Paris, 8th October 2018

POSITION PAPER

KEY ISSUES FOR SUCCESSFUL PSI POLICIES IN EUROPE: IMPROVE THE ARTICULATION WITH GDPR AND ENHANCE DATA QUALITY

2013 PSI directive's French transposition has generated some side-effects that should be considered and corrected during the EU directive recast in order to make a real success of PSI re-use. Funding of public sector entities, Public Sector Information (PSI) including personal data, data quality: such key issues merit extensive debate and coordination within European institutions that should consider them as a high-priority while the PSI directive is being revised.

↗ Focal Points

→ The recast of the PSI directive is a great opportunity to deeply analyse the existing practical contradictions between personal data protection and PSI, otherwise the situation will become more problematic for all parties. **A better GDPR x PSI directive articulation is key for successful PSI policies**.

→ If the EU, Member States and public bodies want the PSI directive to become a real success, it must be clarified that providing PSI files, even free of charge, requiring reformating, data-cleaning etc., is not the right direction. In fact, **development of PSI re-use depends on a set of conditions:**

- Data quality maintenance
- Data dissemination platform
- Economic and financial balance
- Stakeholder's network and licensing

→ **It is essential to have a common definition of "data quality"**. Some sectoral quality standards could and should be developed in that perspective. This is especially important "to ensure the effective cross-border use of public sector documents", as stated in the recital 61 and the development of high value datasets.

About GFII

GFII – Groupement Français de l'Industrie de l'Information, is a French association dedicated to B to B Information and Knowledge.

The GFII is a central source of information, experience-sharing and action, though working groups and a dynamic social network. The GFII provides an opportunity for its members to meet, discuss and exchange points of view on the legal, technical and economic aspects of the sector. ***The wide variety of its membership coming from private and public sectors*** *means that the GFII is uniquely placed to facilitate the transfer of knowledge and professional skills. It updates and analyzes the information about the market in France and Europe, the emerging trends, the latest technological innovations, the evolution of the legislation…*

GFII – 17 rue Castagnary – 75015 Paris – 01 43 72 96 52
gfii@gfii.fr – http://www.gfii.fr
Contact : Margo Dessertenne, General Delegate / +33 (0)1 43 72 96 52 – dg@gfii.fr

Recasting and transposing the PSI directive without *precisely* considering transversally European, national, sectoral and cross-sectoral regulations like GDPR governing the different types of public bodies and the PSI they process and make publicly available, may generate some difficulties for public bodies and re-users. **Both are precisely facing some contradictions that are difficult to manage on both sides.**

The IMCO draft opinion mentions that "Member States should make all reasonable efforts to design policies governing the creation of data *so that their publication is already foreseen".* This principle called "open by design and by default" - a reference to the GDPR - appears as the right direction but the same draft opinion also mentions that **"making PSI accessible and re-usable creates cost for public administration"**.

This cost is not enough taken into consideration and we do not fully agree on the fact that "already today, the benefits outweigh the costs". A transitional period before entering into a full "free of charge principle" may be essential *for some* public bodies for guaranteeing the success of their open data developments. In that case, the fees should of course be "limited to the marginal costs incurred for their reproduction, provision and dissemination…"

↗ The re-use definition x personal data protection: discrepancies between GDPR and PSI directive

GFII is clearly among the 27% of respondents to the online consultation who do not agree that the PSI directive is well aligned with the rules on personal data protection. What if Public Sector documents include personal data?

Indeed, ""Re-use" means the use by persons or legal entities of documents held by public sector bodies, for commercial or non-commercial purposes **other than the initial purpose** within the public task for which the documents were produced." (Art 2.9 of the draft directive).

FIELD 1

Company and Business Registers could be considered as high value datasets. High value datasets may be compared to the "reference data" from the French 2016 "Digital Law". Please note that this idea is originally a GFII proposal of December 2014. Basically, these registers have been created to enable third parties to access publicly available information about their counterparts, customers or suppliers.

When making this data available as PSI in order to develop new services and increase business transparency, it should be first clarified that this kind of registers are "**of general public interest of the Union or of a Member State**, in particular a**n important economic or financial interest of the Union or of a Member State**, including monetary, budgetary and taxation a matters, public health and social security" (art 23-1 (e) of GDPR). As such, when provided as PSI, this data should not be pseudonymised or worse still, not provided at all anymore by public bodies. **But then, re-users must still be compliant with the GDPR and related national laws.**

Business Information Providers, whose job is to help businesses of all sectors / sizes to know more about their potential or existing customers and suppliers to better manage their cash and avoid late payments, process data on the whole population of French businesses. Such a purpose of re-use, through the indirect collection of (personal) data from PSI sources is, from our point of view, fully compatible with the initial purpose of collect / legal task carried out in the public interest or in the exercise of the public bodies.

At present time, an individual acting as a legal representative or sole entrepreneur can exercise its right to correction or right to be forgotten towards the re-user, even if his / her personal data must be filed to / disclosed by registers created to inform third parties.

In that case, the right to correction or to be forgotten **should be exercised to the Public Registers only, as these data should be considered as the official ones in terms of "paternity".** This is why the source and the last update must be provided to the end-user when data are re-used. Additionally, if such data had to be modified or deleted by the re-user, they would come back automatically in the re-user database, because of the updates from the public body.

In its opinion 06/2013 on open data and public sector information ('PSI') re-use, the WG29 had clearly identified some core difficulties:

- **Contradictions around the principle of open data vs the purpose limitation:** "information [ie PSI] should be available for re-use for innovative new products and services, and thus, for purposes that are not previously defined and cannot be clearly foreseen"; at the same time "purpose limitation is a key data protection principle and requires that personal data that have been collected for a specific purpose should not be further used for another, incompatible purpose" (§ 7-6 of the opinion).

- **Commercial vs non-commercial purposes:** "*What needs to be carefully assessed is whether the purposes and the way in which data are further processed are compatible with the initial purposes under the criteria mentioned in Section 7.6. In the case of PSI reuse this will inevitably lead to the consideration of a range of processing scenarios rather than just one."* (§ 7-7 of the opinion).

EXAMPLE: FRENCH ENTERPRISES REGISTER

This has not been taken into account at all at the present time and generated real case studies. As an example, data from the National Repertory of Enterprises managed by the French National Statistical Institute are available PSI for many years. Until 2017, this data was chargeable with different prices for commercial and non-commercial/internal re-use purposes.

PSI regarding sole entrepreneurs, who are also enterprises/businesses, with customers and suppliers, used to be provided with a "flag" signaling to the re-user that some of them had required that their data could not be re-used by third parties **for direct marketing purposes, in particular commercial direct marketing**, according to an article from the French Code of Commerce.

In other words, other categories of re-use (cf recital 41 and article 11 of the draft directive) or purposes of re-use used to be allowed.

From 2017, this data became free of charge according to an open license, for commercial and non-commercial purposes. Since then, these PSI on certain sole entrepreneurs that used to be flagged are not provided **at all** anymore.

Please note how different it is from what you can see on the concerned website of the National Statistical Institute where, when keying in the Unique Identifier you may know from other sources, you get a message telling "this enterprise exercised its opposition right. Information cannot be commercially available".

Conclusion: re-users wanting to develop services for other purposes than direct marketing cannot download a complete file anymore. In addition to this, "historical" re-users cannot provide the same service quality to their end-users than in the past.

The use case mentioned above highlights the two main issues which should be addressed:

1) Personal Data Protection

This "no item" solution is linked to rules applying in terms of personal data protection: a sole entrepreneur, acting in its business capacity **as an enterprise/business**, and having customers and suppliers just as any other business, is considered as an individual in its private/personal capacity, just as if he /she was a consumer "only". From our point of view, GDPR should apply when the individual is acting as a private consumer/individual **only, and not for sole entrepreneurs acting in their business capacity** as they are also an enterprise/a business.

Not to clarify this situation, as well as the one of legal representatives (company law)/beneficial owners (AML directives, anti-corruption laws etc.) of legal persons is generating/will generate difficulties for the concerned individuals, public sector bodies in charge of processing this data and making it available as PSI, re-users and their end-users, especially if Company and Business Registers should become high value datasets as proposed in the IMCO draft opinion.

Considering direct marketing, the "e-privacy" regulation should define the rules. But it is very important to keep in mind that direct marketing is **not** the only purpose of re-use for PSI including personal data.

2) Equal treatment between all re-users

This argument appears as contradictory with the concept of "comparable categories of re-use" mentioned in the article 11 of the draft directive. It is also difficult to understand for re-users having other re-use purposes than direct marketing, these other re-use purposes being de facto not taken into account.
Where is the equality of treatment?

Additionally, since data has to be provided free of charge by the public body, it may be costly for it to develop different files' contents to fit the needs from different comparable categories of re-use / purposes of re-use, which is detrimental to the objectives of the PSI directive.

FIELD 2

Case Law - that could also be considered as high value dataset and could generate important civic and democratic or socio-economic benefits, is another sensitive example. According to the 2016 French "Digital Law", court decisions should also be available as PSI, free of charge. One basic rule before making these PSI available is pseudonymisation. Even if the reasons that motivate this rule are easily understandable, some preliminary questions could be asked:

- Should Justice be secret? Is there no discrepancy between the "open court" principle (PDI) and pseudonymisation (GDPR)?
- What if non-anonymised judicial decisions can still be individually accessed to by contacting the court?
- How to make a constructive analysis of a decision if each data enabling to identify an individual or making him/her identifiable has disappeared, possibly making the decision impossible to understand?
- Is pseudonymisation for individuals only, of for legal persons too? It seems to us necessary to think about it when considering due diligences to make in terms of anti-corruption, AML etc.
- If this data had to be pseudonymised, it will be de facto difficult to combine it with other datasets, since there will be no common indexation key. At present time, the only way to "monitor" the same case all along its judicial path is … parties' names.
- Since national judicial systems and then case law are very different between Member States, how to develop metadata enabling an easy pan-European re-use?

→ **Availability of criminal decisions as PSI seems quite a delicate topic. Beginning with judicial decisions from lower civil courts, administrative and commercial courts could be a first step.**

Please note that case law is another telling example where the "comparable categories of re-use" concept should be taken into account. Legal professions (judges, lawyers, companies' legal advisers etc.) as well as legal

editors and LegalTechs should be authorized to access non-anonymised files. For end-users not having a legal profession, the result of the analysis and, as required by the GDPR, the content of the judgment would be pseudonymised.

→ **One of solutions that could be studied is that the public source provides a pseudonymised and a non-pseudonymised PSI files, according to the different categories of re-use / purposes of re-use.**

Discussions between legislators, DPAs, public bodies, re-users **and end-users** of the latter should then be encouraged in order to efficiently and concretely consider the particular case of each PSI dataset including personal data. GFII will be very pleased to participate and contribute to any workshop or expert group on that topic.

↗ Development of PSI re-use depends on a set of conditions

In 2012, GFII established a list of seven conditions to reach a satisfactory situation in terms of PSI re-use. These conditions are more or less promoted by the present PSI directive and the recast proposal. Based on GFII's experience, some issues still have to be addressed, independently of the personal data question commented here above, that may be summarised through the famous marketing "4Ps":

1- Product: data quality maintenance

Data quality means a set of conditions related to the data, the datasets and the database provided to the re-users: integrity, accuracy, granularity, completeness, consistency, existing metadata, existing specification and documentation, ready and easy-to-use formats, compliance with formal specification, freshness, adequacy to re-users' needs, provision conditions, possibility of feedback to the Public Sector Body in case of error detection, etc. Data must comply with this set of conditions as long as it is available as PSI. The data itself must be updated in numerous sectoral domains.

As an example of the "data quality" key issue, let us examine the use case of declarations to be made in France on the type of crops cultivated on each agricultural parcel, to comply with the EU agricultural policy.

Some years ago, penalties have been required to France because of a lack of precision in its calculation of agricultural areas. France finally resolved the case thanks to a solution provided by IGN (National Institute for Geographic and Forest Information). The main lesson from this use case is that data quality does not come "spontaneously" thanks to good intentions or from a purely administrative review. Data must be *managed* according to a specified and controlled technical process, reviewed by experts and delivered to the re-user/end-user. Of course, these actions have a significant cost and must be financed even if, at the end, data are free of charge for the PSI re-user.

2- Place: data dissemination platform (API…)

Data dissemination is segmented into numerous data repositories, uncomplete catalogues or demonstration websites. It is necessary to make an effort to harmonize the technical access to the data in a comprehensive way. The EU Members public data platforms have to turn a reality for all.

The article 6 of the draft directive provides that "re-use of documents shall be free of charge or limited to the marginal costs …" **But data processing and dissemination have a cost,** which has to be covered by other stakeholders than the re-users.

The remaining stakeholders are the data producers, frequently Public Authorities or Public Sector Bodies, data dissemination platforms, EU Member States and the EU itself.

Today, it remains unclear which stakeholder must finance the set of actions to be done to reach a real "open-data minded" organization, if the marginal costs cannot be considered.

One way to resolve this could be to develop projects at EU level enabling the creation of sectoral networks, mixing Public Bodies and re-users, able to precisely and practically examine the states of play, identify the issues and promote solutions. This kind of work has been done in the past for geospatial data in the framework of the INSPIRE directive and results are now very interesting.

3- Price: economic and financial balance and 4- Promotion: stakeholders' network and licensing

PSI re-use and promotion depends on numerous technical, human and financial conditions making any significant action a very hard way to walk. We also have to consider how PSI can de facto encourage GAFAM to develop their platforms at more or less no cost for them and a very high ROI.

In order to have an effective European Data Industry, as part of the Digital Single Market, it is necessary to enable Public Sectors Bodies and re-users to develop value-added services, the first ones in terms of data quality, the latter in terms of new services, independently of these global actors.

Licences should be "light" and avoid legal terminology as much as possible. GFII encourages very descriptive and practical licences, in order to enable all types of re-users, from big companies having legal departments to SMEs or an individual re-user, to easily understand their rights and duties. One way to do it is to provide examples of what should be done for being compliant. It is especially important when personal data are included in a public body's dataset.