ARTICLE 29 DATA PROTECTION WORKING PARTY

1021/00/EN
WP207

Opinion 06/2013 on open data and public sector information ('PSI') reuse

Adopted on 5 June 2013

THE WORKING PARTY ON THE PROTECTION OF INDIVIDUALS WITH REGARD TO THE PROCESSING OF PERSONAL DATA

set up by Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995,

having regard to Articles 29 and 30 paragraphs 1(a) and 3 of that Directive,

having regard to its Rules of Procedure,

HAS ADOPTED THE PRESENT OPINION:

I. Introduction

1.1. Revision of the PSI Directive

On 26 June 2013, the European Union adopted Directive 2013/37/EU of the European Parliament and of the Council (the 'PSI Amendment') amending Directive 2003/98/EC on the re-use of public sector information (the 'PSI Directive'). [1]

The PSI Directive aims at facilitating the re-use of public sector information by harmonizing the conditions for re-use across the European Union and removing unnecessary barriers to re-use in the internal market.

The initial 2003 text of the PSI Directive harmonized the conditions for re-use but did not require public sector bodies to make available data for re-use. The issue of whether to make data available for reuse was essentially optional: it was left to Member States and the public sector bodies concerned to decide. The result of this was that many public sector bodies across Europe simply chose not to allow their information to be re-used.

Against this background, one of the key policy objectives of the PSI Amendment is to introduce the principle that all public information (that is, all information held by the public sector, which is publicly accessible under national law) is reusable for both commercial and non-commercial purposes. Exceptions from the scope of the amended PSI Directive apply in certain cases, including on grounds of data protection.[2]

The amended PSI Directive thus now makes it mandatory for public sector bodies to allow reuse of all public information they hold. However, as will be shown below, it does not impose an obligation on public sector bodies to publicly disclose personal information. It only mandates reuse of information if it is already publicly accessible under national law, and even then only if reuse would not prejudice provisions of applicable data protection law.

Other relevant new provisions of the PSI Amendment expand the scope of the PSI Directive to include libraries (including university libraries), archives and museums.

In light of the foregoing, the amended PSI Directive has the potential to greatly increase accessibility of information held by public bodies.

1 OJ L 175, 27.6.2013, p. 1.

2 On the scope of the amended PSI Directive and the provisions relating to data protection, see Section V below.

1.2. PSI reuse and personal data

PSI reuse initiatives typically involve (i) making entire databases available (ii) in standardized electronic format (iii) to any applicant without any screening process, (iv) free of charge (or subject to limited fees), and (v) for any commercial or non-commercial purposes without conditions (or under non-restrictive conditions through a licence where appropriate)[3].

This may bring benefits leading to greater transparency and innovative re-use of public sector information. However, the resulting greater accessibility of information is not without risks.

To minimise these risks, wherever personal data are involved, data protection law must help guide the selection process of what personal data can or cannot be made available for reuse and what measures to take to safeguard personal data. In all cases where the protection of privacy and personal data is at stake, a balanced approach needs to be followed. On the one hand, rules for the protection of personal data should not constitute an undue barrier to the development of the re-use market. On the other hand, the right to the protection of personal data and the right to privacy must be respected. It is important to emphasise that as a concept the focus of open data is on transparency and accountability of public sector bodies, and economic growth, not on the transparency of individual citizens.

When applying the PSI Directive and data protection law to the reuse of personal data, a public sector body is likely to make one of three different types of decisions:

1. Decision not make personal information available for re-use under the terms of the PSI Directive
2. Decision to convert personal information into anonymised form (usually into aggregated statistical data)[4] and make only such anonymised data available for re-use
3. Decision to make personal information available for re-use (where necessary, subject to specific conditions and adequate safeguards)

II. Objective of the Opinion

2.1. Consistent guidance and best practice

The objective of this Opinion is to help ensure a common understanding on the applicable legal framework, and offer consistent guidance and best practice examples on how to implement the PSI Directive (as amended) with regard to the processing of personal data.

The aim of this Opinion is not to attempt to harmonise national approaches with regard to the level of transparency, national legislation on access to documents and the availability of information under such national legislation. However, national implementing legislation on the PSI Directive and national interpretation of Directive 95/46/EC[5] with regard to PSI reuse sometimes differ to a degree that goes beyond what may be necessary to cater for diversity in national access regimes and the different levels of transparency.

[3] Note that pursuant to Article 8(1) of the PSI Directive, as amended, license 'conditions shall not unnecessarily restrict possibilities for re-use and shall not be used to restrict competition'.

[4] On the reuse of aggregated and anonymised datasets derived from personal data, see Section VI below.

[5] Directive 95/46/EC of the European Parliament and of the Council of 24.10.1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (OJ L 281,23.11.1995, p. 31).

In this respect the September 2012 Policy Recommendations on Privacy prepared by the Thematic Network LAPSI clearly illustrates the unnecessary disparities in the manner in which the PSI Directive has been transposed in Member States with regard to the protection of personal data.[6] The PSI Directive itself also warns that legislative differences and uncertainties could become more significant with the further development of the information society, which has already greatly increased cross-border exploitation of information.[7]

The lack of a consistent approach may weaken the position of the individuals concerned. It may also pose unnecessary regulatory burdens for businesses and other organisations operating across borders and thus represent an obstacle to development of a common European market for re-use. On the one hand, data subjects must be reassured that their data will be consistently protected irrespective of their transfer to another Member State, for the purposes of re-use. On the other hand, undue complexity and fragmentation should be avoided also to enable free flow of personal data across Europe, another key objective of Directive 95/46/EC.

2.2. Need for an update of Opinion 7/2003

The PSI Amendment follows a decade after the adoption of the PSI Directive in 2003. At that time the WP29 adopted an opinion on the data protection concerns relating to PSI ('Opinion 7/2003')[8]. While the main principles outlined in Opinion 7/2003 remain sound, technological and other developments in the field of PSI and data protection, including the proposed legislative changes in both fields, justify the current efforts to update and complement the 2003 Opinion.

Further, the Opinion can now also take into account other recent and on-going efforts to provide further guidance, in particular:

- the 18 April 2012 Opinion of the European Data Protection Supervisor ('EDPS') on the Commission's Open Data Package[9],
- Opinion 3/2013 of the WP29 on purpose limitation;[10]
- the on-going work in the Technology Subgroup of the WP29 on anonymisation techniques[11];
- the work in some Member States on anonymisation and risk assessment;[12] and
- existing case law and practice on balancing re-use and personal data protection in some Member States.[13]

---

[6] LAPSI is a European Thematic Network on 'Legal Aspects of Public Sector Information', funded by the European Commission, see http://www.lapsi-project.eu/. The Policy Recommendation is available at http://www.lapsi-project.eu/lapsifiles/lapsi_privacy_policy.pdf.

[7] See Recital 7.

[8] See Opinion of the Article 29 Data Protection Working Party of 7/2003 on the re-use of public sector information and the protection of personal data - Striking the balance - adopted on 12 December 2003 (WP 83). See also two earlier, related Opinions of the WP29: Opinion 3/1999 on Public sector information and the protection of personal data adopted on 3 May 1999 (WP20) as well as Opinion 5/2001 concerning a European Ombudsman Special Report, adopted on 17 May 2001.

[9] EDPS Opinion of 18 April 2012 on the 'Open-Data Package' of the European Commission including a Proposal for a Directive amending Directive 2003/98/EC on re-use of public sector information (PSI), a Communication on Open Data and Commission Decision 2011/833/EU on the re-use of Commission documents. Available at: http://www.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2012/12-04-18_Open_data_EN.pdf.

[10] Opinion of the Article 29 Data Protection Working Party of 3/2013 on purpose limitation, adopted on 2 April 2013 (WP 203).

[11] An opinion on this subject is expected to be adopted in the second half of 2013.

[12] See, for example, the anonymization code of practice 'Anonymisation: Managing data protection risk code of practice' issued by the Information Commissioner's Office in the UK in November 2012 and the Risk analysis guidelines issued by the French data protection authority in June 2012.

III. Focus and structure of the Opinion

Opinion 7/2003 focused on the principle of purpose limitation[14], but also addressed other issues such as lawful grounds for public disclosure and re-use of PSI, the special protection provided for sensitive data, transfers to third countries, data quality, and data subjects' rights. These observations are still valid. Considering the previous work already done, this Opinion only updates and complements the conclusions of the Opinion 7/2003 when necessary in the light of new legislative and technological developments.

Section IV helps clarify that the obligation of reuse under the amended PSI Directive is without prejudice to data protection requirements and emphasises the importance of 'data protection by design and default' and 'data protection impact assessments' to help ensure that data protection concerns are addressed before personal data are made available for reuse.

Section V provides guidance, through illustrative examples, of what kind of personal data may come under the scope of the PSI Directive.

Section VI focuses on situations which are currently most common in PSI reuse initiatives: where aggregated statistical data, derived from personal data, are made available in aggregated and anonymised form. Examples include aggregated statistical data about crime rates, government spending, or about how well school children are doing in different geographical regions or in different educational institutions. This being the most common scenario of reuse of public sector information containing personal data, a significant portion of the Opinion will be dedicated to this scenario. The key data protection concern here is to ensure effective aggregation and anonymisation and minimise the risk that any personal data may be re-identified from the aggregated datasets.

Section VII - in less detail - discusses situations where personal data are made publicly available, and thus, may potentially be available for reuse. Although currently this is not the typical scenario for PSI reuse initiatives, it is important to consider that public sector bodies increasingly make personal data publicly available, often on the Internet. Here we are often talking about directly identifiable personal data such as, for example, information in a land registry about who owns a particular real estate, declarations of interests or salaries of certain civil servants or expenses of Members of Parliament. The question here arises to what extent, for what purposes, under which conditions and subject to what safeguards, these data may be available for reuse. It is also important to be clear on whether these data fall under the provisions of the PSI Directive.

In this context, it is important to highlight that any information relating to an identified or identifiable natural person, be it publicly available or not, constitutes personal data. Therefore, access and re-use of personal data that have been made publicly available (e.g. by publishing the data on the Internet) remain subject to applicable data protection law.

Some other specific scenarios, such as the case of research data, and the situation of historical archives - which now come under the scope of the PSI Directive - will be briefly addressed in Sections VIII and IX.

Section X discusses the issue of licensing PSI and the need to integrate a data protection clause into the licenses, whenever relevant.

Finally, Section XI provides a set of conclusions and recommendations.

---

[13] See, for example, LAPSI Policy Recommendation of September 2012 (pp. 4-14).
[14] See Article 6(1)(b) of Directive 95/46/EC.

IV. Not all 'publicly available' personal data should be made available for reuse

4.1. The obligation of reuse under the PSI Directive is without prejudice to data protection requirements

The PSI Directive, when adopted in 2003, did not impose an obligation on the public sector bodies to allow re-use of PSI. The decision whether or not to authorise re-use remained with the Member States or with the public sector body concerned (subject to the national regulatory framework on transparency and access). Opinion 7/2003 was adopted in the light of this 'non-obligation'. Section 2(cc) of Opinion 7/2003 provides that 'It is important to underline that the re-use Directive cannot be invoked as such a legal obligation that has to be complied with, because this Directive does not create an obligation to disclose personal information'.

With the PSI Amendment the analysis becomes more complex but the final conclusion remains the same.

Article 3(1) of the amended PSI Directive provides that '[s]ubject to paragraph 2 Member States shall ensure that documents to which this Directive applies in accordance with Article 1 shall be re-usable for commercial or non-commercial purposes in accordance with the conditions set out in Chapters III and IV.' Unless re-use can be denied for reasons provided in Article 1 (reasons derived from national access regimes and specifically also on grounds of protection of personal data), re-use must be allowed.

At the same time, recital 21 of the PSI Directive notes that the PSI 'Directive should be implemented and applied in full compliance with the principles relating to the protection of personal data'. Further, Article 1(4) provides that the PSI Directive 'leaves intact and in no way affects the level of protection of individuals with regard to the processing of personal data'.

These provisions, taken together, in a combined reading, mean that the 'principle of reuse' is not automatic when the right to the protection of personal data is at stake, and does not override applicable provisions of data protection law. When existing documents held by public sector bodies contain personal data, their reuse falls within the scope of application of Directive 95/46/EC and thus remains subject to the relevant data protection law.

Consequently, in cases in which the reuse includes personal data, the public sector body cannot systematically invoke the need to comply with the PSI Directive as a lawful ground for making available the data for reuse.[15]

4.2. Importance of a data protection impact assessment before opening up data for re-use

Considering the potential risks of PSI reuse, and in particular, the fact that once personal data have been made publicly available for reuse, it will be very difficult to effectively control the use of such data, the WP29 emphasizes the necessity of adhering to the principles of 'data protection by design and by default' and to ensure that data protection concerns are addressed at an early stage. In particular, the WP29 strongly recommends that a thorough data protection impact assessment should be carried out by the public sector body before it makes available personal data for reuse. Member States should also consider making such an impact assessment mandatory under national legislation or promoting it as a best practice. In any case, even if this is not expressly provided by national laws, prior to the disclosure of information and to the decision of making it available for re-use, public

[15] The WP29 also wishes to make it clear that from the perspective of the re-user, the PSI Directive in itself also does not create a lawful ground for processing. (On lawful grounds, see Opinion 7/2003 as well as Section 7.5 below.)

sector bodies should carry out a thorough assessment in order to establish whether personal data may be made available for reuse and if so, under what conditions and subject to what specific data protection safeguards reuse is permissible.

The assessment should, among others, establish a legal basis for the disclosure (and potential legal basis for reuse), assess the principles of purpose limitation, proportionality and data minimisation, and consider the special protection required for sensitive data. In carrying out this evaluation the potential impact on the data subjects should be carefully considered.

This assessment should help decide what, if any, personal data may be made available for re-use, and subject to what safeguards.[16] It is to be highlighted that the proposed Data Protection Regulation[17] encourages and in some cases requires data protection impact assessments as a key tool to help ensure the accountability of data controllers.[18]

Whenever possible, the analysis prior to the re-use decision should be based on an informed debate and the representation of diverse stakeholders, including the data controller wishing to release the data but also those demanding the data, and who therefore can provide context for the discussion, as well as the representatives of the individuals whose personal data are at stake (for example, consumer protection organisations, patients' rights organisations, teachers' unions). When the outcome is not clear, the competent data protection authority and the national authorities responsible for freedom of information may be able to offer guidance.

Member States should also consider establishing and providing support to knowledge networks/ centres of excellence and thereby enable sharing of good practice related to anonymisation and open data. These may be of particular importance for smaller public sector bodies that may lack the necessary expertise to carry out anonymisation, a data protection impact assessment and to assess and test the risks of re-identification.[19]

Finally, an impact assessment is also strongly recommended before new legislation is put in place that calls for public disclosure of personal data.

---

[16] In case the assessment leads to a decision not to make available for re-use personal data as such, but rather, to make available anonymised datasets derived from personal data, a re-identification risk assessment should be carried out. See Section VI on anonymisation and re-identification risk assessment.

[17] On 25 January 2012, the Commission adopted a package for reforming the European data protection framework. The package includes (i) a 'Communication' (COM(2012)9 final), (ii) a 'proposed Data Protection Regulation' (COM(2012)11 final), and (iii) a 'proposed Data Protection Directive' (COM(2012)10 final).

[18] For further guidance on how to carry out a data protection impact assessment, see, for example, the website of the PIAF project (A Privacy Impact Assessment Framework for data protection and privacy rights) at http://www.piafproject.eu/Index.html. PIAF is a European Commission co-funded project that aims to encourage the EU and its Member States to adopt a progressive privacy impact assessment policy as a means of addressing needs and challenges related to privacy and to the processing of personal data. Guidance is also available in some Member States. See, for example, the privacy impact assessment (PIA) handbook issued by the UK's Information Commissioner at http://ico.org.uk/for_organisations/data_protection/topic_guides/privacy_impact_assessment; the Risk analysis guidelines issued by the French data protection authority, already referred to in footnote 12 above, and the guidance provided by the Slovenian Information Commissioner, specifically on 'Privacy Impact Assessments in e-Government Projects', available at https://webmail.europarl.europa.eu/exchweb/bin/redir.asp?URL=https://www.ip-rs.si/fileadmin/user_upload/Pdf/smernice/PIASmernice__ENG_Lektorirano_10._6._2011.pdf

[19] As an example, in the United Kingdom a consortium led by the University of Manchester, with the University of Southampton, Office for National Statistics and the government's new Open Data Institute (ODI), runs the UK Anonymisation Network (UKAN) to enable sharing of good practice related to anonymisation, across the public and private sector. The network includes a website at https://webmail.europarl.europa.eu/exchweb/bin/redir.asp?URL=http://www.ukanon.net, case studies, clinics and seminars.

V. Scope of the PSI Directive: exceptions on grounds of protection of personal data

This Section provides guidance on the scope of the PSI Directive and in particular, on exceptions on grounds of data protection.

5.1. Applicability of the general data protection framework to PSI reuse

Recital 21 of the PSI Directive notes that the PSI 'Directive should be implemented and applied in full compliance with the principles relating to the protection of personal data'. Further, Article 1(4) provides that the PSI Directive 'leaves intact and in no way affects the level of protection of individuals with regard to the processing of personal data'.

5.2. Exceptions on grounds of protection of personal data

The PSI Directive provides that '[t]his Directive shall not apply to: … documents which are excluded from access by virtue of the access regimes in the Member States …'[20]

In addition, the PSI Directive, as amended, also provides exceptions on grounds of data protection. Article 1(2)(cc) addresses the following three situations, all three of which are excluded from the scope of the PSI Directive:

- documents access to which is excluded by virtue of the access regimes on the grounds of protection of personal data;
- documents access to which is restricted by virtue of the access regimes on the grounds of protection of personal data, and
- 'parts of documents accessible by virtue of those regimes which contain personal data the re-use of which has been defined by law as being incompatible with the law concerning the protection of individuals with regard to the processing of personal data'.

5.3. General comments

The WP29 underlines that irrespective of the 'principle of reuse' formulated in the PSI Amendment, reuse for any commercial or non-commercial purposes under the terms of the PSI Directive is not always appropriate in cases when the PSI to be reused contains personal data. Decisions regarding reuse of personal data under the terms of the PSI Directive will need to be made on a case-by-case basis and there is also a need to put in place additional legal, technical or organisational measures to protect the individuals concerned.

The reuse of publicly available personal data is and should be limited by

- general provisions of applicable data protection law,
- (where applicable) specific additional legal restrictions and
- technical and organisational safeguards that have been put in place to protect the personal data.

5.4. Documents access to which is excluded

This provision excludes from the scope of the PSI Directive all documents that are excluded by virtue of the access regimes of the Member State concerned on the grounds of protection of personal data.

[20] See PSI Directive, Article 1(2)(c).

Unlike data protection laws, which are harmonised to a large degree based on Directive 95/46/EC, access to information laws significantly diverge across EU Member States. Access regimes typically call for a balancing test that compares the interests protected by privacy and data protection rules against the benefits of openness and transparency. Considering the divergences, the outcome of the balancing exercise may be different in the different EU Member States. For example, tax authorities is some Member States may publish certain parts of the income tax declarations of taxpayers (subject to legal, technical and organisational measures to minimise risks of misuse), whereas another Member State would consider this as information that would fall under the exception and should, in general, be kept private.

That being said, national legislation must comply with Article 8 of the European Convention on Human Rights ('ECHR') and Articles 7 and 8 of the Charter of Fundamental Rights of the European Union ('EU Charter'). This implies, as the European Court of Justice held in the *Österreichischer Rundfunk* and *Schecke* rulings[21], that it should be ascertained that the disclosure is necessary for and proportionate to the legitimate aim pursued by the law.

In any event, once the personal data contained in a document is excluded from access under the laws of the relevant Member State (including situations where the national legislation on transparency and openness do not provide for the general accessibility of the personal data concerned), it will also be excluded from the scope of the PSI Directive.

To ensure legal certainty and transparency towards the data subjects, it is good practice, whenever possible, to take a proactive approach and define in advance the personal data that could be made publicly available. Data subjects can then be informed, at the time of collection of the data, whether any part of the personal data they provide, or that will be further processed during the administrative procedure, will become publicly available as a result of freedom of information laws.

5.5. Documents access to which is restricted

This provision excludes from the scope of the PSI Directive all documents access to which is restricted by virtue of the access regimes of the Member State concerned on the grounds of protection of personal data. Again, access regimes in the different Member States may vary as to what data may be subject to restricted access and what type of restrictions there may be. Some examples are as follows:

- Collections of national archives containing personal data that are accessible only subject to specific conditions of access, and additional safeguards (see Section IX below).
- Collections of research data containing personal data that are accessible only subject to specific conditions of access, and additional safeguards (see Section VIII below),
- Certain information in public registers, court files, or other administrative documents containing personal data that may be accessible only to individuals or organisations who show a legitimate interest, or only subject to other specific conditions of access, and additional safeguards.

5.6. Parts of documents accessible but where reuse is incompatible

This provision excludes from the scope of the PSI Directive

[21] See ECJ 20 May 2003, Rundfunk, Joined Cases C-465/00, C-138/01 and C-139/01 and ECJ 9 November 2010, Volker und Markus Schecke, Joined Cases C-92/09 and C-93/09.

- parts of documents
- accessible by virtue of national access regimes
- which contain personal data 'the re-use of which has been defined by law as being incompatible with the law concerning the protection of individuals with regard to the processing of personal data'.

This provision confirms that even in cases where certain documents containing personal data are fully accessible, their reuse may nevertheless be restricted on data protection grounds.

The WP29 emphasises that this provision in the PSI Directive should be interpreted in accordance with Article 1(4) of the PSI Directive, which declares that the PSI Directive 'leaves intact and in no way affects the level of protection of individuals with regard to the processing of personal data'.

The WP29 would welcome as a good practice the adoption of specific legal provisions in national law that clearly describe (i) what data are made publicly available, (ii) for what purposes, and (iii) where appropriate specify to what extent and under which conditions reuse is allowed. However, when such specific provisions are not in place, this does not mean that publicly available personal data can always be reused under the PSI Directive.

Instead, in these cases data protection law (applied together with other relevant law such as access to documents legislation) determines whether personal data may be made available for reuse in the specific case, and if so, subject to what additional safeguards. If the outcome of this assessment is positive, reuse is authorised, subject to specific data protection safeguards and all other conditions set forth in the PSI Directive (so long as they are without prejudice to data protection law). If the outcome of the assessment is negative, reuse will be outside the scope of the PSI Directive.

The following examples may help illustrate when this exception from the scope of the PSI Directive may apply. In the first example, restrictions on reuse are clearly specified in law.

- Tax laws in a Member State may provide that the income tax declarations of all residents of the country are publicly available for review by any other resident on request on the premises of the tax authorities, without the need to show legitimate interest. The law also clearly specifies that the data cannot be further processed, for example, published on the Internet, combined with other data, or redacted further. An NGO requests access and the right to re-use the tax declaration database in order to publish them on their website. In this case the tax data are outside of the scope of the PSI Directive and there is no obligation on the public sector body to make the dataset available for re-use under the PSI Directive.

In many other cases, however, legal restrictions are likely to be less clearly expressed and less categorical in terms of reuse. Typically, various civil, commercial and population registers, and other databases allow consultation of personal data by the public, increasingly in digital form via the Internet. Accessibility is often subject to specific safeguards, including technical restrictions on search capabilities and bulk download. Users may also be asked to agree to terms and conditions for access.

- Tax laws in a Member State may provide that the names of those residents that have had tax arrears over a certain threshold for an extended period of time are published on a dedicated internet site, for a limited period of time, subject to additional technical safeguards including limitations on bulk download and search capabilities. The purpose of this publication is to encourage timely payment of income tax and to serve as an additional (reputational)

punishment for those who fail to do so. A consortium of banks requests access for reuse to feed the data into their credit reporting system.

- Specific laws in the health care sector in a Member State may allow, subject to safeguards, for patients to verify, on a dedicated website, whether a particular doctor or other professional is banned from practice. Technical safeguards apply, including limitations on bulk download and search capabilities. A patients' rights organisation seeks access for reuse in view of creating a multi-lingual and more user-friendly website to access the same data.
- Specific laws in a Member State may require publication of names of donors to political parties over a certain threshold. The information which can reveal the donors' political opinions is made public via a dedicated website. Technical safeguards apply, including limitations on bulk download and search capabilities. An activist group requests access to the data in bulk for reuse under the PSI Directive with a view to create a new website with additional features and better search capabilities.
- The name and address of the owner of a real estate is public in the land registry of a Member State, but browsing in the publicly accessible database is limited so that only searching for a certain real estate is possible and not for a certain individual. Bulk download is also limited. A commercial company requests access to the data in bulk for reuse with a view to create a more user-friendly website at a more competitive price.
- Business registers in a Member State allow public access to a broad range of personal data, including names, addresses and specimen signatures of directors, and information regarding the ownership of certain types of companies. There are some restrictions on search capabilities and limits on the number of items that can be downloaded. The information is available via a dedicated Internet site and subject to payment of a fee. A commercial company requests access to the data in bulk for reuse with a view to create a website that combines information from several different types of registers and to offer enhanced information at a more competitive price.

In all cases, the public sector body concerned must make a careful data protection impact assessment to decide whether the data may be made available for reuse under the PSI Directive and if so, whether data protection law requires any specific conditions and safeguards. The 'principle of reuse' is not automatic, and cannot override applicable provisions of data protection law.

This careful assessment is all the more important as under the PSI Directive, the public sector body, in principle, must not consider who the particular re-user requesting access is. Pursuant to Article 10 (Non-discrimination), 'any applicable conditions for the re-use of documents shall be non-discriminatory for comparable categories of re-use'. Further, pursuant to Article 11 (Prohibition of exclusive arrangements), 'the re-use of documents shall be open to all potential actors in the market …. Contracts or other arrangements between the public sector bodies holding the documents and third parties shall not grant exclusive rights.'

Therefore, when deciding whether or not to authorize reuse, public sector bodies must consider the compatibility of allowing re-use under an open licence not just to the applicant but also to anyone requesting the data. This requires a high level of confidence that none of the potential re-users will be able to misuse the personal data made available.

The PSI Directive does not exclude that the terms and conditions could authorise processing only for specific purposes. The question for the public sector body is then whether re-use, by any 'potential actor in the market', for these purposes, is compatible with the purposes specified by the public sector body. The potential re-use of tax payment information by financial institutions, for example, for credit reporting purposes, is relevant as they are still a potential re-user, under the 'any person'

test. Therefore, to meet data protection concerns, in particular, to ensure that the principle of purpose limitation is adhered to, the public sector body (or the legislator) must be allowed to limit, where relevant, the purposes of reuse.

VI. Reuse of aggregated and anonymised datasets derived from personal data

6.1. What are the benefits of aggregation and anonymisation for the re-use of PSI?

To date, PSI reuse initiatives launched by public sector bodies through 'open data portals' or other platforms typically have aimed at making aggregated and anonymised data available for reuse, rather than personal data as such. This approach is indeed safer and should be encouraged.

Data protection laws usually do not allow that public sector bodies publicly disclose personal data collected for another, usually administrative, purpose[22]. Thus, in these cases, their reuse as part of PSI reuse initiatives is also not possible. Rather than personal data, it is typically statistical data derived from personal data that are and that should - in principle - be made available for reuse. This is the most effective solution to minimize the risks of inadvertent disclosure of personal data. These anonymised and aggregated datasets should not allow re-identification of individuals and thus, should not contain personal data.

Deciding what level of aggregation may be appropriate and what specific anonymisation techniques to use is a challenging task. If aggregation and anonymisation are not done effectively, this carries the risk that individuals may nevertheless be re-identified from these datasets. Therefore, data protection law has an important role to play in helping to determine the threshold at which it is 'safe' to release anonymised and aggregated data as part of a PSI initiative.

Directive 95/46/EC sets a high level of threshold for anonymisation

When used in this document the term 'anonymisation' refers to data that can no longer be considered personal data under Article 2(a) of Directive 95/46/EC. Article 2(a) defines 'personal data' as 'any information relating to an identified or identifiable natural person ("data subject"). An identifiable person is 'one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, psychological, mental, economic, cultural or social identity'.[23]

Recital 26 of Directive 95/46/EC is also relevant and further provides that in order 'to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person'.

It must be emphasised that this sets a high threshold, as will be discussed further in this Opinion. Unless data can be anonymised to meet this threshold, data protection law continues to apply. This means, among others, that unless the threshold is met, the public release of the information (and any further use) must be 'compatible' with the initial purposes of data collection under Article 6(1)(b) of the Directive 95/46/EC. In addition, there must also be an appropriate legal basis for the processing

[22] Of course, where applicable, freedom of information legislation may require disclosure of personal data, and the interest in transparency and the availability of information in some situations may override data protection and privacy concerns. This is an evolving area that may bring future changes.

[23] In its statement made on 27 February 2013 on 'current discussions on the data protection reform package', the WP29 emphasised that 'a natural person can be considered identifiable when, within a group of persons, (s)he can be distinguished from others and consequently be treated differently. This means that the notion of identifiability includes singling out'. The statement also clarifies that 'identification numbers, location data, IP-addresses, online identifiers or other specific factors relating to an individual should be considered personal data'.

under Article 7(a) through (f) of Directive 95/46/EC (for example, consent, or necessity to comply with the law). In contrast, if the data have been anonymised in the meaning of Article 2(a) and recital 26 of Directive 95/46/EC, data protection rules will no longer apply and re-users may be able to reuse the data without these constraints.

Once again, it is to be emphasised that 'anonymised data', as used in this Opinion, refer to data that are no longer considered personal data. Anonymised data should, in particular, be distinguished from data, that have been manipulated using various techniques to mitigate risks of re-identification of the individuals concerned, but have not attained the threshold required by Article 2(a) and recital 26 of Directive 95/46/EC.[24] In many scenarios these techniques are only appropriate for limited disclosure for re-use by screened third parties but not full public disclosure and re-use under open licence.

It is also important to emphasise that once data are publicly released for reuse, there will be no control over who can access to the data. The likelihood that 'any other person' will have the means and will use those means to re-identify the data subjects will increase very significantly. Therefore, and irrespective of the interpretation of recital 26 in other contexts, when it comes to making data available for reuse under the PSI Directive, the WP29 wishes to make it absolutely clear that utmost care should be taken to ensure that the datasets to be disclosed should not include data that can be re-identified by means likely reasonably to be used by any person, including potential re-users, but also other parties that may have an interest in obtaining the data, including law enforcement.

Further guidance on anonymisation and the concept of personal data

For further guidance on anonymisation and the concept of personal data, see Opinion 4/2007 of the WP29 on the concept of personal data, adopted on 20 June 2007 (WP 136). The WP29 may also provide further guidance on anonymisation techniques in a separate document in the second half of 2013.

6.2. What are the challenges and limits of anonymisation for the re-use of PSI?

Anonymisation is increasingly difficult to achieve with the advance of modern computer technology and the ubiquitous availability of information. Re-identification of individuals is an increasingly common and present threat.[25] In practice, there is a very significant grey area, where a data controller releasing the data might believe a dataset is anonymised, but a third party may still be able to identify at least some of the individuals from the data, for example, by using other publicly available information, or other information that is available to him/her.

One of the major risk factors is the increasing amount of online and offline data, both publicly available and concentrated in the hands of business organisations, which then can be used for

[24] The 27 February 2013 statement emphasises that 'where it is possible to backtrack an individual or (indirectly) identify an individual by other means, data protection rules continue to apply.'

[25] See, for example, 'Transparent Government, Not transparent Citizens', a report prepared for the UK Cabinet office by Kieron O'Hara of Southampton University in 2011, in which the author warned of the ability to identify individuals from anonymised data, using, among others, 'jigsaw identification' and saying that there are no complete technical solutions to the de-anonymisation problem. Available at: http://www.cabinetoffice.gov.uk/sites/default/files/resources/transparency-and-privacy-review-annex-b.pdf See also Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization by Paul Ohm of University of Colorado Law School, 57 UCLA Law Review 1701 (2010), available online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006

profiling individuals for behavioural advertising and for an increasing array of other purposes. When matched against the realities of 'big data' already available to these organisations, PSI derived from personal data and made available for reuse could increase the likelihood that individuals could now be identified or that their profiles can be further enriched, often without the individuals being aware that this is happening.

6.3. Who should carry out the aggregation and anonymization and when?

Aggregation and anonymisation should occur at the earliest opportunity – by the data controller or by a trusted third party acting on behalf of the controller or several controllers (and who is also in possession of the necessary specialized skills). It cannot be left to the re-user to carry out the anonymisation, for example as a licensing condition. Further, it is important to ensure that the possible third party organisation carrying out the aggregation and anonymisation has no conflict of interest and is clearly held accountable that the personal data will only be used to carry out the anonymisation and that all the necessary safeguards are put in place to this effect. The third party should also be able to guarantee that the personal data from which the aggregated and anonymised datasets are derived should be deleted as soon as they are no longer required for that purpose.

6.4. Assessing the risks of re-identification

Unless data can be anonymised in the meaning of Article 2(a) and recital 26 of Directive 95/46/EC, data protection law continues to apply.

Controllers should assess whether an individual can be reasonably identified from the 'anonymised' dataset to be made available for re-use and from other data. In other words whether any organisation or individual could identify any individual from the data being released – either in itself or in combination with other available information.

As explained in Section 6.1, this Opinion does not aim to provide comprehensive and conclusive guidance on how to assess the risks of re-identification. Neither does it aim to provide a conclusive definition for 'anonymisation' or 'anonymised data'. However, it reiterates that the reader may find further guidance in existing documents (including those referred to in Section 6.1) and there is also work on-going in the Technology Subgroup of the WP29 on anonymisation techniques, as noted in Section 6.1 and Section 2.2.

That said, and without aiming for comprehensiveness, the WP29 wishes to highlight some of the factors/concepts that are helpful to consider when assessing the risks of re-identification, including, in particular:

- what other data are available, either to the public at large, or to other individuals or organisations; and whether the data to be published could be linked to other datasets;
- the likelihood of re-identification being attempted (some types of data will be more attractive to potential intruders than others); and
- the likelihood that the re-identification, if attempted, would be successful, considering the effectiveness of the anonymisation techniques proposed[26].

[26] On anonymisation techniques, see upcoming Opinion of the WP29 on this specific subject.

What 'other' information is out there?

When determining whether an individual may be indirectly identified, it must be considered whether identification is possible using the data in question (in our case the 'anonymised' dataset), or from those data and *other information* which is in the possession of, or may/likely to come into the possession of, the organisation or individual that attempts re-identification.

The 'other information' needed to perform re-identification could be information available to certain businesses or other organisations including law enforcement authorities or other public sector bodies, to certain individuals or to everyone because it has been published on the Internet, for example. An obvious example is where publicly available data – such as the electoral roll, telephone book or other data easily retrievable from a web-search – can be combined with the (inadequately) 'anonymised' data, allowing an individual to be identified (e.g. using his/her birth-date and postcode).

Re-identification risks can increase where one individual or group of individuals already knows a great deal about another individual, for example, a family member, a colleague, contact on a social networking site, a doctor, teacher, a law enforcement agent, or another professional.

What matters here, however, is not simply whether the individual with prior knowledge can identify the data subject concerned but whether he/she will learn something new from the information obtained through re-identification. The two examples below will illustrate the importance of this distinction.

Example one: measles statistics. In one case, anonymised statistical data may reveal that in town A in the year 2012 X number of people contracted measles. No further breakdown or information is given. A doctor who contributed to the statistics by providing information about his own patients to the relevant health authorities holds more complete records of these patients in his office, subject to medical confidentiality. The doctor would be able to easily re-identify several of the patients from the statistical dataset. Similarly, a mother who knew her child contracted measles that year could easily re-identify her child in the dataset. Nevertheless, neither the mother nor the doctor would learn anything that they have not known before from the anonymised dataset that has been made publicly available.

Example two: drug and alcohol abuse, sexual abuse, and school performance. This example can be contrasted with the following. Research is conducted into correlations between drug and alcohol abuse of parents, sexual abuse of children, and school performance. Purportedly 'anonymised' research data are published with good intentions but without a careful assessment of re-identification risks.

The statistics reveal, among others, that at School A, where a total of 500 pupils are enrolled, in the year 2012 20 % of pupils (100 pupils) lived in a household where at least one parent is an alcoholic or a drug addict. Of these, in 8 % of the cases (8 pupils) the child has been sexually abused. The report also specifies that no other pupils were sexually abused at School A.

The figures also show that in 96 % of cases (96 pupils) the children whose parents were alcoholic or drug-addicts significantly struggled with their school performance ('poor performers' defined by an appropriate academic standard), however, at this particular school only 50% of those sexually abused (4 pupils) had significant difficulties with school work.

At the school it is common knowledge that AA, a bright and hard-working boy has a difficult family background, and his mother is an alcoholic. He is often bullied by some of his classmates. These same classmates now detect from the statistics re-published in the school paper that AA must fall into the 50% of sexually abused children who are not struggling at school ('good performers'). Thus, they have learned new (and in this case very sensitive) information from an ineffectively anonymised dataset.

The risk of combining information to produce personal data increases as data linkage techniques and computing power develop, and as more potentially 'match-able' information becomes publicly available. Indeed, computational power is doubling every year and data storage, due also to the availability of cloud services, is likely to become commodity. Thus, the risk of re-identification through data linkage is unpredictable because it can never be assessed with certainty what data are already available or what data may be released in the future.

Despite all the uncertainty, re-identification risks can usually be at least to some degree mitigated by adhering to the principle of data minimisation, that is, by ensuring that only the data necessary for a particular purpose are released.

The likelihood of re-identification being successfully attempted: the 'motivated intruder' test

A 'motivated intruder' test is an emerging concept, which is yet to be fully tested. It may be helpful to determine whether:

- anyone would have the motivation to carry out re-identification and
- whether the re-identification may/is likely to be successful.

The motivated intruder test essentially involves considering whether an 'intruder' would be able to achieve re-identification *if* motivated to attempt this. The 'motivated intruder' is a person (an individual or an organisation) who wishes to identify the individual from whose personal data the anonymised data have been derived. This test is meant to assess whether the motivated intruder would be successful. The approach assumes that the 'motivated intruder' is competent and has access to resources commensurate with the motivation it may have for the re-identification.

Some sorts of data will be more attractive to a 'motivated intruder' than others. For example, an intruder – in general - might be more motivated to re-identify personal data if such data:

- have a significant commercial value (including on the black-market or outside the European Union) and can thus be bought and sold for financial gain[27];
- can be used for law enforcement or intelligence purposes;
- reveal newsworthy information about public figures;
- can be used for political or activist purposes (e.g. as part of a campaign against a particular organisation or person);
- could be used for bad-intentioned personal reasons (e.g. stalking, harassment, bullying, or just to embarrass others);

[27] This may include, for example, transactional or other behavioural data from which individual consumer profiles could be inferred, which then could be used for advertisement purposes or for price discrimination; financial or other information enabling identity theft; sensitive information that could be used to blackmail individuals or to discriminate against them; medical information that could be used by insurance companies, for example, to deny coverage based on a pre-existing medical condition; information allowing inferences about creditworthiness that could be used to assess credit risks; etc.

- could raise curiosity (e.g. a local person's desire to find out who has been involved in an incident shown on a crime map).

While it is useful to think through the potential motivations of the potential intruders, the WP29 emphasises that there are also considerable limits to this approach:

- The exercise may be to some degree speculative.
- In the absence of obvious 'motivating factors' such as those described above the exercise may lead to false reassurances and may suggest that personal data that are relatively innocuous can be made available for reuse without effective anonymisation.
- Intruders can be sophisticated and innovative and 'ahead of the game', finding uses for de-identified data that are not obvious for others.
- With the increasing tendencies towards 'big data' analytics, there is an increasing risk that once de-identified, seemingly innocuous data may, once combined with other information, ultimately pose more serious risks.

6.5. Re-identification testing

In some circumstances it can be difficult to establish the risk of re-identification, particularly where complex statistical methods might be used by a third party to match various pieces of anonymised data. Therefore, as part of the overall assessment to identify the risk of re-identification, it is good practice to use re-identification testing – a type of 'penetration' or 'pen' testing - to detect and deal with re-identification vulnerabilities. This consists of attempting to re-identify individuals from the datasets that are planned to be released.

The first stage of a re-identification testing process should be to take stock of the datasets that the public sector body has published or intends to publish. The next stage should be to try to determine what other data - personal data or not - are available that could be linked to the data to result in re-identification. Targeted 'penetration tests', in particular, should help assess what are the risks of jigsaw identification, i.e. piecing different bits of information together to create a more complete picture of someone.

Of course, re-identification testing should not be considered as a panacea and should not lead to a false sense of security. First, the testing could be difficult to perform since it often requires significant technical expertise and adequate tools, as well as awareness of what other data may be available. Second, data controllers must also be aware that the risk of re-identification can change over time. For example, increasingly powerful and affordable data analysis techniques and tools are now available and correlation with other datasets becomes easier and easier as more and more data are generated. Therefore, organisations should carry out a periodic review of their policy on the release of data and of the techniques used to anonymise the data. In addition, decisions should never be based solely on current threats - but also on foreseeable future threats.

Once an assessment has been made under Section 6.4 regarding the risks of re-identification, and - where necessary - after carrying out re-identification testing, the public sector body can establish whether or not the dataset can be considered anonymised, in other words, whether it no longer contains personal data in the meaning of Article 2(a) and recital 26 of Directive 95/46/EC. If so, the dataset can be released without data protection constraints.[28] On the other hand if a test is successful these data may not (or may no longer) be made available as anonymised data but have to be

[28] See, however, Section 10.3 on 'License terms for anonymised data-sets', and in particular, the need to put in place safeguards to help continue to ensure that individuals will not be re-identified.

considered personal data (and thus, their release may not be possible, or may only be possible subject to the requirements discussed in Section VII).

6.6. Recall of compromised datasets

In the event of proven re-identification of data from an open dataset, the public sector body providing the dataset must be able to turn off the feed or remove the dataset from the open data website. In case of removing the dataset from the website, the public sector body must also inform re-users and call them to stop processing and delete all data coming from the compromised dataset. As informing all re-users will be difficult under an open licensing regime required by the PSI Directive public bodies must implement reasonably effective steps to address this issue. While a recall may often be too late to avoid the damage, it is a necessary step to help mitigate any adverse impact on the data subjects.

VII. Opening up personal data for re-use

7.1. Examples of publicly accessible personal data released by public sector bodies

While making available anonymised datasets is the typical scenario for PSI reuse initiatives, in some cases public sector bodies may also make available personal data for re-use.

Many publicly available registries such as land registries or business registries contain large amounts of personal data and are, because of e-government initiatives, increasingly available also online. There are also many other examples where legislators in particular Member States established a legal basis for making available personal data of individuals on the Internet or upon request for access to documents. These may include, for example[29],

- expenses, salaries or conflict of interest declarations of certain public officials, or beneficiaries of state aid (for example agricultural subsidies),
- names of organisations or individuals donating to political parties,
- tax declarations of individuals[30],
- court decisions (with the names of the parties or other individuals sometimes deleted or replaced by initials to reduce the risk of re-identification),
- electoral lists,
- court lists (i.e. schedule of cases to be heard before the court on particular days).

In each of these cases the public sector bodies or the legislators may proactively consider whether they wish to make these data available for reuse (for example, to improve public services such as provision of access to business or land registers). Potential re-users may also contact the public sector bodies to request re-use of the data. In some other cases, it is also possible that potential re-users will simply take the personal data, which are already available on-line, and use them without necessarily contacting the public sector body that released the information. In all three cases re-users would have to comply, of course, with data protection law as they are dealing with personal data.

7.2. Differences in national access regimes

The legal obligations to make publicly available certain personal data vary greatly in Member States due to different legal and cultural traditions. In some Member States there is a legal basis to make

[29] See also the examples provided in Section V, when discussing the scope of the PSI Directive.

[30] See, e.g. judgment of the European Court of Justice of 16 December 2008 in Case C-73/07 Tietosuojavaltuutettu vs Satakunnan Markkinapörssi Oy en Satamedia Oy.

certain personal data available, while other Member States would prohibit disclosure of the same personal data in the same situation. The PSI Directive acknowledges and makes it clear that it builds on the existing access regimes in the Member States and does not change the national rules for access to documents.[31]

7.3. Need for a data protection impact assessment and appropriate safeguards

Whenever personal data are considered to be made available for reuse – as a general rule – a cautious approach is absolutely necessary. The WP29 in particular, recommends that a thorough data protection impact assessment must be carried out before publication of the dataset (or before adopting a law requiring publication), which also assesses the possibilities and potential impact of reuse. In general, opening up personal data for re-use under an open license without any technical and legal restrictions on re-use is to be avoided.

7.4. Importance of a licensing regime

Additionally the WP29 recommends that a rigorous licensing regime should be put in place, which must also be appropriately enforced to ensure that the personal data will not be used for incompatible purposes - for example, for unsolicited commercial messages or otherwise in a way that the data subjects would find unexpected, inappropriate or otherwise objectionable.

7.5. Importance of a firm legal basis for publication and also for re-use

The WP29 reiterates the importance of establishing a firm legal basis for making personal data publicly available, taking into account relevant data protection rules, including the principle of proportionality, data minimisation and purpose limitation.

The WP29 recommends that any legislation calling for public access to data clearly specify the purposes for disclosing personal data. If this is not done, or only done in vague and broad terms, legal certainty and predictability will suffer. In particular, with regard to any request for reuse, it will be very difficult for the public sector body and potential re-users concerned to determine, what were the intended initial purposes of the publication, and subsequently, what further purposes would be compatible with these initial purposes. As it was already mentioned, even if personal data are published on the Internet, it is not to be assumed that they can be further processed for any possible purposes.

Any further re-use must, in these cases, have an appropriate legal basis (e.g. consent or requirement of law) under Article 7(a) through (f) of Directive 95/46/EC and comply with all other data protection principles.

7.6. Purpose limitation

It is challenging to implement the principle of purpose limitation effectively in case of PSI re-use. On the one hand, the very idea and driving force for innovation behind the concept of 'open data' and PSI re-use is that the information 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. The PSI Directive also requires licensing not be unnecessarily restrictive to re-use.

[31] That being said, as explained in Section 5.4, national legislation must still comply with Article 8 of the ECHR and Articles 7 and 8 of the of the EU Charter, as interpreted by relevant case law.

On the other hand, 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.[32] This principle equally applies to personal data that are publicly available. The mere fact that personal data are publicly available for a specific purpose does not mean that such personal data are open for re-use for any other purpose.

For example - senior public officials' expenses are made available on the Internet to provide for transparency but enabling re-use, by any member of the public, for other purposes may not be compatible.

As discussed in more detail in Opinion 3/2013 of the WP29 on purpose limitation (see Section III.2.2 and Annex 1), assessing whether further processing of personal data is incompatible with the purposes for which those data have been collected requires a multi-factor assessment. Account shall be taken in particular of:

(a) the relationship between the purposes for which the personal data have been collected and the purposes of further processing;
(b) the context in which the personal data have been collected and the reasonable expectations of the data subjects as to their further use;
(c) the nature of the personal data and the impact of the further processing on the data subjects;
(d) the safeguards applied by the controller to ensure fair processing and to prevent any undue impact on the data subjects.

These key factors need to be assessed when the decision is made whether to publicly disclose any personal data as well as in each case when personal data will be re-used. Some examples are provided below:

- A public sector body publishes contact information, including name, title, work address and work telephone number, about its civil servants in a directory. The obvious - although not specifically stated - purpose of this directory is to help the public identify whom to contact with official enquiries and other official business. A re-user wishes to 'harvest' the content of this directory, combine it with the employees' home addresses and telephone numbers (when such is publicly available, for example, in a telephone book), and make both home and work addresses and phone numbers available on an interactive map to showcase where different civil servants live and work. This data combination and reuse has to be considered incompatible with the initial purpose. A civil servant, whose work contact information is disclosed in order to be contacted by the public, would not have reasonably expected that this information will then be correlated with other data that she has made publicly available for another purpose not related to her work.
- In some Member States under the national law announcements about a planned marriage are public and can be consulted by anyone. Such publication is aimed at giving notice of the will of the engaged couple to marry and to allow interested persons to make an opposition. The fact that the personal data contained in the publication of the marriage announcements are available to anyone, however, does not allow third parties to use this information to send commercial communications to the couple. This additional use would be incompatible, having regard to the aim of the public disclosure of the marriage announcements, which is to allow the submission of any objection to the marriage as provided by law.

[32] It is only exceptionally - subject to the strict safeguards under Article 13 of Directive 95/46/EC - that data can be used in a way incompatible with the purposes specified at collection. See Section III.3 of Opinion 3/2013 of the WP29 on purpose limitation.

7.7. Commercial versus non-commercial purposes

Opinion 7/2003 highlights commercial activities as the main incentive for re-use of PSI in contrast with access to information, where the purpose of freedom of information laws is to ensure transparency, openness and accountability to the citizens.

Opinion 7/2003 also stresses that 'in the normal case [citizens] use the information for their own, non-commercial, purposes'. This statement needs to be updated in light of experience gained in the meantime with PSI reuse. Experience with open data initiatives has shown that re-use of PSI may also significantly contribute to enhancing transparency and accountability and may also lead to better use of public services. The distinction between re-use for commercial or non-commercial purposes should not be decisive when considering the compatibility of further use of personal data. The assessment of compatibility should not be primarily based on whether the economic model of a potential re-user is based on profit or not.

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.

7.8. Proportionality and other concerns

Another key principle provided for in Directive 95/46/EC is proportionality[33]. There are many different methods and modalities of making personal data publicly available. Some of these may be more intrusive than others and present more risks. Consequently, some may be considered proportionate, while others may not.

As with the purpose, there is concern how to control further processing of data and ensure compliance with other principles of data protection law, including but not limited to proportionality. Once data has been made publicly available, especially on the Internet, it is very difficult to effectively limit its use and ensure compliance with data protection laws.

Some of the challenges of ensuring compliance with data protection law include:

- how to ensure up-dating and accuracy of data that are disconnected from the primary source;
- how to ensure that the use of personal data remains limited to the functionalities foreseen for the initial purpose of publication;
- how to ensure timely deletion of data if the publication of personal data was foreseen for a limited amount of time only[34];
- how to exercise individuals' rights with regard to personal data made available for reuse (including the right to request correction, update or erasure).

7.9. Legal and/or technical limitations on re-use

Sometimes, legislation or technical design of the systems limit specific processing operations or establish other safeguards that limit the use of public registries (e.g. limiting the possibility of downloading the entire content of the registry or limiting search queries, for example, based on an

[33] See Article 6(1)(c) of Directive 95/46/EC.

[34] See, for example, the case before the European Court of Justice Volker und Markus Schecke GbR v. Land Hessen (Joined Cases C 92/09 and C 93/09), para 31: '[I]t is not possible to withdraw the data from the internet after the expiry of the two-year period laid down in Article 3(3) of Regulation No 259/2008'.

individual's name and surname). In this case reuse should in principle be allowed only in accordance with these specific limitations and conditions.

In this context it is important to carefully consider what measures - including both legal and technical measures - could be put in place to help ensure that data protection concerns, including the concerns outlined in Section 7.8, will be addressed. It is particularly important to consider how re-users will access the data - for example, through a bulk download function or through a customised interface featuring limited access capabilities subject to certain conditions. In this respect it is crucial what additional security controls will be put in place, such as, for example, a 'captcha[35]' verification system to prevent automated access and minimise the risk of harvesting an entire database. Using specific technical measures could help reduce misuse of personal data and negative impacts on data subjects that otherwise could be made possible by unlimited and unconditional access of reusers to entire datasets.

Importantly, in many cases it may be necessary to ensure that re-users will only be able to make targeted queries through technologies aiming at preventing bulk downloads of data records, such as through custom-designed Application programming interfaces ('APIs'). This may help ensure proportionality of use and reduce risks of misuse of entire databases. In addition, such customised interfaces may also help ensure that data are always updated, and also, that data will no longer be available through the API once a decision has been made to this effect by the public sector body concerned. On the other hand, it may limit the ways in which a re-user may re-use the data.

7.10. Accuracy, updates and deletion

Another specific question is what happens if personal data are published or made otherwise publicly available only for a limited period of time. Article 6(1)(e) of Directive 95/46/EC provides that personal data must be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed. Recital 18 of the PSI Directive also provides that if a 'competent authority decides to no longer make available certain documents for re-use, or to cease updating these documents, it should make these decisions publicly known, at the earliest opportunity, via electronic means whenever possible'.

It is however difficult or sometimes impossible to make sure that data are deleted or removed once they have been published and made available for re-use.

In this regard, it may provide some - though by no means complete – solution if data are not made available in a downloadable form, only via a customised API and subject to certain restrictions and security measures, as noted above.

VIII. Research data

Here it is important to draw a distinction between the publication of anonymised data on one hand (see Section VI) and limited access on the other hand. Clearly the open data agenda relies on the public availability of data. However, much research (importantly, scientific research, for both

[35] A CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) is a challenge-response system test designed to differentiate humans from automated programs. A CAPTCHA differentiates between a human and a computer by setting some task that is easy for most humans to perform but is more difficult for current computer programs to complete.

commercial or non-commercial purposes, but also other research) takes place by releasing data within a closed community, i.e. where a finite number of researchers or institutions have access to the data and where it is possible to restrict the further disclosure or use of the data and its security can be guaranteed.

Limited access is particularly important for the handling of personal data (often in pseudonymised form[36]) derived from sensitive source material or where there is a significant risk of re-identification. There can still be risks associated with limited access disclosure - but these are lower and can be better mitigated where data is disclosed within a closed community working with established rules.

A problem often faced by those using data for research purposes is that on the one hand they want data that are rich, granular, and usable enough for their purposes; on the other hand, they want to ensure that re-identification of individuals does not occur. At one end of the spectrum individual-level pseudonymised (for example, simply key-coded) data may be very valuable to researchers because of its individual-level granularity and because pseudonymised records from different sources can be matched relatively easy. However, this also means that there is a high re-identification risk: the possibility of linking several datasets (pseudonymised or not) to the same individual can be a precursor to identification or may enable direct identification.

Therefore, a higher level of scrutiny and additional caution is needed before any publication or making available for reuse of pseudonymised datasets. In general, the more detailed, linkable and individual-level the data are, the more limited and controlled the access to the data should be. The more aggregated and less linkable the data are, the more likely it is that they may be published and made available for re-use without significant risks.

This is a complex and evolving area and it would be inappropriate to categorically exclude the publication and reuse of all datasets that fall short of the high threshold of 'anonymisation' described in Section VI. That said, and while a case by case analysis and a careful assessment is always called for, as a rule of thumb, the WP29 considers that in general, release under the terms of the PSI Directive of individual-level datasets, or other datasets posing a significant risk of re-identification will often not be appropriate.

In addition, it is important to emphasise that should some such datasets nevertheless be published and made available, after a careful assessment of risks and benefits, the disclosure and any further reuse must be made in full compliance with data protection law (see Section VII). This is because these data, despite some (sometimes very significant) measures taken to decrease the risks of their re-identification, nevertheless continue to be considered personal data.

IX. Historical archives

Historical archives and museums also have specific characteristics requiring specific safeguards. In many cases, and depending on factors such as the age and sensitivity of the data and the context of collection, other options - such as allowing restricted access only subject to confidentiality obligations - may be more appropriate than digitalising and making the data available for re-use over the Internet without restrictions.

[36] See again Opinion 4/2007 on the concept of personal data, adopted on 20.06.2007 (WP 136), especially on p. 12-21 (discussing 'pseudonymised data', 'key-coded data' and 'anonymous data' on p. 18-21). The issue of information 'relating to' an individual is discussed on p. 9-12. It is also relevant, as noted on page 3, that the WP29 is currently working on providing further guidance on anonymisation techniques.

With regard to archives, it is also important to emphasise that although the sensitivity of data will generally decrease with the passage of time, the inappropriate release of many decades old records could still have a severely detrimental effect on the individual directly concerned but also on other individuals, such as members of his/her family, or descendants. This is particularly true for highly sensitive data. For example, released criminal records would continue to stigmatise an individual and hinder his/her rehabilitation. . Further, information that a deceased individual has been a secret agent or collaborator of an oppressive regime, a paedophile, perpetrator of crimes, suffered from a mental illness giving rise to a stigma, or suffered from a hereditary disease, may also all have a negative impact on the family (e.g. surviving spouse, children, or other descendants) of the deceased individual. DNA samples of deceased individuals, sometimes retained in the archives of public hospitals, could also require protection for similar reasons. Therefore, such information, even if it is relating to deceased persons, may require protection under data protection laws and/or other laws protecting fundamental rights, as the case may be.

Member States often have specific laws governing access to national archives, archives of recent historical periods of particular interest (such as archives evidencing collaboration with oppressive regimes), and files kept by the judiciary.[37] These laws often call for appropriate security measures and restrictions on access and other safeguards aimed at balancing the interests at stake and ensuring accessibility of certain personal data for historical research, transparency, and journalistic enquiries, while at the same time guarding that disclosures, when necessary, be limited so as not to prejudice the private and family life and dignity of the parties concerned.

With regard to 'purpose limitation', it is to be noted that historical archives typically store information for historical research purposes. These purposes are different from the original purposes for which data were collected. The materials that will ultimately end in archival collections were initially created for specific administrative purposes by the different public sector bodies. Typically, after a certain period of time, when the document will no longer be necessary for the initial administrative purposes, a selection process will be carried out, and the documents which are considered to have 'historical' value, will be transferred to the historical archives. The question here that arises is for what purposes the personal data stored in the archives should be available for reuse. In this context, it is important to make a careful assessment – considering the potential value of making archival material available for reuse, but also the potential impact on the rights, freedoms and dignity of the persons concerned.

Overall, it can be concluded that while digitalization of certain records containing personal data, and making them available for re-use may be appropriate in some situations, and some data may also be released in an anonymised form, in other cases limitations on the disclosure and re-use of personal data, and adequate security measures to protect such data are paramount. A thorough data protection impact assessment should ensure that no archival collection is made available for reuse unless any potential negative impact on the individuals concerned are excluded or any such risks are reduced to an acceptable minimum. The Archives sector could also consider developing Codes of Conduct or amending existing Codes to explain good practice.

[37] Other examples could include the archives of civil status registers, which contain, in some Member States, among others, cause of death, gender modification, name of partner (from which sexual orientation can be inferred), or the fact that an individual has been adopted. Access to these archives is also subject to specific conditions.

X. Licensing personal data for re-use

10.1. Relevant provisions of the PSI Directive

Recital 15 of the PSI Directive provides that 'ensuring that the conditions for re-use of public sector documents are clear and publicly available is a pre-condition for the development of a Community-wide information market. Therefore all applicable conditions for the re-use of the documents should be made clear to the potential re-users. Member States should encourage the creation of indices accessible on line, where appropriate, of available documents so as to promote and facilitate requests for re-use'.

Recital 26 of the PSI Amendment further provides that 'in relation to any re-use that is made of the document, public sector bodies may impose conditions, where appropriate through a license ...' and that 'Member States should where appropriate encourage the use of open, machine-readable formats'.

Further, Article 8(1) provides that '[p]ublic sector bodies may allow re-use without conditions or may impose conditions, where appropriate through a licence. These conditions shall not unnecessarily restrict possibilities for re-use and shall not be used to restrict competition.'

10.2. Licensing and data protection

Licences are a core part of the PSI regime. They can also affect the way personal data are processed and should be among the safeguards to be applied when making personal data (or anonymised data derived from personal data) available for reuse. Licences do not remove the need for compliance with data protection law but a data protection clause in license conditions would help to ensure compliance with data protection law by adding a layer of 'enforceability'. Such a clause could also help raise awareness by reminding re-users of their obligations as data controllers.

With regard to the content of the licenses, it is useful to distinguish between two different scenarios.

10.3. License terms for anonymised data-sets

First, with regard to anonymised data (that is, datasets that no longer contain personal data), license conditions should

- reiterate that the datasets have been anonymised;
- prohibit license-holders from re-identifying any individuals[38];
- prohibit license-holders from using the data to take any measure or decision with regard to the individuals concerned; and
- should also contain an obligation on the license-holder to notify the licensor in case it is detected that individuals can be or have been re-identified.

As an alternative to a license condition, a warning message could be brought to the attention of the re-users, in a prominent manner, on the open data portal. However, the adoption of license conditions should be promoted because it would have the added benefit of contractual enforceability.

[38] Limited exceptions might apply, for example, in bona fide cases of re-identification testing. Even in such cases, however, the results of the tests should be brought to the attention of the controller and the public sector body concerned, and the re-identified data should not be published or otherwise disseminated more broadly.

Recall of compromised datasets

The possibility to alert the licensor of the fact that re-identification has taken place or can take place must be available for all other web-users, including the data subjects themselves. When an increased risk of re-identification is discovered by the licensor, a procedure should be foreseen in the license whereby the licensor can 'recall' the 'compromised' dataset. In other words, the data protection clause should give the licensor the right to suspend or terminate accessibility of data (for example, the right to turn off the API or remove the file from the platform). The licensor should make all reasonable efforts to require all re-users to delete all or parts of the datasets that have been compromised (have become re-identifiable). This should include prominent notices on websites such as open data portals and forums/email lists/social media accessed by groups or individuals who are likely to be re-using the data. Requiring registration may be the most effective means of recalling datasets but this should not be encouraged if it will require the collection of new personal data from re-users and would have a general effect of discouraging use of PSI websites and other services.

10.4. License terms for personal data

When personal data are licensed, there is a need to define the limits of the use of such data. Here the key concern is to ensure that any reuse will be limited to what is 'compatible with the purposes for which the data has been initially collected'.[39] To achieve this, the license conditions must at least make it clear for what purposes data was first published and give indication of what would and what would not be considered as compatible use of personal data.

It needs to be noted, however, that this should not 'unnecessarily restrict possibilities for re-use' (Article 8(1) of the PSI Amendment). This may often mean that the generic terms of standard open licences are not suitable and specific licences may need to be developed for certain personal data, or templates may be used, which could be adapted.

At present some standard open licences (such as the UK open government licence) exclude personal data – they are not licensed at all under the terms.

10.5. Robust enforcement should follow in case of re-identification or incompatible use

Once data have been published under a licence - such as an open government licence - it may be difficult to protect them from further incompatible use, disclosure or to keep them secure. Monitoring the reuse and enforcing any violations, be it in the form of re-identification of data subjects, or further use for an incompatible purpose by the licensor, is in this context very important.

While the WP29 reiterates the important role that public sector bodies should play, it also emphasises that where a re-user collects personal data through a re-identification process, the reuser will most likely be considered to be processing personal data unlawfully and could be subject to enforcement action by data protection authorities. This includes severe fines under the proposed Data Protection Regulation.

[39] See again Opinion 3/2013 of the WP29 on purpose limitation.

XI. Conclusions

In conclusion, the WP29 reiterates that PSI reuse may bring benefits leading to greater transparency and innovative re-use of public sector information. However, the resulting greater accessibility of information is not without risks. In order to ensure the protection of the privacy and personal data of individuals, a balanced approach needs to be followed and data protection law must help guide the selection process of what personal data can or cannot be made available for reuse and what measures to take to safeguard personal data.

Irrespective of the 'principle of reuse' formulated in the PSI Amendment, reuse for any commercial or non-commercial purposes under the terms of the PSI Directive is not always appropriate in cases when the PSI to be reused contains personal data. Rather than personal data, it is often statistical data derived from personal data that are and that should be made available for reuse.

Nevertheless, it may also be possible, in some situations, that personal data may be considered available for reuse under the terms of the PSI Directive, where necessary, subject to additional legal, technical or organisational measures to protect the individuals concerned. For these cases the WP29 reiterates the importance of establishing a firm legal basis for making personal data publicly available, taking into account relevant data protection rules, including the principle of proportionality, data minimisation and purpose limitation. In this context, it is also important to highlight again that any information relating to an identified or identifiable natural person, be it publicly available or not, constitutes personal data. Therefore, access and re-use of personal data that have been made publicly available remain subject to applicable data protection law.

In light of these considerations, the WP29 recommends that:

- The fact that some PSI may contain personal data should be taken into account at the earliest occasion when considering making PSI publicly available, following the principles of 'data protection by design and default';
- With this in mind, the public sector body concerned (or the legislator, as the case may be) should carry out a data protection impact assessment before any PSI containing personal data may be made available for reuse (or before adopting a law allowing publication of personal data and thus making them potentially available for reuse); a data protection impact assessment should also be carried out in situations where anonymised datasets derived from personal data will be made available for reuse;
- When datasets are anonymised, it is essential to assess the risk of re-identification, and a good practice to carry out re-identification testing;
- The outcome of the assessment could help identify appropriate safeguards to minimise risks including, without limitation, technical, legal and organisational measures, such as appropriate license terms and technical measures to avoid bulk download of data, and adequate anonymisation techniques; it may also lead to a decision to refrain from publication and/or making available for re-use;
- The terms of the licence to re-use PSI should include a data protection clause, whenever personal data are processed including also situations where anonymised datasets derived from personal data will be made available for reuse;
- Where the data protection impact assessment concludes that an open license is not sufficient to address data protection risks, public sector bodies should not make personal data available under the PSI Directive. (However, the public sector body may still use its discretion to consider re-use outside the terms and scope of the PSI Directive and may also require applicants to demonstrate that any risks to the protection of personal data are adequately

addressed and that the applicant will process data in compliance with applicable data protection law);

- Where appropriate, public sector bodies should ensure that personal data are anonymised and license conditions specifically prohibit re-identification of individuals and re-use of personal data for purposes that may affect the data subjects;
- Finally, Member States should also consider establishing and providing support to knowledge networks/centres of excellence and thereby enable sharing of good practice related to anonymisation and open data.

Done at Brussels, on 5 June 2013

For the Working Party
The Chairman
Jacob KOHNSTAMM