**GFII Contribution on the proposal for**

**a Data Governance Act (01/02/2021)**

****

* Making “protected data” available for reuse should be mandatory, in compliance with DGA requirements
* The possibility to have different anonymization levels according to the end-users’ categories of the reuser and the purpose of reuse should be considered
* BtoB datasharing requires more than interoperability; it requires to improve all characteristics of holder's data (format, completion, update frequencies, documentation..)

**GFII Contact:**

17, rue Castagnary Anne-Christine Carpentier

75015 Paris Executive Officer

France Phone: +33 1 43 72 96 52

 Mail: ac.carpentier@gfii.fr

**1. Re-use of certain categories of protected data held by public sector bodies (chapter II of the draft DGA)**

 The idea of enabling the re-use of protected Public Sector Information looks appealing as it takes into account the purpose of re-use of data, which is very important from GFII’s point of view.

 1 - **Data value and re-use categories**

 As well mentioned in the “Strategy for Data”, “*the value of data lies in its use and re-use*”, which means that the public body should know why “its” data are re-used, and not only the number of downloaded files or the number of accounts opened to access APIs. Which is not the case at present time with “open” Public Sector Information, with advantages as well as inconveniences.

Among the latter, we can notice that one same and unique right, identical for all re-users, generates *de facto* a restrictive right for some of the use cases, which looks quite counterproductive according to the objective of development of the data economy.

There are different categories of re-users, whose economic and legal environments are different, which can / could justify that data could be made available differently.

Of course, rights and duties should remain equitable, practical and explicit and data should circulate in the EU as well as between sectors.

This is why we would consider pertinent to have different levels of access / re-use rights as well as different levels of legal rights and duties according to the purpose of re-use and reusers’ end-users.

**2 - No obligation on public sector bodies to allow re-use of data**

The point (3) of the article 3 provides that “*the provisions of [this] Chapter [II] do not create any obligation on public sector bodies to allow re-use of data* … ”, which is quite confusing.

Does it mean that there is no obligation to allow the re-use of existing protected datasets or does it mean that there is no obligation when / if the purpose of re-use does cannot be accepted?

The first interpretation would immediately reduce access to protected datasets; as a matter of fact, public bodies already have organizational, technical and financial difficulties to make their “open” Public Sector Information available for re-use according to high quality standards.

3 - **Conditions for allowing the reuse of protected datasets**

 Public sector bodies shall make publicly available the conditions for allowing the re-use. As it will probably be difficult for each public body to anticipate all potential purposes of reuse, the “candidate” should be able to contact it or the competent body described in article 7 to explain what its objectives are. Of course, the explanations should remain confidential.

**4 - Data catalogue**

In order to receive explicit and detailed requests for reuse, the public body will have to make publicly available the characteristics of its protected datasets (content / fields / technical options to access data …), i.e what we could call its “data catalogue”

**5 - Trial datasets**

As making available the data catalogue described here above will already require an investment from the public body, we suggest to first of all provide a free trial dataset and a description of all fields available. If there are expressions of interest from reusers (that should remain confidential), proving that there is an economic potential for one or more of the protected datasets, then, it will justify making costly developments for making data available with a high quality of service.

**6 - Anonymisation: means and organisation required**

The topic of anonymization / pseudonymization should be more precisely explained. We understand from the point (3) of the article 5 that the public body could require from the re-users that they anonymise / pseudonymize the data or suppress some data.

When the point 2 of the article (7) lets us understand that anonymization / pseudonymization will be done by the public body.

In both cases, we suggest considering the possibility to have different anonymization levels according to the end-users’ categories of the reuser and the purpose of reuse, that the public body will have to consider well in advance in the data production and distribution chain.

 Except some very well-structured databases enabling a fully automated anonymization, a semi-manual anonymization is the rule at present time, which generates costs that are too often minimized or ignored. An “over-anonymisation” can also be required *a posteriori*, that requires the development of very specific workflows.

As the existing automatic anonymization is quite “binary”, the result is quite difficult to use as a lot of contextual information necessary to the correct understanding of transferred data has been lost.

In order to achieve an anonymisation which is well adapted to the purpose of reuse, it is then necessary to have well-developed anonymisation processes which can embark:

- the tagging of data to be anonymised,

- programs enabling the management of differentiated anonymisation levels.

Before launching any industrialization project for data anonymisation, it is essential to have well in advance:

- defined processes,

- identified the different anonymization levels

- adjusted the anonymization algorithms

- made the algorithms “run”

- control the results against expectations

This conception phase is key, before launching the developments.

**7 - Verification of results of data processing undertaken by the reuser**

The point (5) of article 5 provides that:”*The public sector body shall be able to verify any results of processing of data undertaken by the re-user and reserve the right to prohibit the use of results that contain information jeopardising the rights and interests of third parties*”. According to what has been described here above, it will be necessary to explain what may be the terms of these controls.

**8 - Article 7**

We suggest having one competent body only in each Member State. Each public body intending to make protected data available for re-use should work first with this competent body, in order to capitalize on other experiences.

**9 - Article 8**

We support the idea of a single information point, which should be consulted by each re-user and where the data catalogue described here above should be available.

But granting access or not, which should be duly motivated, should remain in the hands of the public body holding the protected data.

Public resources, and not only fees, will be necessary for enabling public bodies as well as the competent body to make protected data available, as legal and technical requirements will be *de facto* stricter than the ones for open Public Sector Information.

Like the latter, the quality of data (all meanings of the word) that will be made available is key.

As a matter of fact, start-ups as well as businesses existing for a long time cannot develop projects or new services based on data if the availability continuity is not confirmed by the public body and the essential qualities are not maintained, not to say improved, including in terms of technical access.

**2 - Data sharing services (chapter III of the draft DGA)**

If noticed well, no mention is made on the fact that data sharing service providers will have to be remunerated. Maybe should it be specified that it is / will be a “commercial” service that will have to be paid by the data holder and/or re-user.

 We suggest specifying that it will be a regulated activity.

**1 -  Intermediation services between data holders who are legal persons and potential data users – BtoB data sharing**

 The GFII strongly supports this proposal. It is essential to recall that, for protected Public Sector Information as well as for BtoB data sharing, access and re-use of data by the re-user does not mean necessarily display “in clear” by the latter.

Data can feed algorithms without being displayed as raw data.

We agree that the activity of “data sharing service provider” should be separated from other activities, in order to avoid potential conflicts of interest.

 We understand from article 9 that the data sharing service provider will also have to provide technical means and skills to make holders’ data available.

From our point of view, it is essential. As a matter of fact,  making high quality data available for re-use requires very often some data cleaning / reformatting as well as the conception of the documentation (file structure, fields, format, definitions …).

That said, this work, that *must* be done, may generate a kind of loss of rights for the data holder as the data sharing service provider will have provided an essential technical value to the holder’s data “usability”.

That is why  requiring from the data sharing service provider to “convert the data into specific formats *only* to enhance interoperability within and across sectors or if requested by the data user or where mandated by Union law or to ensure harmonisation with international or European data standards” is quite limitative from our point of view.

The point (6) from the article 11 is key. No re-user will make developments without guarantees on that particular topic.

Please note that insolvency is not the only case that may be a “continuity breach”. If a dataset is not available anymore, if the maintenance is not well done, if APIs cannot manage all requests are other examples of potential “continuity breach”.

The point (7) from article 11 seems to us too concise according to the concrete requirements the data sharing service provider will have to comply with. These requirements should be more precisely described in order for the data sharing service provider not to underestimate them. They should also be very well explained in order for non-legal experts to perfectly understand them.

The point (10) of article 11 seems to be dedicated to data cooperatives and it should be specified;

 For data sharing services too, we suggest having one competent authority only in each Member State.

It will enable to have a complete view of the data sharing service providers and to work with other competent authorities like the DPA etc.

It will be necessary for this competent authority to have legal and technical expertise.

**2 - Data cooperatives**

The present definition provided in the point (c) of article 11 seems to be the one of a legal adviser.

If well understood, data cooperatives should be a kind of legal “go between” between data holders being data subjects, one-person companies or micro, small and medium-sized enterprises and data sharing service providers

Then, we suggest dedicating one article to data cooperatives.

**3. Data altruism (chapter IV of the draft DGA)**

The article 16 refers to the “general interest”, which – to our knowledge – is not defined. We believe the Commission should work on a more precised definition.

We understand that a public body and an association could be a data altruism organization. And that a private-for profit entity could possibly create a non-profit subsidiary. But how will it be financed? By some corporate sponsorship from the private-for profit entity? Thanks to financial donations, in addition to fees to be paid by natural or legal persons processing the data, if any ?

The idea is, of course, interesting but seems to us not solid enough from an economic point of view. As a matter of fact, before getting some potential fees, an initial investment will be required to create the data altruism organization.

In the article 18, the (c) from the point (2), will the anonymization / pseudonymization process have to be done by the data altruism organization or by the re-user?

 In that case too, we suggest having one competent authority by Member State.

 **4. The European Data Innovation Board** **(chapter VI of the draft DGA)**

 It is provided that:” *Stakeholders and relevant third parties may be invited to attend meetings of the Board and to participate in its work*.”

We consider that it should be mandatory and not an option.

As a matter of fact, it seems to us essential to include some representatives of data sharing service providers and data altruism organisations being in charge of the operational implementation of this regulation.

Of course, participation arrangements will have to be well defined.

**5. Extra-EU issues and digital sovereignty**

We are in line with the idea as the GFII has already alerted on the potential consequences of “open Public Sector Information” for the EU digital sovereignty.

It will be necessary to work on this at a very early stage in order to have precisely and easily understandable rules and not only for legal experts.

**About GFII**

GFII, created in 1979, is an independent and unique French organization of information professionals working in the BtoB data landscape **that brings together private and public data producers and/or reusers** such as the French Ministry of Interior, Bibliothèque Nationale de France, French chambers of Commerce, Orange Business Services, many legal and scientific publishers, Total, Crédit Agricole, Engie and SME as Starzdata, OpenDatasoft, Cairn-info,…

GFII aims to promote a sustainable and equitable economy of data/content. It acknowledges the important costs that implies manufacturing, maintenance, development and dissemination of data in an assumed commercial environment. It stresses the major need of interoperability of data in the process of valorisation and reusability power.

GFII promotes a sustainable and ethical use of data and works closely with lawyers, strategy experts, data-managers to bring the most expert and efficient feedback about the implementation of data policies.