. To build good models for tackling undeclared work, it is important to know what information other authorities have access to.

In this respect, higher levels meet and talk in **Belgium**, but new data-sharing opportunities are often better understood bottom up through joint inspections. Then people in the field become aware that the different agencies have different tools and information, which can be shared to produce better results. It is thus important that knowledge about data on the management level is combined with such from the field.

In **Greece**, the Ministry of Digital Governance has taken important steps in this direction. The ministry has created an **Interoperability Centre**, an information system that provides a unified infrastructure for the installation and use of online services, through which business data is exchanged between public authorities, ensuring high information security. By law, all authorities that need to exchange any data, have to proceed through the Interoperability Centre, which ensures the unification of all interoperability services in the public sector. Yet it might create delays in establishing a required connection due to increased demand for data exchange throughout the public sector. The Centre has a website detailing available web services, allowing authorities to seek ways to use them, aiming to increase their productivity and their online services provided to citizens and businesses.

An important element of data sharing is how labour inspectorates are organised within the rest of the public administration. For example, in Germany and the Netherlands, they are under one authority with the fiscal specialists and this could foster the cross-usage of data.

Presentation and discussions of the following actions:

- ▶ Meeting with other authorities (e.g., social security) to discuss which and how data could be better shared based on the experiences gathered from this first meetings (e.g. for joint action to do cross-checks of data) (**Action: PT**)

One of the follow-up activities in **Portugal** was a meeting with social security colleagues. They are close to the labour inspectorate and share some information related to labour matters. Labour and social security authorities are in the same ministry in Portugal. Hence, meetings and data sharing are easier. At the same time, calling upon regional contacts in the country in tax and border authorities is not that easy. These authorities raise questions about legality and GDPR. So, it is a matter for building trust between decision-makers from different institutions to bring down institutional silos.

In this respect in **Latvia**, sometimes the **use of personal networks** could help speed up data sharing and the development of follow up steps.

Presentation and discussions of the following actions:

- ▶ Check what information on existing experiences could be shared with the group (e.g. on data mining, data sharing, proof of concept for web scraping and what are the needs related to it or look into an example of working with data scientists) (**Action: BE**)

The social security authority in Belgium has been experimenting with new ways to **acquire and match data from the internet**. Its early experience demonstrates the need for keeping an edge on the technical understanding of modern data sources and the limitations of such an approach. They have **tested web-scraping** as a potential tool to better target inspections. For the time being, the development of the pandemic and the lack of precise legal basis and data protection/privacy rules, this effort has been temporarily postponed and its results not yet completed.

Four ideas were considered:

- ▷ Google search for finding missing companies that are not present in the social security databases;
- ▷ HoReCa platforms' search to compare activity online and cross-check with data reported to the social security authorities;
- ▷ collaborative economy platforms' checks for detection of self-employed or unregistered jobs;

- ▷ foreign registers: search the data for foreign employers involved in "network fraud case".

The Belgium authorities did a **proof-of-concept on HoReCa sector**. They tried to link companies established at popular platforms (for reservations and assessment of services well known by the general public within the HoReCa sector) to the information in DIMONA – application for the declaration of a labour relation.

The **methodology** is essentially three-fold. First, link entities found on platforms with those declared in the official databases. Then, evaluate the activity of these establishments (frequency of reviews, number of rooms, menu prices, etc.). And finally, compare with the activity declared in the databases (number of workers, turnover, etc.).

There is a need to **start small** when such data collection techniques are piloted. The first choice is of the right technique for data scraping, as those that are easier to use or better at collecting larger quantities of data are also those that are more detectable or outright prohibited on certain websites. To further narrow down the experiment, the focus was put only on the Brussels area and on certain type of information on the websites, which can be cross-checked easiest with data from the Crossroads Bank for Enterprise (CBE). The following data was collected: company name(s) and address; telephone number; and number of visits / comments. Despite the narrowing the focus, more than 7 500 units of data were collected. The data is then matched to official records through a "**keying**" **technique**, which allows for different matching levels, as the data from the two databases can never be identical. For example, the data label in the official database could be "address", while in the scraped data "contact", etc.

As a result, **three options** are established **for each row of data** (usually denoting a certain company): (i) match – data in both databases is identical, so most probably no undeclared work; (ii) doubt – some of the data matches but other pieces, not; this could be the result of non-declaration or simply an artefact from the database; and (iii) no match – lack of official data to corroborate the scraped data, which could mean entities are completely off the books (and would warrant an inspection), or that they operate under different legal and brand name, etc.

The **results** of the experiment suggested that:

- ▷ data scraping and matching proved effective for HoReCa, which is a very visible sector, i.e. has a lot of online interactions;
- ▷ it is unclear whether more "underground" sectors (such as nail shops, car washes, etc.), which often do not have any websites, would produce similar results;
- ▷ The experiment linked 70-75% of businesses found in the official database with high reliability.

What some **possible next steps** could be in taking this experiment further is to search for other use cases and/or confront the data with findings in the field through inspections. In addition, in terms of better technical preparedness for such exercises in the future some of the lessons are to better manage addresses, add geocoding (use of geographic coordinates) to the collected data from the internet, and employ GIS / API sources: OpenStreetMap, Google.

However, the social security authority in Belgium **decided to not go forward with the exercise** as the discovered that:

- ▷ there is no solid legal basis for data scraping;
- ▷ often data that was thought of as not personal, turned out to actually personal details, which entails privacy protection;
- ▷ websites have restricting terms of use, and in most, it is stated explicitly that the information contained on them cannot be used for web-scraping or other purposes than the intended;

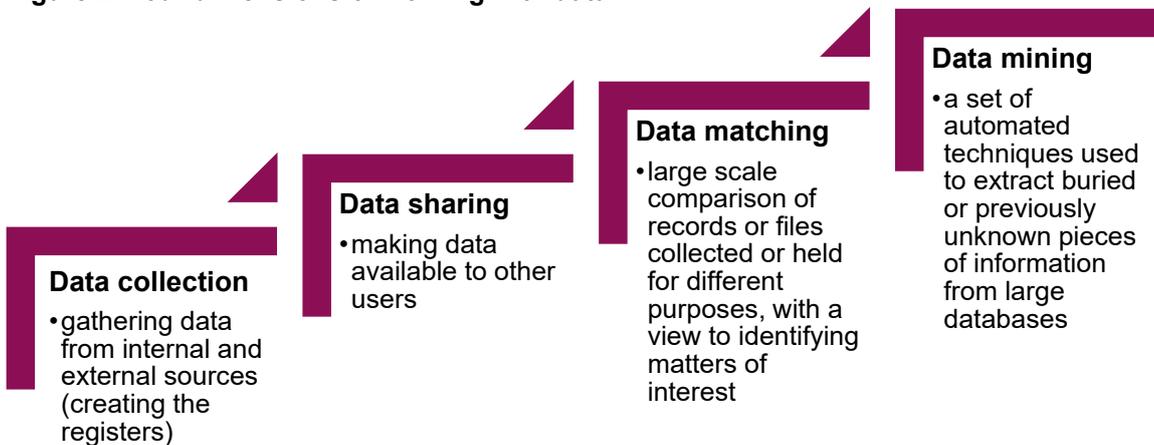
- ▷ the initial data findings could not be verified in practice, as there was no possibility for action on the field due to Covid-19 restrictions.

It is important to introduce some specific legislation to cover the possibility for labour and social security organisations and authorities to use scraping and internet data use, so that they are not vulnerable to law suits. For example, In Belgium the police can use such data from scraping in criminal cases but social inspection services, not. Either the inspection services need to get the same authority as the police or they would need to introduce legislation that would provide them with authority for web-scraping.

3.0 IT INFRASTRUCTURE AND TECHNICAL REQUIREMENTS

Sharing data across government departments requires a cross-government information technology infrastructure that actively supports standardised processes. The design and the architecture of the information technology infrastructure must reflect the operational needs of the enforcement authorities and be capable of being updated without prohibitively high effort and cost. It also needs to take into account if the data in the individual public authority is stored centrally, or if it is decentralised in different departments, and needs to be gathered and consolidated first.⁴ According to the 2017 Platform survey only 13% of the countries state that the database in their ministry/enforcement authority is inter-operable with other databases of public authorities also involved in tackling undeclared work (Belgium, Estonia and Latvia).⁵

Figure 2. Four dimensions of working with data



Source: European Platform tackling undeclared work, (2017). Data Mining for More Efficient Enforcement. A Learning Resource Paper.

The peer-learning dialogue discussions focused on identifying the challenges and solutions for:

- ▶ Achieving improvements in the current IT infrastructure, and the way data is classified and input into databases; and

⁴ European Platform tackling undeclared work. (2021). Dialogues for peer learning and cross border action: Improving cross-government data and information exchange and identifying good practices, Background report.

⁵ European Platform tackling undeclared work. (2017). 2017 Platform Survey report: organisational characteristics of enforcement bodies, measures adopted to tackle undeclared work, and the use of databases and digital tools.

- ▶ Taking steps towards introducing or improving data mining and the use of more complex data sharing techniques.

3.1 IT infrastructure and technical requirement challenges

Key challenges identified by the participants

- ▶ Lack of access to each other's databases across authorities
- ▶ Lack of knowledge and sometimes hardware and software for remote access to databases
- ▶ Silo culture and lack of budget, training, human resources
- ▶ Different inspection authorities rely on different unique identifiers, which hampers proper data matching and data mining
- ▶ The cost for contractors to implement interoperability is high
- ▶ The amount of work needed to be invested in coordinating with other authorities on what to exchange and how is considerable over time
- ▶ Continuous emerging technical issues
- ▶ Centralised systems do not automatically lead to more data matching capacities.

The participants reached the following conclusions:

- ▶ Although, there are countries like Belgium, which are very advanced legally and technically in data sharing, silos still constrain many other Member States. Typically, labour inspectorates have access to and develop their own internal databases only (e.g. Portugal, Greece, Latvia). The access to social security databases provides a significant boost in their understanding and capabilities to tackle undeclared work; however, the GDPR has rendered authorities to be more careful when sharing data and slowed down the speed of data exchange. With modest technical capabilities, organisations typically only receive data from other bodies on an informal basis through emails or phone calls. There is rarely remote access to databases even if it is technically possible (even on smartphones). The **biggest roadblocks** are the lack of human resources, training and budget, as well as awareness that data could help fulfil the organisation's mission.
- ▶ **Data strategies** are rare among labour inspectorates and not included in annual reports. Typically, tax authorities are the first to dispose of integrated database with access to different data resources. For most countries using artificial intelligence to collect and analyse data is a very long way to reach.
- ▶ **Only Belgium**, has an **integrated long-term data sharing strategy**. Its data infrastructure relies on the **DOLSIS platform**: the different institutions have their own databases, and through DOLSIS employees can access all the databases they are entitled to. For example, they can access the data of a person throughout all the provided government databases, based on personal or company ID. DOLSIS leaves a trace who checked the data and when, which is a critical component for data protection and trust. The DPO, citizens and companies can ask who had access to what database, and the institution must justify data handling.
- ▶ In 2021 Belgium introduced 'My Digital Information System (MDIA)', which allows using the DOLSIS platform on a **smartphone**, e.g. during an inspection of a construction site. Previously inspectors had to perform the checks on a personal computer only before or after the inspection, and then work on paper on site. With the MDIA this can be done on the spot. Yet, Internet access might be difficult with some remote rural sites.

- ▶ One of the technical issues on data sharing is having **different IDs for an individual or a company in different databases**. For posted workers for example, one might not have an ID number or it can differ, which might not allow to identify the person uniquely. Hence, Belgium introduced the LIMOSA database, which ascribes posted workers a BIS number (social security number). But sometimes doubles are created, e.g. when there are more names of foreigners, and the AI might attribute two different BIS numbers for every two names of the same person.
- ▶ DOLSIS has the labour and social security data **but not tax**. Belgium has a good legal framework, and social security staff are allowed to check with tax authorities and police, e.g., a criminal investigation-related information. On the other hand, tax authorities can use some modules of the DOLSIS platform but with DPO consent, who has to be able to check who entered the database, when, etc.
- ▶ Whenever, access is provided to DOLSIS, there is a **need for training** for the people who have not have access before.
- ▶ There are **limits to the number of data cells** inspectors can access through DOLSIS, e.g. in case of company with thousands of workers, export has to be provided by the IT company servicing the system.
- ▶ **Permission** for a specific external service to have access to DOLSIS is requested on a continuous basis. Whenever such authorisation is requested, it has to specify the number of staff and what kind of staff would have the access, not in terms of names of people but in terms of the specific functions they have, so that when the actual people change, the new recruits would still have access to the databases.
- ▶ The system has a **cadastre of enquiries**, allowing social Inspectors to see whether someone else from a different inspection service has been or currently is on inspection in the same site or company. This reduces the burden and allows synergies of inspections.
- ▶ The **COVID-19 crisis boosted financing for IT infrastructure and technical solutions**. For example, in France, like in many other countries, the State provided funds to the companies for keeping personnel employed throughout the crisis, and labour inspectors had to check if the companies indeed needed the money and used it correctly. Therefore, inspectors needed access to different company data to be able to match it. The existing strategy for data quality came into force, and also there was new impetus about data sharing and data matching with the COVID-19 crisis.
- ▶ Having undeclared work defined in the legislation as a **criminal offence**, like in France, can help open doors to access data. Also, more institutions outside the labour inspectorate can initiate an investigation into undeclared work cases, such as the gendarmerie and the judiciary. This calls for more data sharing and cooperation.
- ▶ Centralised employment data systems, like **the ERGANI in Greece**, push for more data sharing and access from different authorities. The system has established interoperability connections with systems of the tax and social security authorities to provide to employers several online services for declaring all their employment details before these are put in place. However, no steps have been taken in the field of data matching or data mining using the data of these authorities towards tackling undeclared work.
- ▶ The transfer to **the cloud** and maintenance of the environment requires a lot of financial resources and additional training. There are some specific protocols and procedures, which have to be followed, and learning them is time-consuming. In addition, the choice of **technological platform** locks in the authorities for the long-term and needs to be carefully considered to allow for the best balance between market flexibility and security of supply/upgrade. Work with framework contractors across the public administration might be useful in this respect.

3.2 Overcoming IT infrastructure and technical requirement challenges

Key solutions proposed by the participants

- ▶ Have established principles of interoperability, infrastructure and budget availability
- ▶ Bring inspectors and data scientists together
- ▶ Be patient and have a long-term strategy for capacity building – Belgium first started in 1990
- ▶ Ensure prioritisation, so that data miners are not swamped with requests for data analysis, which are not aligned with central priorities
- ▶ Use current crises as triggers for asking for higher tech budgets but think about providing cost-benefit and ROI justification
- ▶ Think about economies of scale: if a solution exists elsewhere in the public administration, explore it and use it if possible.

The participants reached the following conclusions:

The participants agreed, there are some **basic principles** for advancing IT solutions for data sharing: (i) interoperability; (ii) solid infrastructure; and (iii) budget availability.

Belgium presented **five principles** for managing its infrastructure and technical needs:

- ▶ *only ask principle*: you can only ask for something that is not available somewhere else; if you need information, which another institution got, then you need to have it;
 - ▶ *reuse principle*: e.g. if you did some technical and IT development for one institution, then you need to apply it others and not re-invent the wheel;
 - ▶ have general *common protocols* on development, maintenance, etc.; have interoperability keys between databases;
 - ▶ use *project management*;
 - ▶ *budget control*: think about cost-benefit, and about ROI, etc.
- ▶ Europe has also started to develop its **common protocols**. For example, BRIS is the EU's common enterprise number, a sort of European crossroads bank for enterprises. The BRIS could be an important building block for developing EU-wide solutions for secure data and information sharing, including in tackling cross-border undeclared work.
 - ▶ There is a **need for capacity building** to implement each of the principles. It is a process and it needs a lot of effort to be completed. And there is also a need to re-start the cycle with feedback again and again. So, in many countries at this point, there is only data matching on single cases of UDW or inspections.
 - ▶ The most difficult part in developing new technological solutions is the cooperation and knowledge sharing with data scientists. It is good to have data scientists in house and to have a team of inspectors working with them.

- ▶ There could also be (false) expectations that once there is a data mining department, every inspector could ask for any single case to be looked into by data scientists. Therefore, the data miners must reject some of the requests. There is a need for central decision making on the priorities, e.g., if social dumping is important, then the authority could define that 1 000 inspections would be carried out of which 1/3 would be based on data matching and data mining.
- ▶ One possible solution to budget constraints is to start asking whether other organisations exist in the country that already have the needed infrastructure, and to use **economies of scale** from other institutions, which already have such solutions.

3.3 Pilot testing of actions: key takeaways

This section showcases the agreed actions from the second day of the PLD and provides an overview of the additional insights gathered through them.

Presentation and discussions of the following actions:

- ▶ Find out about data mining in other authorities (do you employ data scientist, do you do modelling) (**Action: BE, EL, FR, LV, PT**) – future activity could be also to get insight from ELA analysis team.

Most models used by the **Belgian** social security authority are based on data-matching (not so much data mining), which allows it to find discrepancies, irregularities and undeclared work. Data matching requires specialised IT knowledge, which is both scarce and expensive on the market. Hence, the Belgian social security authority has signed a contract with an IT company for data mining. Six people in the IT team take a considerable share of that annual budget for training.

Greece also does not employ data mining and only very rudimentary data matching. Consultations with the social security authority have shown that they are combining data manually to produce some results and indications of risks. Approached tax authority was not willing to disclose if and what data matching they are doing. They could potentially open up if they are approached by someone they know personally.

In **Portugal** there is also no data mining and only limited data matching within the labour authorities. Information about the capacity of other administrations to do data matching and data mining is not readily available. The labour inspectorate has a technical team at its disposal but they are working on software and hardware support and are not engaged in data mining and modelling.

In **Latvia**, there are no data specialists in the labour inspectorate, but there are such at the tax administration. However, data scientists with the tax administration do not share information.

According to **Belgium**, in smaller EU countries with limited resources, only some institutions do data matching or data mining without sharing data with the rest of the administration. Such countries need an overall solution for the administration and more cooperation internally to make the best use of the limited available resources.

Presentation and discussions of the following actions:

- ▶ Assess guidelines for inspectors to use data and databases, what are the gaps, what works well, etc. (**Action: BE, EL, FR, LV, PT**).

Greece does not have guidelines for data mining but these could be very useful. The **data matching guidelines** in **Belgium** are prepared in collaboration between data scientists and inspectors. Every two months, data scientists

meet inspectors to finetune data models that could be used for better targeting and inspections. In addition, they are always on the lookout for solutions used by other administrations to replicate, adapt and reuse them if possible.

Sometimes, inspectors might not be interested in data mining, as they might erroneously perceive it as an additional assignment. Hence, the role of data models in easing inspection work needs to be carefully explained. For example, a data model might return 100 targets for inspection. But that does not necessarily mean inspectors have to visit all of them. Instead, they need to focus only on the highest risk ones.

An important element of ensuring data mining is not seen as a threat but as an opportunity in the society is **transparency**. For example, in Belgium the social security authority is obliged to disclose (i.e. on website) that it is engaged in data mining. The explanation provides information on how the authority process the data. Citizens and others can openly challenge data mining and data matching practices, if they feel their rights are breached. In Belgium data-matching has even been challenged in Parliament. An additional instrument for trust-building in society introduced in Belgium is the possibility for citizens to check online who from the public administration has looked into their records. Citizens can ask for a justification of such investigation.

Portugal reported not having guidelines for use of data and databases and the need for creating appetite among policy-makers for developing such. Similarly, Latvia reported that labour authorities face a similar situation with regard to guidelines.

Presentation and discussions of the following actions:

- ▶ Check if data strategy exists (e.g., in mission statement of own Ministry/organisation, in government, in e-government strategy) (**Action: BE, EL, FR, LV, PT**).

No guidance documentation on data strategies exists in **Greece**. There are specific manuals on software and applications but not on data use. However, the Greek Ministry of Digital Governance has established an Interoperability Centre, which has a strategy about data exchange and guides other administrations on the web services available throughout the public sector authorities. But separate administrations do not have data strategies.

The **Belgian** social security authority has a clear strategy on data use. The authority provides a lot of information on data sharing and data matching to its employees. Information is readily available on the organisation's intranet, and people working with data are well-trained on data handling and sharing. There are checks upon employees about why they looked into specific data records. Furthermore, different departments have their own specific requirements about data access, which also helps avoid unwanted mistakes.

Moreover, in Belgium, through annual appraisal, data exchange knowledge and training are actively promoted in career development. Every department and its staff have annual objectives to fulfil, including objectives on learning and skills. For example, anyone who has access to the intranet must follow compliance training on GDPR. Managers then regularly check whether employees follow the e-learning module. In addition, the human resources department has access to the information on GDPR training and they can also check if all people have not done the training.

4.0 REFLECTIONS AND SUGGESTIONS

The following commonalities emerge from the meetings:

- ▶ Most participants use data and data sharing on a national level on a case-by-case basis;
- ▶ They do not perform more advanced data operations, like data matching and data mining;
- ▶ Training on data handling and GDPR is key, as it provides the foundation for understanding the value of data and the legal constraints for its use; this way also a critical barrier to data sharing is overcome;
- ▶ There is a need to access national tax authorities' data across participating countries;
- ▶ A critical requirement for better data use and data sharing nationally is that policy-makers understand the return on investment and support legislative and practical steps to better usage of data.

Participants agreed on the following recommendations and suggested actions (1) which the Platform members can implement in their national practices and (2) which could be the subject of ELA future activities, including mutual learning at Platform Working Group level to prevent all authorities having to "reinvent the wheel".

4.1 Suggestions for national authorities

- ▶ Identify data needs and availability by reviewing and mapping existing data strategies or reports. Data audit can help establish state of play in data availability and data procedures.
- ▶ Meet other national authorities that have a vested interest in tackling undeclared work to discuss possible data sharing cooperation. This could be done by setting up multilateral working groups at national level on what kind of information labour inspectors need from other enforcement authorities and what kind of information they can share with them.
- ▶ Raise awareness and note return-on-investment to policy makers on better data gathering and data exchange opportunities. Think of practical examples showing the value of data sharing to spark interest among policy makers;
- ▶ Assess capacity needs/gaps for data sharing, and in what areas, and build a strategy on how to meet them;
- ▶ Focus on covering minimum data protection standards as guaranteed by GDPR;
- ▶ Think long term and work on a building political support and trust, legislative base, internal rules, capacity, training and IT infrastructure for data gathering and data exchange;
- ▶ Identify peers and work with national authorities and with ELA to reduce existing capacity gaps;
- ▶ In terms of countries, which could offer the best examples of data sharing, participants mentioned the following:
 - ▷ Belgium comes across as an advanced data gathering and data sharing country.
 - ▷ Finland has a central authority gathering data from many other authorities, which does data matching and data mining;
 - ▷ Spain's National Anti-Fraud Office is known to have access to a lot of information and they are doing some meaningful analytical work and data mining.

4.2 Suggestions for ELA

- ▶ The Platform could promote better understanding of data sharing benefits at an EU level. At a further stage, ELA could play an important role in the field of data analysis and exchange of information among relevant national level enforcement institutions. Upon requests of Member States, these activities could be linked to the harmonisation of data analysis practices, proposing common solutions to information and analytical work challenges. While conducting the analytical work, as mentioned at article 10 of the founding Regulation, ELA could examine existing data at national or EU level and report possible solutions to national administrations in order to overcome common labour mobility related obstacles.
- ▶ ELA could offer **GDPR compliance training** (including on how to make it effective on a national level and related issues of privacy), train the trainers how to apply GDPR, share good practices from different Member States and how these can be adapted. ELA would have to take note of the differences on a national level and know the specificities of each Member State. Also, ELA could support a common understanding in Member States that inspection services share data for the good public purpose.
- ▶ The best way for ELA to propose a train the trainers module in GDPR is to focus on **specific practical examples of data mining**. GDPR in itself is very legalistic, and it would be more beneficial if ELA training modules focus on applying actual data mining. This would help participants from different countries relate to a specific practical issue.
- ▶ ELA could raise **awareness on data mining and data-matching and the benefits associated with them**. ELA could showcase that there is a real return on investment from data mining, to justify investing in data analysis and related technical support, software applications, and continuous training. ELA could help focus Member States' authorities' attention on data sharing by organising and carrying out **data matching and data mining workshops**, for example by bringing together more and less advanced countries. In such workshops participants could share good practices and explain/showcase success stories with high return on investment for management level and raise interest in building data bases/systems at national level.

ELA could propose a model **multilateral workshop with tax and migration (and/or other law enforcement) authorities**, with the aim of bringing together authorities expected to benefit from data exchange both nationally and/or cross-border. This could be a first step in building trust across institutions. It can be a workshop among all or some Member States, at which good practices can be shared.

- ▶ ELA could invite Belgium, which is one of the pioneer Member State in implementing preventative approach enforcement measures and data analysis practices, to demonstrate an example with an experimental but real dataset, cleaned from privacy issues. Such a presentation should focus on more technical issues to show in numbers and with a specific open-source tool to demonstrate the added value of data mining. This will be very convincing for other countries to follow suit. Labour and social security authorities across the EU will be much more convinced and inspired by a demonstration for practical solution, a real-life system.

Some of the topics identified during the current PLD could be taken forward in future PLDs or other Platform formats:

- ▶ Integrating data and information exchange into strategic policy documents at labour inspectorate and national level. Developing a sound legal underpinning for data exchange at national level.
- ▶ Creating interoperable databases, which allow better tackling of undeclared work at national level. Developing possible cross-border elements.
- ▶ Data scraping and novel Internet tools for improving the fight against undeclared work.

- ▶ Building model GDPR compliance as a basis for effective exchange of data and information on national and cross-border level.
- ▶ Training on data management, e.g. workshop(s) on how inspectors could cooperate with data scientists and how to develop practice and strategy guidelines out of such cooperation.

5.0 BIBLIOGRAPHY

European Platform tackling undeclared work, (2017). Data Mining for More Efficient Enforcement. A Learning Resource Paper. Available at: <https://ec.europa.eu/social/BlobServlet?docId=18746&langId=en>

European Platform tackling undeclared work, (October 2017). 2017 Platform Survey report: organisational characteristics of enforcement bodies, measures adopted to tackle undeclared work, and the use of databases and digital tools. Available at: <https://ec.europa.eu/social/BlobServlet?docId=18747&langId=en>

European Platform tackling undeclared work, (September 2019). 2019 Annual Platform Survey: Tackling undeclared work in the collaborative economy and bogus self-employment, data exchange and data protection, and cross-border sanctions. Available at: <https://ec.europa.eu/social/BlobServlet?docId=21902&langId=en>