Originally published in: Porsiel, Jörg (Ed.):
Maschinelle Übersetzung für Übersetzungsprofis. -
Berlin: BDÜ, Weiterbildungs- und Fachverlagsgesellschaft mbh, 2020. S. 42-57.

"Cover this data that I cannot see"[20]

Privacy by Design in Machine Translation

Dr. iur. Paweł Kamocki, Dr. jur. Marc Stauch

1 Introduction

It is not uncommon, and indeed quite commonsensical, to believe that as long as one does not do anything morally wrong or evil, one does not have to worry about the legality of one's acts. This, however, is not entirely true; to quote what is perhaps the most obvious example, income tax rates and deadlines for submitting tax declarations are not set according to natural law or other moral guidance, but rather as a result of purely economic analysis and cold calculation; and yet, everybody seems to understand and accept that failing to pay income tax on time will result in a sanction, and potentially even time in prison — not so much because it is morally wrong or evil, but simply because it violates the rules by which we have all agreed to play. We may lack the perspective necessary to determine the fairness of some of these rules, but deep inside we want to believe that they are indeed fair. This is why most of us pay our tax on time.

The same reasoning should apply to privacy laws — they are strict, relatively little-known to the general public, and may appear to defy common sense — but they are laws nevertheless. Even if they do seem unfair in certain contexts, they are in fact the cornerstone of democratic society in the information age.

Perhaps more specifically in the research and development community, some tend to believe (often for plausible reasons) that their actions contribute to the progress of humanity, and as such are covered by some sort of 'public interest' justification. This is true to some extent — intellectual property rules and privacy laws usually come with some sort of 'research exemptions' (relaxing some of the legal requirements that operate in the ordinary,

20 This is a paraphrase of the machine translation of *"Couvrez ce sein que je ne saurais voir"* (from Tartuffe by Molière); a literary translation of this passage reads: *"Cover this breast which I cannot behold: / Such a sight can offend one's soul. / And it brings forth guilty thoughts"*

Publikationsserver des Leibniz-Instituts für Deutsche Sprache
URN: http://nbn-resolving.de/urn:nbn:de:bsz:mh39-99267

non-research context). However, this does not excuse researchers from knowing what the residual requirements on them are.

Quite the contrary — rather than relying on the imaginary 'public interest' (which, at some level, boils down to saying 'my research is more important than your rights and freedoms'), researchers need to know what the law allows them to do and under which specific circumstances.

Hopefully, even if not yet fully convinced, the reader is now more interested to know whether privacy and data protection laws can really apply to Machine Translation. In this paper, we will focus mostly on the European perspective on data protection. It should be noted, however, that Europe, especially with the adoption of the General Data Protection Regulation (GDPR), set the tone for the debate about privacy protection worldwide, and we believe that, a few intricacies aside, this chapter could also be interesting for readers who are not based in the European Union.

2 The Relevance of Privacy and Data Protection Laws for Machine Translation

The relevance of privacy and data protection laws for machine translation has already been discussed in greater detail in our previous papers[21]; we will provide just a quick overview here.

Since 2018, the General Data Protection Regulation (EU Regulation 2016/679 ('GDPR')) is the key European legislation applicable in a general way (across multiple different areas of data use) to the processing of personal data. Here, the starting point for demonstrating the relevance of Privacy Data Protection Laws for Machine Translation is thus the definition of personal data.

In fact, the GDPR (in its Article 4(1)) defines such data very broadly as 'any information relating to an identified or identifiable natural person'. This definition is not new — it was also present in the Data Protection Directive of 1995 that preceded the GDPR. It covers not only personal details (name, surname,

21 *Kamocki, P. and M. Stauch (2017). Data Protection in Machine Translation under the GDPR, in: J. Porsiel (ed.) Maschinelle Übersetzung, BDÜ Fachverlag, Berlin, and Kamocki, P., J. O'Regan and M. Stauch (2016). All Your Data Are Be-long to us. European Perspectives on Privacy Issues in 'Free' Online Machine Translation Services, in: D. Aspinall et al. (eds.), Privacy and Identity Management. Time for a Revolution?, Springer.*

gender, date of birth...) or contact information (e-mail, postal address, phone number...), but also any other information related to a living person, either because it says something about the person (i.e. related via the *content*), or because it can be used to evaluate the person and treat him/her in a certain way (i.e. related via the *purpose*), or because it is likely to have an impact on the person's rights and freedoms (i.e. related via the *result*). The form of the information (digital or analogue) is irrelevant.[22]

Therefore, pieces of information such as 'X was born on July 5', 'X's mother was British', 'X speaks fluent Swedish and some Romanian, but only very little English', 'X tends to spell "honey" with a "u"', 'X is afraid of spiders' and 'X does not like goat's cheese' are all potentially personal data, depending on whether X is identified directly (i.e. by a sufficiently distinctive name and surname) or indirectly (e.g. via a social security number or an IP address[23]), or can be identified by any means reasonably likely to be used (e.g. by cross-referencing available datasets)[24].

Looking next at 'processing', this is also very broadly defined as "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means (GDPR, Art. 4(2))". In particular, the purpose of the processing is irrelevant at this stage — in principle, the GDPR applies regardless of whether the processing is carried out in the 'public interest', for commercial purposes, or even crudely for spying on people.

In the light of the above definitions, it seems obvious that Machine Translation may indeed involve the processing of personal data, namely at several stages:

- at the development stage, the parallel corpora collected from various sources and used to train MT engines may contain personal data, especially if they were obtained via web crawling;
- at the deployment stage, when personal data may be entered into an MT system or otherwise provided by the user; the data can relate to the user him- or herself, but also to third persons (e.g. when the user inputs an email

22 Article 29 Data Protection Working Party, Opinion 4/2007 on the concept of personal data, adopted on 20 June 2007 (WP 136).

23 The Court of Justice of the European Union confirmed that IP addresses constitute personal data in its decision C-582/14 – Patrick Breyer v Germany of 19 October 2016.

24 Cf. Recital 26 of the GDPR.

that he/she intends to translate for sending, it is likely to contain information about the addressee);

- at the evaluation stage, where user-entered data are used to qualitatively or quantitatively evaluate the performance of the system
- potentially (in some systems) also at the repurposing stage, in which user-entered data, possibly cross-referenced with other datasets, can be used for a purpose unrelated to machine translation, such as profiling or direct marketing (e.g. targeting advertisements for cookbooks or kitchenware at a user who regularly translates recipes).

Within the scope of the present chapter, we have opted to focus on a key requirement introduced into this area of law by the GDPR, namely 'Privacy by Design', which brings together many aspects of data protection that are of importance for MT systems.

3 The Concept of Privacy by Design

The principle of Privacy by Design has drawn considerable public attention since its explicit legal embodiment in the GDPR. One should not forget, however, that the principle as such has actually been around for decades, and that it originated outside of Europe.

Although some sources would trace the origins of the concept back to the 1970s[25], the paternity of privacy by design is commonly attributed to Ann Cavoukian, the Privacy Commissioner of Ontario, Canada, co-author of the 1995 international report on Privacy Enhancing Technologies (PET). Cavoukian famously argued *"that the future of privacy cannot be assured solely by compliance with regulatory frameworks; rather, privacy assurance must ideally become an organization's default mode of operation"*[26]. In this approach, to use a common metaphor, privacy protection should be 'baked into' a technology or a product, rather than just 'sprinkled over' it.

More recently, in 2009, Cavoukian listed what she called 'Seven Foundational Principles of Privacy by Design'. These principles are as follows:

---

25 European Data Protection Supervisor (EDPS) (2018). Opinion 5/2018. Preliminary Opinion on privacy by design, pp. 3-4.

26 Cavoukian, A. (2009), Privacy by Design. The 7 Foundational Principles, https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf (retrieved 23 January 2020).

1. *Proactive not Reactive; Preventive, not Remedial.* According to this principle, one should make an effort to anticipate privacy threats and take actions to prevent them from materialising.

2. *Privacy as the Default Setting.* The default settings of a system should therefore be set to the level that provides maximum privacy; any deviation from maximum privacy necessitates the user's explicit action (modification of the settings). (In fact the GDPR tends to view Data Protection by Default as a separate requirement, related to but distinct from Data Protection by Design.)

3. *Privacy Embedded into Design.* This principle seems to be the cornerstone of Data Protection by Design, as required by Article 25 of the GDPR (see below).

4. *Full Functionality: Positive-Sum not Zero-Sum.* According to this principle, privacy should ideally have no detrimental effect on functionality or security of the system.

5. *End-to-End Security - Lifecycle Protection.* Privacy and security must be guaranteed from the conception phase and through the entire lifecycle of the data (including, where applicable, their long-term archiving).

6. *Visibility and Transparency.* This principle requires that users be informed about various aspects of the processing, and that they are able to verify the accuracy of the information they have been given.

7. *Respect for User Privacy: Keep it User-Centric.* The user must play a central and active role in the processing; in the GDPR, this is guaranteed by the rights granted to data subjects.

By the time these rules were formulated, the importance of embedding privacy into design was already acknowledged in EU legislation, albeit rather timidly: Recital 46 of the 1995 Data Protection Directive (95/46/EC) stated that *"the protection of (...) personal data requires that appropriate technical and organizational measures be taken, both at the time of the design of the processing system and at the time of the processing itself"*. In the past decade, however, the concept has gained much more traction on both sides of the Atlantic.

In 2010, Privacy by Design was mentioned in a major EU policy document, Digital Agenda for Europe[27], as essential for practical enforcement of the right to privacy and to the protection of personal data in the EU. The Commission

27 COM(2010)245

defined Privacy by Design as an approach in which *"privacy and data protection are embedded throughout the entire life cycle of technologies, from the early design stage to their deployment, use and ultimate disposal"*. This definition was also mentioned in a 2010 Commission communication entitled 'A comprehensive approach on personal data protection in the European Union'[28].

Privacy by Design has also been discussed by US policymakers. In 2012, the Federal Trade Commission adopted its 'Recommendations for Businesses and Policymakers concerning Protecting Consumer Privacy in an Era of Rapid Change'. The document listed Privacy by Design (defined as an approach in which *"companies should promote consumer privacy throughout their organizations and at every stage of the development of their products and services"*) as one of the key recommendations.

The heyday of Privacy by Design is arguably yet to come, as it is now an important part of the EU's General Data Protection Regulation, which entered into force on 25 May 2018.

4 Analysis of the Privacy (Data Protection) by Design Requirement in the General Data Protection Regulation

Article 25(1) of the GDPR reads "Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects".

28 COM(2010)609

In essence, this provision requires that data controllers: (1) take appropriate technical and organisational measures designed to implement the data protection principles; and (2) integrate the necessary safeguards into the processing stage in order to meet the requirements of the GDPR and protect the rights of data subjects.

Perhaps the most important part of Privacy by Design is that these measures and safeguards must be implemented not only 'at the time of the processing itself', but also, crucially, 'at the time of determination of the means for processing', i.e. at the time when the decisions concerning the system design are made.

The GDPR also emphasises 'effectiveness' of measures and safeguards; in other words, the measures should be appropriate and really contribute to achieving the desired goals. Implementing a measure that does not really minimise any privacy risks for the users would therefore not meet the Privacy by Design obligation.

Article 25(1) lists 'costs of implementation' among the factors to be taken into account when implementing Privacy by Design. This means that the costs of implementation should be weighed against the risks that they seek to mitigate or avoid; there is no need to implement a very costly measure to mitigate a low risk, especially where there is a cheaper way to achieve a similar result. On the other hand, the European Data Protection Board emphasises that "incapacity to bear costs is no excuse for non-compliance with the GDPR"[29].

At first glance, only controllers are required to implement Privacy by Design. However, other actors in the process may also be indirectly concerned. For example, if a controller outsources certain processing tasks to others, it must choose a contractor/processor 'providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements [of the GDPR]' (Article 28(1) of the GDPR), including Privacy by Design. Moreover, Recital 78 of the GDPR states that "When developing, designing, selecting and using applications, services and products that (...) process personal data (...), producers of the products, services and applications should be encouraged to take into account the right to data protection when developing and designing such products, services and applications and, with due regard to the state of the art, to make sure that controllers and processors are able to fulfil their data protection obligations (...)". Therefore, MT developers, even if they don't qualify as controllers

29 European Data Protection Board (EDPB) (2019). Guidelines 4/2019 on Article 25 Data Protection by Design and by Default adopted on 13 November 2019 (version for public consultation), p. 8.

or processors (which, especially in the latter case, is debatable), are also expected to observe Privacy by Design.

According to Article 83(4) of the GDPR, any infringement of the Privacy by Design principle can be sanctioned with an administrative fine of up to 10,000,000 EUR. Data Protection Authorities may also assess compliance with Privacy by Design and issue corrective measures, such as warnings or orders to comply. Quite recently, the French Data Protection Authority fined a small translation company (Uniontrad) 20,000 EUR for (among other violations) failing to observe Privacy by Design.[30]

Article 25(1) of the GDPR seems to promote what can be described as a principle-based approach to implementing privacy in design. Another possible approach to the problem, hinted at by the reference to safeguards for protecting the rights of data subjects (and more fully articulated in Article 24 of the GDPR, see below) is risk-based. The former approach is more static and perhaps more universal; the latter seems to require more practical experience with protecting personal data (which is still a rare commodity), but can, if properly applied, lead to more specific, tailor-made results.

Below we consider these two approaches in turn with particular reference to how they may operate in relation to the design and running of MT applications.

4.1 Principle-based Approach to Implementing Privacy by Design

The principle-based approach focuses on a set of pre-defined objectives, or goals to achieve. These goals, like in Article 25(1) of the GDPR, seem to correspond to the 'data protection principles' defined in Article 5 of the Regulation. These principles are as follows:

Lawfulness, i.e. the principle according to which processing can only be carried out if it is based on one of the grounds listed in Article 6 of the GDPR. From the point of view of MT, as we have stated previously, the most relevant grounds for processing are the data subject's consent (which ideally should be obtained for the 'primary' processing of personal data in MT, i.e. the translation itself) or legitimate interest of the data controller (which may justify certain 'secondary' purposes of processing, such as evaluation). From the design

30 CNIL, Délibération de la formation restreinte n° SAN-2019-006 du 13 juin 2019 prononçant une sanction à l'encontre de la société UNIONTRAD COMPANY

perspective, technical and organisational measures that can be taken to ensure adherence to this principle include:

- the legal basis should be determined before the processing starts;
- clear differentiation between processing operations based on consent of the data subject and those based on the controller's legitimate interest;
- for processing based on legitimate interest: carrying out a documented assessment ('balancing test') of the interests at stake, and how they outweigh the interest of the data subject; if applicable, measures such as pseudonymisation of data (e.g. replacing IP addresses with 'pseudonyms' such as code numbers, in a consistent manner) can be taken to mitigate any negative impact that the processing may have on the data subject;
- for processing based on consent, the MT user should be provided with means to withdraw his or her consent at any time (Article 7 of the GDPR requires that it should be as easy to withdraw consent as it was to give in the first place).

Fairness. This very general and abstract principle seems to be of particular importance in online MT applications, where any secondary use of the data, especially for commercial purposes (e.g. building user profiles), may seem unfair and misleading to an ordinary user. Designing the tool, including the user interface, in such a way as to make it correspond to the expectations of an average user, is of particular importance here. Measures that can be taken include e.g. providing clear and easily accessible information that the data provided by the user are analysed not only at the word level, but also in broader contexts (sentence or even paragraph level) in order to improve the translation; making it clear for the user that the data he or she inputs in the system are not immediately deleted, but stored and reused in order to evaluate and improve the engine.

The GDPR (in its Articles 15 through to 22) also gives specific rights to data subjects that are closely linked to the principle of fairness, including the right to the erasure of their data (right to be forgotten), to object to processing, and to be protected against automated decision-making. Specific internal procedures should be designed, and contact points designated to facilitate the exercise of these rights by the data subject.

The principle of **Transparency** requires the data controller to provide data subjects with information about the processing. Of particular relevance for MT is Article 13 of the GDPR, which lists the information that shall be provided to the data subject when the data is collected directly from him or her. The required elements are the following:

- identity and contact details of the controller (or, if the controller is based outside the EU, of its representative in the Union);
- contact details of the data protection officer, if a data protection officer has been appointed (which is mandatory in cases listed in Article 37 of the GDPR);
- all the purposes of the processing (those apparent to the user, i.e. translation, but especially those that may not be obvious, e.g. MT system evaluation, research in the field of MT, user profiling, etc.);
- where the processing is based on consent, the option of withdrawing consent at any time;
- where the processing is based on legitimate interest, the interest pursued by the controller;
- if applicable, the entities to which the data may be disclosed (e.g. project partners) and intended transfers outside of the European Economic Area;
- the storage period (see below about storage limitation) or the criteria used to determine this period;
- the existence of the rights of data subjects;
- if applicable, the existence of automated decision-making, including profiling (see Article 22 of the GDPR for more information);
- the right to lodge a complaint with a data protection authority.

In practice, this information is usually provided in a Privacy Policy. The Policy should be designed in a *"concise, transparent, intelligible and easily accessible form, using clear and plain language"* (Article 12 of the GDPR). The Policy should be easily accessible and visible (e.g. a link can be added to the website's footer, so that it is always one click away). It should be shown to the user at the appropriate time, taking into account the context, and in any case before the processing starts (e.g. an excerpt of the Policy may appear before the user creates an account, and another excerpt before he or she inputs data into the system and hits the 'Translate' button). It is good practice to use various media other than text, and to make the policy layered (e.g. bullet points with the most essential elements with additional information available when the user clicks 'Learn more'). An interesting way to present some information in the Privacy Policy is by means of a table, e.g.:

For what purpose do we process your data?	What data do we process for this purpose?	On what grounds do we process the data?	For how long do we store the data for this purpose?
Creation of a user account	Username, password, e-mail address	Your consent	A year after your last login, after which the data are deleted
Translation	Only the data you input into the system	Your consent	24 h, after which the data are anonymised
Generation of user statistics	Your IP address, operating system and time spent on our website	Our legitimate interest in improving the service	1 month, after which the data are anonymised

Purpose limitation is the principle according to which the controller must collect the data for specified, explicit and legitimate purposes, and not further reuse them for purposes incompatible with those for which they were collected. These purposes should be defined as early as the conception stage, and when a new purpose appears, its compatibility with the original purposes should be assessed in a documented manner, in particular taking into account user expectations. The use of technical measures such as hashing in order to prevent data from being reused for another purpose may be an effective way to safeguard this principle. Moreover, functional separation of data used for different purposes (e.g. user account details and data entered into the MT system are stored in separate databases) may also be used to ensure purpose limitation.

From the perspective of MT and other data-intensive technologies, **Data Minimisation** is probably the most problematic data protection principle. It requires that the data processed are limited to those 'adequate, relevant and necessary' for achieving the purposes of processing. It is particularly important to observe this principle in the development phase, i.e. when building MT engines. Rather than randomly crawl the entire (multilingual) Internet, as some projects aim to do, the sources of data should be carefully selected in order to avoid collecting unnecessary personal data e.g. from social media profiles (*'Select before you Collect'*). The developers should assess (again, preferably in a documented manner) the necessity of the collection by asking themselves if the same result could be achieved in a reasonably practical way without collecting personal data (and cost-effectiveness is not a primary consideration). For development and evaluation, it may be best to avoid personal data altogether (instead using data anonymised with an appropriate anonymisation technique, such as random permutation of named entities of the same class within a dataset). If personal data are used, the processing operations themselves should be minimised, i.e. the workflow should be designed in such a way as to minimise the number of copies and entry points.

According to the **Accuracy** principle, the data should be accurate and kept up to date; inaccurate data should be rectified or deleted. This is of particular relevance for data collected from the user in the process of creating an account, if applicable, especially if such data as name, surname, age or country of residence are collected (which would pass the necessity part of the data minimisation principle with difficulty). Adherence to the accuracy principle is safeguarded by the data subject's rights of access and rectification (Articles 15 and 16 of the GDPR). It is important to design the tool in such a way as to enable the user to consult his or her profile data, and to edit them at any time.

Storage limitation is another principle that has caused consternation among those involved in the field of data-intensive technologies, such as MT, who would like to be able to keep their hard-earned data forever. Unfortunately, the GDPR states that personal data can in principle only be stored ('in the form which permits identification of data subjects', i.e. before anonymisation) for the duration necessary to achieve the purposes of the processing[31]. Many sector- and country- specific rules, either stemming directly from various statutes or emanating from data protection authorities, require a specific minimum or maximum data storage period (data retention periods). When designing an MT system, it is important to predefine the storage periods, or at least the criteria used to determine them, taking into account the applicable rules. It is also advisable to create internal procedures for enforcing data retention policies, and for deleting or anonymising datasets at the end of the retention period. If possible, the process of deletion or anonymisation can be automated, and its performance periodically evaluated.

The principle of **Integrity and Confidentiality** refers to the technical and organisational security of the data and the environment in which they are processed. This principle is best safeguarded by the use of state-of-the-art storage techniques, with backups and strict access controls. The adopted measures should safeguard various aspects of data security, in accordance with the 'CIA triad', Confidentiality, Integrity and Access. Pursuant to Articles 32-34 of the GDPR, an internal procedure for containing, notifying and possibly communicating data breaches (e.g. the loss or theft of personal data held by the controller) should also be adopted, and periodic drills may be run to assess its performance.

Finally, the **Accountability** principle requires controllers to be able to demonstrate their compliance with the GDPR to responsible supervisory authorities.

---

31 Data processed solely for research purposes can be stored for longer periods of time if the processing is subject to 'appropriate safeguards' (cf. Article 5(1)(e) and Article 89 of the GDPR).

Documenting the design process and the efforts made to ensure adherence to the data protection principles is in itself an accountability measure. Other such measures include clear definition of roles in processing (e.g. controller vs. processor); this is particularly relevant in the case of joint controllership (e.g. where the MT service is developed, deployed and run by a consortium), where the responsibilities should be clearly assigned to each controller in a joint controllers agreement.

The principle-based approach to Privacy by Design can also be centred around objectives other than GDPR principles. Such alternative approaches underline the international pedigree of Privacy by Design, and its technical and organisational, as opposed to purely legal, character. For example, a relatively established approach distinguishes three privacy-related objectives: transparency, unlinkability and control. These objectives are achieved through design strategies labelled with action verbs (inform, control, minimise, abstract, separate, hide, enforce, demonstrate), which in turn are embodied via different 'design patterns' (specific measures or safeguards, such as encryption)[32].

Although framed differently, these objectives are essentially the same as those set by the GDPR principles[33].

4.2 Risk-based Approach to Implementing Privacy by Design

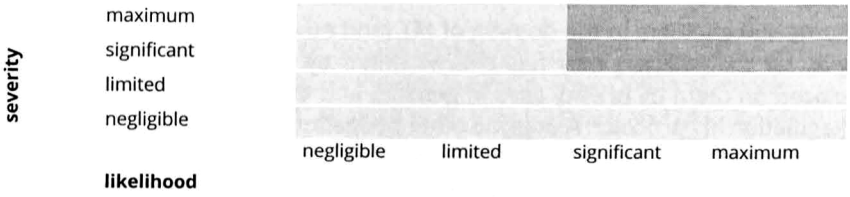
As noted earlier, another possible approach to Privacy by Design consists in identifying potential privacy risks associated with the tool, and then adopting specific measures to prevent them from materialising. This approach stems directly from Article 24 of the GDPR, according to which the controller shall implement appropriate technical and organisational measures to ensure compliance with the GDPR, taking into account, *"the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons"*.

An interesting methodology for assessing privacy risks, proposed by the French Data Protection Authority (CNIL), consists of representing the threats

32 More about the privacy design patterns can be found at https://privacypatterns.org/

33 For a more thorough comparison between the GDPR principles and the goals of Transparency, Unlinkability and Control, see: European Data Protection Supervisor (EDPS) (2018). Opinion 5/2018. Preliminary Opinion on privacy by design.

in a coordinate system, evaluating its severity and likelihood (as illustrated below).[34]



For example, unauthorised access to MT data can, at least in some MT systems, be a risk of maximum severity to the data subjects (as they may have used the system to translate their confidential or sensitive information), and if the data are not stored in a secure environment (e.g. they can be accessed via URL links or otherwise retrieved from the system by another user), then the likelihood of the risk is at least significant.

The role of the controller in this approach is to anticipate the risk and adopt preventive measures before the processing starts. For example, if the controller stores sensitive personal data in which a malicious third party may foreseeably have an interest, the controller should strongly consider use of encryption or anonymisation techniques: any unauthorised access to the data will then have far less serious consequences for data subject privacy.

Other privacy risks that may materialise in MT systems include unauthorised cross-referencing or profiling, or gathering information about the user which may be used to his or her disadvantage (e.g. to demonstrate that the claims about his or her language skills on his or her LinkedIn page are exaggerated). The severity and likelihood of such risks should be evaluated on a case-by-case basis, and the preventive measures should directly address the identified risks.

It should be kept in mind that if the processing is 'likely to result in a high risk' to the rights and freedoms of individuals (i.e. risk that falls within the dark zone on the graph above), the controller will be required to carry out a Data Protection Impact Assessment (under Article 35 of the GDPR). Even where the DPIA is not mandatory, some form of risk assessment should always be carried out as an integral part of Privacy by Design.

---

34 Cf. CNIL, Privacy Impact Assessment (PIA) templates, February 2018, available at: https://www.cnil.fr/sites/default/files/atoms/files/cnil-pia-2-en-templates.pdf (retrieved 23 January 2020), p. 23, for more information about this methodology, see also CNIL, *Autorisation unique AU-053 – Contrôle d'accès biométrique avec base centrale*, p. 9.

5 Conclusion

Some stakeholders in the domain of MT (and other data-intensive technologies for that matter) may feel overwhelmed by the organisational burden placed on them by privacy laws in general, and the General Data Protection Regulation in particular. Alongside other obligations, such as keeping a record of processing activities, carrying out an impact analysis, or appointing a Data Protection Officer, implementing Privacy by Design may initially appear as another complex exercise in futility.

If implemented appropriately, however, Privacy by Design may indeed turn into a guiding principle, a real breadcrumb trail through privacy principles, allowing the controller to better understand the risks associated with the processing. Ultimately, if taken seriously, Privacy by Design facilitates creation of better, more robust and user-friendly tools and applications.

With an appropriate approach, implementing Privacy by Design may ultimately offer a significant competitive advantage, not only because it leads to better design, but because it should (as per Recital 78 of the GDPR) be taken into account in the context of public tenders. Above all, it is an essential means for businesses to gain and retain customer trust in line with increasing expectations of safe and ethical data use, and where the costs (legal and reputational) of non-compliance can be very high.

6 Bibliography

Agencia Espanola Proteccion Datos (AEPD) (2019). A Guide to Privacy by Design.

Cavoukian, A. (2009), Privacy by Design. The 7 Foundational Principles. Available at: https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf (retrieved 23 January 2020).

European Data Protection Board (EDPB) (2019). Guidelines 4/2019 on Article 25 Data Protection by Design and by Default adopted on 13 November 2019 (version for public consultation).

European Data Protection Supervisor (EDPS) (2018). Opinion 5/2018. Preliminary Opinion on privacy by design.

European Union Agency for Network and Information Security (ENISA) (2014). Privacy and Data Protection by Design -- from policy to engineering.

Kamocki, P. and M. Stauch (2017). Data Protection in Machine Translation under the GDPR, in: J. Porsiel (ed.) Maschinelle Übersetzung, BDÜ Fachverlag, Berlin.

Kamocki, P., J. O'Regan and M. Stauch (2016). 'All Your Data Are Be-long to us'. European Perspectives on Privacy Issues in 'Free' Online Machine Translation Services, in: D. Aspinall et al. (eds.), Privacy and Identity Management. Time for a Revolution?, Springer.