1. Introduction

1.1. Context

The revised Directive on the re-use of public sector information (2013/37/EU) calls upon the European Commission to assist the Member States in implementing this Directive in a consistent way by issuing guidelines, particularly on recommended standard licences, datasets and charging for the re-use of documents, after consulting interested parties.

The present report analyses the feedback received in the period between 30.08.2013 and 22.11.2013 via an online survey, as a part of a public consultation on the future Commission guidelines on recommended standard licensing, datasets and charging for the re-use of public sector information.

In addition to the online survey, the consultation exercise included a public hearing open to all interested parties and a meeting of the Public Sector Information Group composed of the representatives of all EU Member States, held on 25-26th of November 2013 at the Commission's premises in Luxembourg. The reports from those meetings are available on the Commission's web pages1.

1.2. Objective:

The revision of the Directive on the re-use of public sector information (PSI Directive) is one of the key actions of the Digital Agenda for Europe. The rules on the re-use of public information were partly harmonised by the PSI Directive adopted in 2003. Nevertheless, the difficulties and uncertainties surrounding compliance with the licensing and charging provisions have since been often identified as the main persisting obstacles to wider re-use. This led to and adoption of a package of legislative and soft-law measures within the Commission's proposal to amend the original PSI Directive (COM(2011) 877 final).

As a consequence, and in line with Recital 36 of the revised PSI Directive, the objective of the consultation was to seek the views of stakeholders on specific issues to be addressed in the future Commission guidelines that should cover the three issues of importance for the promotion of wider PSI re-use in Europe: recommended standard licensing, datasets and charging for the re-use of public sector information.

2. Online Survey

2.1. Methodology:

The online survey was targeted at all parties interested in the re-use of PSI. This included primarily the public authorities, citizens, experts as well as professional PSI re-users. The questionnaire was prepared by DG CONNECT and delivered via the Commission’s standard interactive policy-making tool. It combined closed and open ended questions, where the respondents could further elaborate their opinions by providing free-text answers. The hyperlink to the consultation website was distributed by DG CONNECT through existing mailing lists with identified stakeholder partners (e.g. Public Sector Information Group, Member States Experts Group on Digitisation and Digital Preservation) and by stakeholder networking platforms (e.g. ePSI Platform, PSI Alliance). The recipients were invited to distribute the link further. The consultation was also advertised on the website of DG CONNECT, on the EU's consultation portal 'Your Voice in Europe' and at a number of events to which DG CONNECT representatives participated.

The analysis of the results of the public consultation was performed by DG CONNECT.

2.2. Profile of the respondents:

The consultation yielded 355 replies, composed of 304 online submissions and 51 offline questionnaires. In addition to this, 8 separate position papers or extended replies were received.

Whilst all five groups of stakeholders were well represented in the survey, the citizens and public authorities' categories were dominant (note: choosing more than one option was allowed).

![Respondents by categories](image)

The geographic coverage of the survey is satisfactory, representing respondents from all EU Member States (apart from Cyprus). 14 replies were submitted by entities from outside of the EU.
As the table below indicates, some member states (Austria, Bulgaria, Germany, Spain and United Kingdom) were particularly well represented in the survey.

<table>
<thead>
<tr>
<th>Country</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>50</td>
</tr>
<tr>
<td>Belgium</td>
<td>50</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>50</td>
</tr>
<tr>
<td>Cyprus</td>
<td>50</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>50</td>
</tr>
<tr>
<td>Denmark</td>
<td>50</td>
</tr>
<tr>
<td>Estonia</td>
<td>50</td>
</tr>
<tr>
<td>Finland</td>
<td>50</td>
</tr>
<tr>
<td>France</td>
<td>50</td>
</tr>
<tr>
<td>Germany</td>
<td>50</td>
</tr>
<tr>
<td>Greece</td>
<td>50</td>
</tr>
<tr>
<td>Hungary</td>
<td>50</td>
</tr>
<tr>
<td>Ireland</td>
<td>50</td>
</tr>
<tr>
<td>Italy</td>
<td>50</td>
</tr>
<tr>
<td>Lithuania</td>
<td>50</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>50</td>
</tr>
<tr>
<td>Malta</td>
<td>50</td>
</tr>
<tr>
<td>Netherlands</td>
<td>50</td>
</tr>
<tr>
<td>Poland</td>
<td>50</td>
</tr>
<tr>
<td>Portugal</td>
<td>50</td>
</tr>
<tr>
<td>Romania</td>
<td>50</td>
</tr>
<tr>
<td>Slovakia</td>
<td>50</td>
</tr>
<tr>
<td>Slovenia</td>
<td>50</td>
</tr>
<tr>
<td>Spain</td>
<td>50</td>
</tr>
<tr>
<td>Sweden</td>
<td>50</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>50</td>
</tr>
<tr>
<td>Iceland</td>
<td>50</td>
</tr>
<tr>
<td>Norway</td>
<td>50</td>
</tr>
<tr>
<td>Switzerland</td>
<td>50</td>
</tr>
</tbody>
</table>

3. Analysis of the results

3.1. Recommended datasets

In its Communication on Open Data published in December 2011 the European Commission called on the EU Member States to formulate and implement open data policies, which would include releasing a wide range of datasets likely to have the strongest socio-economic impact.

The overall purpose of the first part of the online consultation was to gather stakeholders' views on the categories of data that should be made openly available for re-use across Europe as a matter of priority.

The first question gathered feedback on the issue of the desired consequences of prioritising certain datasets ('core datasets'). The results show that the views on the matter appear to be evenly spread, with a slight majority of the respondents being in favour of machine-readability and guaranteed quality as the most important requirements that the core datasets should satisfy.

Some respondents suggested other possible consequences of qualifying datasets as 'core'. Most of these included a reference to the 'openness' of such data, for example a requirement to release under an open licence that respects the Open Definition or exempting the data from copyright protection by using a CCO dedication. Other group of comments focused on the quality and completeness of data, by pleading for a commitment to ensure regular updates of public information or to include rich metadata, as well as for guaranteeing the provenance of data, for instance by electronic signatures.
A closely related question is that of identifying certain datasets as 'core'. Hence, a question was included in the survey to gather feedback on the factors the national authorities should take into account when labelling certain data as 'core data'. The result shows an exceptionally even split of opinions between those favouring high value for commercial re-use and those opting for a non-commercial re-use objective. Apart from these two dominant groups of opinions, a relatively high number of submissions favoured other factors, such as 'usefulness of information', 'transparency', 'high value for use within the government', 'existing links to other sets of open data' and 'scientific as well as educational value'. Finally, some of the respondents expressed the view that all data produced by a body defined as a 'refined data monopoly' should be considered 'core' by default.
The last element of this section of the online questionnaire included a pre-defined list of 'high-value' datasets, based on the Technical Annex of the G8 Open Data Charter\(^2\). The respondents were asked to indicate which of those predefined categories could indeed be considered 'core' and therefore released with the highest priority. The results below show that there is a general consensus as to the inclusion of virtually all of the suggested datasets in the category of 'core data sets' so as to ensure their immediate availability. Some categories stand out as particularly relevant. These include above all geospatial data, closely followed by information in the field of transport, statistics, earth observation, environment and public finances.

![Core datasets chart]

A small minority of respondents chose to submit additional suggestions regarding the choice of themes for core datasets. Such suggestions covered various fields and included 'cultural heritage data', 'outcomes of publicly funded research projects', 'national registers of academic researchers', 'surveys and evaluations commissioned by public authorities', 'planning applications and land use', 'archival and bibliographic information'.

3.2. Licencing

This section of the survey was designed to gather input on the methods by which the public sector bodies should establish rules governing the re-use of data, on the choice of the possible conditions to be laid down and on other important aspects of licensing, such as interoperability and standards.

With regard to the fundamental question of a method that should be used to describe the applicable re-use conditions, a light-weight approach, limited to a mere disclaimer or

consisting of allowing the re-use of data without any particular restrictions was chosen by a vast majority of the respondents. The feedback received is visualised in the graph below:

When asked about the conditions that comply with the PSI Directive's requirement of 'not unnecessarily restricting possibilities for re-use', the biggest share of replies indicated the requirement to acknowledge the source of data. Slightly less popular was an obligation to acknowledge that the original data was not altered or the inclusion of a liability waiver. Overall, the respondents favour a reduced number of licensing conditions, generally those already suggested within the text of the PSI Directive. This is evidenced by a low number of suggestions for different conditions under 'other' category. At the same time however, the answers may also be interpreted as leaning towards the use of licensing conditions as a tool to counter the risk of the data being misrepresented or distorted.
Given the general objectives of the PSI Directive, which include the creation of an environment conducive to wider PSI re-use in Europe, it was important to gather the views of the stakeholders regarding the existence of provisions that should be avoided at all cost, due to their particularly strong role in discouraging or hampering re-use. The answers provided by the respondents to the two related questions on the topic are rather mixed. On the one hand, there is a generally low support towards a complete freedom in shaping the licences at the will of the licensor (i.e. high number of opinions confirming the existence of conditions that are particularly harmful to re-use). On the other hand, a sizeable share of the replies fell in the 'no opinion' category which indicates either a low interest of the respondents in replying to this question or their lack of understanding. In addition, only a small number of respondents chose to put forward examples of potentially 'black-listed' conditions.

Among the suggestions submitted, the following ones are most often mentioned: 'no derivative conditions', 'no redistribution', 'no commercial use', 'obligation not to distort the original meaning or message of the documents', 'any term restricting competition', 'limits on time over which data can be used', 'registration of (end) users'.

The apparent indecisiveness in relation to the existence of 'black-listed' conditions is not at all the case when it comes to the issue of existing standard licences. Here, we can note a dominant trend with the vast majority of the respondents answering in the affirmative to the question whether a greater interoperability would be best achieved through the use of standard licences, such as Creative Commons. This may be interpreted as both a high awareness of the availability of standard licences and a genuine understanding of their role in ensuring licencing interoperability across jurisdictions.
As a follow-up to this question, the respondents were asked to specify whether national licensing models, such as the UK's Open Government Licence would be preferable over other types of standard licences, notably 'Creative Commons'. In case of other preferences, an open comment box was provided for extra feedback. It was nevertheless seldom filled-in, as the overwhelming majority of respondents voiced their preference for 'Creative Commons' with only 16% percent favouring national licencing standards. Within the 13% of replies that found other licensing models to be of more use, the stakeholders mentioned primarily licencing schemes developed for particular thematic areas, such as geo-spatial information (German GeoLizenz) or metadata (ISA Open Metadata Licence), or expressed a wish for completely new licensing standards to be developed (pan-EU licence). Other suggestions included an Open Data Commons Database Licence, Flemish Open Data Licence as well as the French Open Data Licence.
Many respondents chose to further narrow down their preference for Creative Commons licencing models by indicating the type of CC licence most suitable for PSI re-use. The most quoted CC licence was CC0 and CC-BY, sometimes with a distinction ('CCO for datasets, CC-BY for text') or remark ('CC-BY only if CC0 cannot be used', 'CC BY 4.0 for datasets, possibly CC0 for metadata-as in the Europeana database'). Fewer respondents opted for CC-BY-SA, while some commented in more generic terms: 'The most open CC license could be used. But BEST OPTION is no use of any of license: public domain'

The next question referred to the practice of applying separate licences according to the purpose of re-use (commercial or non-commercial). Here, the predominant answer was negative, meaning that the majority of respondents were not in favour of drafting and using two different sets of licensing conditions solely on the basis of the distinction between commercial and non-commercial re-use.

Among the 29% of respondents who were in favour of applying such a distinction, many were of the view that additional conditions for 'commercial licences' might be needed. These choices suggest a wide array of circumstances in which such licences are already used, and a high degree of variation among the conditions applied across different institutions. The most relevant and representative choices of the additional conditions include: 'Shared revenue from the reuse, clear attribution, conditions on alterations', 'verification and control of data, data authentication', 'share of revenue of re-use with provider', 'previous approval of provider', 'possibility to publish partial results in another dataset with less information', 'ensure that the object cannot be altered or put under a different context', 'a time limit or delay post release on the use of datasets with commercial value in that a creator of the said data set has an interest', 'requirement of data holder to offer SLAs or certain guarantees of service for some data', 'disclosure of source code used to process the derivative dataset'.

---

**Is it justified to use different licences (commercial/non-commercial)?**

- Yes: 29%
- No: 60%
- No opinion: 11%
The final issue discussed in the section devoted to licensing was that of interoperability. Several possible characteristics of an interoperable licence were put forward and the respondents were asked to indicate which of those should be obligatorily used in licences to make them interoperable. The results, as visualised in the graph below, are rather evenly split among the four possibilities offered. Machine readability and common terminology were slightly more popular among the answers, although 'common sets of conditions' are not far behind. This suggests that most respondents are of an opinion that ensuring interoperability is a complex matter that can only be tackled by targeting various aspects of licencing at the same time, including the technical side (machine-readability) as well as cautious drafting (terminology, harmonised conditions). Some pertinent remarks were also provided on the licencing terms that would require a harmonised definition in order to ensure interoperability. The following terms were most often mentioned as candidates for harmonisation: 'public task', 'reasonable return on investment', 'anti-competitive', 'raw and refined data', 'manifestly unreasonable', 'attribution', 'derivative work', 'database', 're-usable format' and 'dataset'. The choice of terms shows the respondents' concerns about the fact that these key drafting components of licences (but also of re-use regulations in general) may cause interoperability problems if defined or interpreted differently.

In order to gather more detailed feedback on the measures that could be taken to ensure higher interoperability of licences across Europe, the respondents were asked to indicate whether other important issues should also be taken into account when preparing the future Commission guidelines. The input received shows that most participants to the survey had no particular opinion on that point or were outright sceptical about the existence of such issues.

However, among the 34% of affirmative answers, a number of specific additional measures were recommended. They mostly concern solutions for ensuring interoperability on a pan-EU
or even international levels: 'Single point national coordination for licenses or coordination at EU level', 'Domestic regulators must all have powers of sanction against PSIH's', 'EU coordination of licences and map/registry of licence conditions are needed', 'Publishing PSI as "public domain" would eliminate all these issues and be best!', 'Interoperability must go beyond the European to Global level!'.

3.3. Charging.

The last part of the online survey focused on the topic of charging for the re-use of public sector information, the rules for which are one of the key novelties of the recently adopted revised PSI Directive. Given the radical departure from the default charging rule based on cost recovery to that based on marginal cost of dissemination, the stakeholders' feedback was particularly important in relation to topics that are likely candidates for inclusion within the scope of the future guidelines. Those topics include the cost elements to be taken into account in the calculation of the marginal cost, cost elements that can be taken into account whenever a full-cost recovery is possible, the meaning and calculation of the 'reasonable return on investment' and the practical application of the charging rules with regard to the cultural institutions.

The first question in that area presented a long list of cost elements that might be taken into account when calculating the marginal cost of reproduction, provision and dissemination, in line with Article 6.1. of the Directive. Answers

The graph below presents the outcome of the online survey by focusing on the four possible replies: 'always', 'until amortised', 'never' and 'no opinion' (which proved to be very rarely chosen by the survey participants).

The main conclusions that could be drawn from the results of the survey on the topic of marginal cost charging are the following:

- The cost elements considered least appropriate for inclusion in the scope of 'marginal costs' (i.e. receiving the highest share of replies in the 'never' category) are those
related to archiving, data maintenance, data creation, database development, duplication/copying and value added services.

- The cost elements most readily accepted as a basis for calculation of marginal costs are those related to customer service, database modification, hardware enhancement and general hardware costs. However, for costs related to hardware and its enhancement, as well as for those needed to modify databases, a large proportion of replies made this inclusion dependent on staying below the threshold of amortisation.

- Very few respondents are of the view that other cost categories deserve to be included in the list. However, almost 40% of survey participants are of a different opinion and suggest the following cost elements to be added: 'administration cost', 'development tools to access information', 'data handling', 'long term preservation', 'digital asset management', 'verification and control of data' or 'provision of digital copies of analogue documents'. Finally, a number of submissions made it clear that discussion on cost elements is not all that important, since 'any genuinely “marginal” cost may be charged', or because 'data that is paid for by the public must be available to the public free of additional fees'.

![Cost elements - marginal cost charging](image-url)
Regarding the opinions voiced in relation to cost elements that should be taken into account in the cost-recovery method of charging, there appears to be a consensus concerning all four categories included in the consultation: overhead costs, non-incremental database development, non-incremental hardware costs and data maintenance. The most significant conclusion to draw from the results is that there is a general tendency not to include in the calculation of fees of the costs listed in the survey, with the majority of participants opting for 'never' or 'no opinion' as their reply on each of the sub-questions asked.

Furthermore, the stakeholders consulted clearly reject the idea of including any other potential costs, which were not listed in the survey. This reading of the results is confirmed by a high proportion of 'no opinion' and 'never' votes in the last column.

In the rare cases where 'other costs' were deemed necessary in the calculation of fees, the examples given included a wide range of costs, such as 'IT infrastructure investment', 'administration costs', 'metadata creation', 'website development', 'customer service', 'licencing', 'billing' or 'acquisition and preservation of documents'.

Apart from the crucial issue of deciding on the scope of the costs when calculating charges, public institutions may also be faced with a doubt as to the level at which charges should be calculated: item level, organisation level, database or other.

The survey included a question to that effect and the results show that the majority of respondents favour the calculation of fees on a level which takes into account whole groups of datasets (organisation, database) rather than individual documents. More than a half of all respondents chose these options, with some voices within the 'other' category calling for an even higher level, such as 'state level'.
An interesting outcome was yielded in reply to a question of the survey which intended to
gather views on the recommended scope of the term 'reasonable return on investment'. This
term has been included in the definition of the cost-recovery charging mechanism since the
entry into force of the PSI Directive in 2003. However, despite the fact that the term has been
in use for over a decade, there is a surprisingly large share of undecided answers, with 33%
declaring 'no opinion' on the topic. From the remaining answers, the pattern that can be
observed is that of confining the 'reasonable return on investment' to the lower ranges of
percentage points above the fixed interest rate on ECB's main refinancing operations, with
43% of all replies limiting it to less than 5% above the interest rate.
Among the replies suggesting different methods of calculating a reasonable return on investment, many included an outright rejection of such cost elements: 'It is unclear why ROI should be a factor in government operations', 'Zero - this should be not-for-profit' or pointed to the complex nature of charging for PSI in general: 'Depends on % each division contributes to service and costs involved', 'to be decided at the national/executive level'.

One of the last questions of the survey referred to the particular situation of the cultural institutions, for which the PSI Directive permits the inclusion of 'preservation and rights clearance costs' in the setting of their charging fees. The respondents were asked to list cost elements that should be considered when calculating such costs.

On the one hand, the answers recommended the inclusion of various costs items: 'security of originals', 'restoration/cleaning', 'access to originals/visits', 'acquisitions/sales', 'works for mandatory conformance to norms', 'property rights and contracts', 'development cost', 'personal costs', 'security costs', 'transports and insurance costs', 'digitization and copyright searches'. On the other hand, a significant proportion of opinions voiced dissatisfaction with the fact that the PSI Directive allows for any charging at all in relation to cultural heritage data: 'none - the costs of preservation and rights clearance should not be covered as they are not related to the re-use of PSI, but to libraries' mission', 'the information collected in those agencies should be available at no cost'.

Finally, in view of ensuring greater transparency, the survey tested the stakeholders' views on the necessity of publishing the revenues generated through the re-use of PSI. A vast majority of respondents clearly supported such a requirement, with an equally large share of opinions being in favour of an annual interval in the publishing of the generated income.

![Pie chart showing public opinion on whether public bodies should publish revenue from re-use.

<table>
<thead>
<tr>
<th>Option</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>86%</td>
</tr>
<tr>
<td>No</td>
<td>7%</td>
</tr>
<tr>
<td>No opinion</td>
<td>7%</td>
</tr>
</tbody>
</table>
4. Conclusion.

The results of the survey confirm the existence of a very rich landscape of PSI re-use in Europe, which is characterised by a multitude of actors and a corresponding multitude of views and attitudes. The fact that such a wide range of opinions could be found in the sample of stakeholders that took part in the survey makes it likely that the target audience of the survey was reached successfully and guarantees the relevance of the final results for the whole PSI re-use community. At the same time, the visible differences in interests and opinions that characterise the main groups of respondents (citizens, re-users, public sector bodies) make it extremely difficult to draw unanimous conclusions, which on many occasions have to rely on clusters of opinions rather than on any dominant tendencies.

Nevertheless, the picture that emerges from the survey is of a community increasingly characterised by expectations of an open and transparent system in which all the relevant parts cooperate rather than compete. Whilst it is understandable that the expectations of commercial re-users will be different to those of public data holders, there is an increasing trend in making sure that the rights of both sides are respected. It is for instance telling that there has been a wide acceptance of the new provisions of the revised PSI Directive, with few comments openly challenging the compromise reached as a result of a lengthy legislative process.

Hardly any controversies could be detected in the part of the survey devoted to datasets, where most respondents seem to support a wide release of rich sets of data with a view of stimulating commercial and non-commercial re-use alike.

There is also a widespread acceptance of the need to offer interoperable solutions, both on the technical and licencing levels. And even if opinions differ as to the exact shape of re-use conditions, the answers show that a general trend towards a more open and interoperable licencing system in Europe, largely based on available standard licences is gaining ground.

As could be expected, the part of the survey devoted to charging generated the most varied feedback, with the supporters of low charging being slightly better represented. At the same time, the more 'technical' questions, focused on the methods of fee calculation or the level of data to which they should be applied, showed a considerable level of hesitation and indecision. This suggests that the PSI re-use market in Europe is still under development and that a number of practices, procedures and mechanisms that could become standard one day, have not yet emerged.

All in all, a significant part of the feedback can serve as a good overview of the stakeholders' expectations towards the scope and content of the future Commission guidelines on PSI re-use. This is particularly true in the areas of datasets and licencing. Where the outcome is less clear-cut, the guidance of the Commission will have to strike the right balance between the views expressed on the one hand and the constraints of the revised PSI Directive underpinned by the wider principles of Open Data on the other.