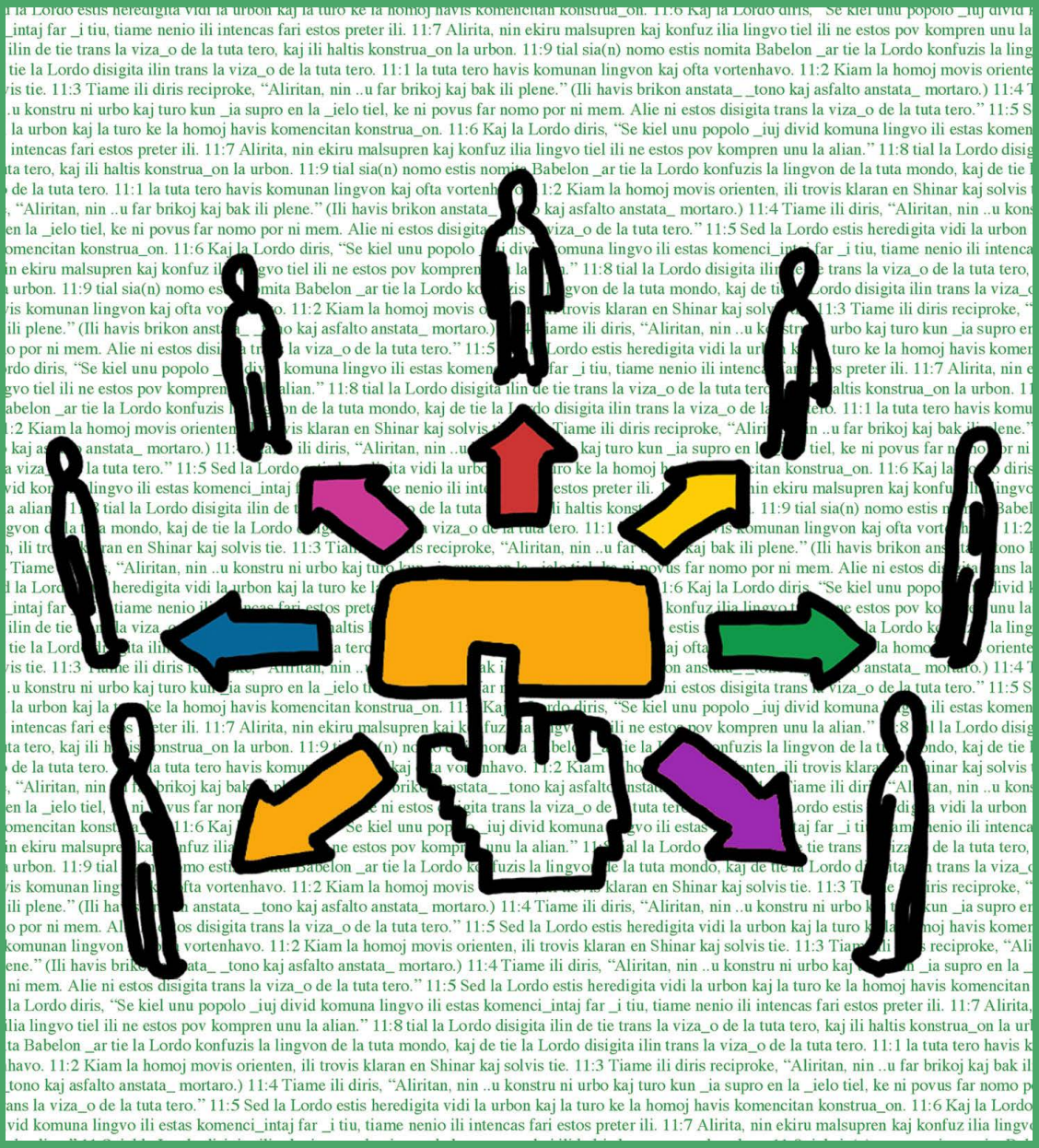
Who Owns My Language Data?

Realities, Rules and Recommendations

A White Paper



February, 2020

TAUS
the language data network

Baker McKenzie.

Authors: Wouter Seinen, Jaap van der Meer

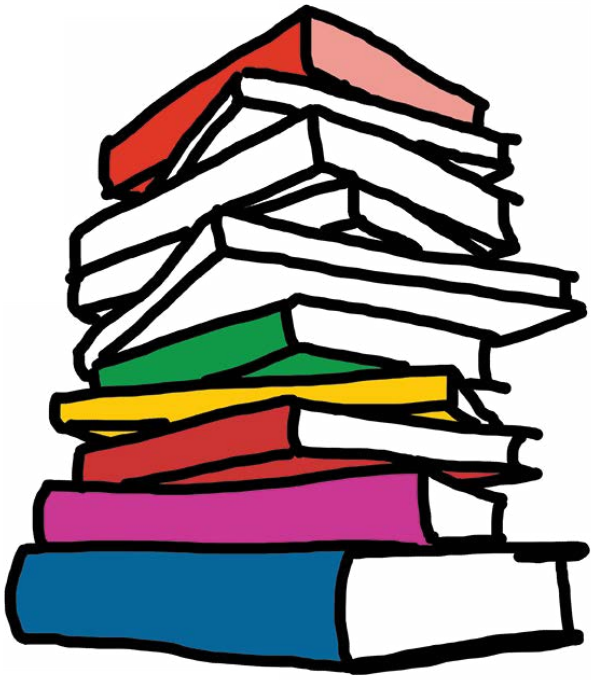
Reviewers: Andrew Joscelyne, Şölen Aslan

Published by TAUS Signature Editions, Danzigerkade 65A, 1013AP Amsterdam, The Netherlands
E-mail: memberservices@taus.net

www.taus.net

Design: Anne-Maj van der Meer

Table of Contents



Preamble

Who owns my language (data)? The title of this white paper is a bit of a teaser, because as everyone knows: nobody can own your language really, or anyone else's language for that matter. Language by its very nature is meant to be shared to help us, humans, to communicate and advance our evolution.

TAUS is a pioneer and leader in the space of language data. From the start of the TAUS Data Cloud in 2008, TAUS has worked with Baker McKenzie as its legal advisors. The IP/IT Commercial Practice of Baker McKenzie works with many large corporations consulting them on the use and sharing of data.

Throughout 2018 and 2019 Wouter Seinen, Partner at Baker McKenzie and Jaap van der Meer, Director of TAUS, have had many brainstorming sessions to address questions from users about the rules, regulations and best practices around sharing of data. Together they modernized the TAUS legal framework to accommodate the new use scenarios of Matching Data and the Human Language Project and to make them compliant with the latest privacy rulings.

In this joint White Paper Baker McKenzie and TAUS address concepts that are part of the legal foundation of sharing language data and share takeaways from the brainstorm sessions.

Introduction

Barriers to the free flow of data have various causes: the legal uncertainty surrounding the emerging issues on 'data ownership' or control, (re)usability and access to/transfer of data, and liability arising from the use of data.

There is a similar feeling of uncertainty in the translation industry, where language data are in great demand. This white paper is therefore intended to provide essential help for everyone who is actively involved in translation management or producing translations.

Chances are that you are sharing your translations, for instance by using cloud-based translation tools with integrated MT engines, as well as by exchanging translation memory files in email attachments and through file transfer. This is the very nature of the business. But from a legal point of view, does this matter?

Questions around privacy of language data are becoming particularly pressing now that Artificial Intelligence and Machine Learning are playing such a major role in business processes of all kinds. Buyers and providers of language related services find themselves increasingly working in not-business-as-usual circumstances: annotation, validation, crawling, clustering and the cleaning of voice and text data are becoming as important as traditional translation work. We are entering uncharted territory here. Which use cases are legitimate and which are unlawful? Are language data a proprietary good, and if so who owns what under which legal regime?

This white paper explains some international concepts of intellectual property law and data protection laws and applies them to language data and to the realities as we experience them in our daily practice. But, we need to be humble and realistic as well: intellectual property laws and data protection laws are not only complex, they also vary from country to country. For the data protection analysis the GDPR has been used as the starting point, as this regulation is known as one of the strictest privacy laws in the world and many other countries are implementing similar concepts in their own privacy laws. The Intellectual Property rights analysis is based on international treaties (such as the Bern convention, to which 177 countries are party), but also looks at European copyright law. The latter is in some respects rather strict or rigid as compared to copyrights in e.g. the US, whose "fair use" doctrine allows for more flexibility. We have followed the strictest regime, and as a result some of the below topics may be less of an issue if you only deal with data from contributors in, for instance the US or South America.

We hope that this summary will help to ease the minds of both practitioners and buyers of translation and advocate the productive use of data in the translation sector. We also recommend the Clarifying Copyright on Translation Data article published by TAUS on January 16, 2013.

1. Realities

1.1 The Translation Ecosystem: Naming the Data

Let's determine first what we are talking about here with some nomenclature and definitions. We propose to distinguish between *language data* and *translation data*. Language data are pure text data, bilingual or monolingual, such as source and target segments in pairs or text in a single language. Translation data, on the other hand, are data *about* the translations - i.e. metadata. Translation data are attributes such as who is the translator, the customer, what is the content type, the industry domain, how much time was spent on translating and editing the segment, type of edits, which CAT tool was used, which MT engine was used, date, and location. A lot of these translation data or metadata are collected by CAT tools, or they are added by a project manager during project set-up. The data points on productivity, edit distance, MT performance and translation quality as collected by the TAUS DQF plugin in translation tools are another example of translation data.

1.2 Translation Ecosystem: Naming the Actors

Let's also take a look at the actors involved in a typical translation transaction. It all starts with the organization that has a document and wishes it to be translated. We will assume that this organization is a legal entity which had the source text written and hence is the owner of (the copyrights in) the source text; we will refer to this actor as the "**Customer**". Then there is the translation agency, which receives the instruction to have the source text translated and will be referred to as the "**Agency**". Third, there is the individual creating or reviewing the translation. This can be an employee of the Agency but is often an independent contractor. We will call this individual the "**Translator**".

We realize that the three-layer, Customer-Agency-Translator, supply chain is a simplified representation of reality. Very often the Agency function is split between several global vendors who subcontract to regional vendors who then use freelance translators. It is also not uncommon for regional vendors to use sub-vendors. On the other hand, there are also many cases where customers contract directly with freelance translators. In this White Paper we describe the typical translation supply chain as a three-layer supply chain.

1.3 More Data is always Better

The reality today is that we live in a data-driven world, and this is especially true for the world of translation. Language data are used to train MT engines. *The more data the better*, is what you hear engineers often say. That was certainly true for the statistical phrase-based MT engines that have been most popular in the last decade. The new generation of Neural MT engines seem to be a little less data-hungry, but they thrive better on high-quality data and interestingly they also learn from the translation data (i.e. from the metadata).

The most common way of collecting language data has been through web crawling. This means that tools are being used to scrape text from translated websites and align the segments and apply some basic cleaning. From a technical perspective crawling websites is possible unless the website owners / operators have protected the content from their sites from being copied. Theoretically this may not always be allowed, but in practice website content is being used by website visitors as they like, as well as by search engines who cache and index content and translation businesses who wish to train their employees. Hundreds of billions of words have been collected through this process by all the big MT developers. Another way of collecting both language data and translation data is by simply using one's own translation memory data or asking permission from customers and platform-users to share their data.

A typical translation supply chain
with legal rights

CUSTOMER

Owns IP of source

Controller of PII in Language Data

AGENCY

License of IP

Processor of PII in Language Data

Controller of PII in Translation Data

TRANSLATORS EDITORS

Owns IP of target, usually transfers IP to customer

Sub-processor of PII in Language and Translation Data

2. Rules

2.1 Copyright on Language Data

If a source text has been written by a human being and the writing has required at least some creative choices, such a text is generally protected by copyrights. This means that many source texts that are later translated are copyrighted works. In most countries, copyrights come into existence when a text (or other work) is created. The copyright owner is typically the individual who authored the text, but in corporate environments, the copyrights are often vested in or are transferred to the company that employed or instructed the author.

The target text (i.e. the translation of a source text) is subject to multiple copyright claims: it is a modified reproduction of the original work, so the copyright owner of that source text will also have copyrights for the target text. In addition, the translator holds the copyrights of the translation.

In the translation business, however, ownership is often transferred to the Customer as part of the purchase of the service. This means that there are two possible scenarios. In the first scenario, the source segments are owned by the Customer and the copyrights in the target segments are owned partly by the Customer and partly by the Translator. In the second scenario (i.e. if the Customer has translators and the authors agree to transfer copyright), then the copyrights in the language data are solely owned by the Customer.

2.2 Copyright on Individual Segments

This raises an interesting question. When translated, a document is broken down into individual segments and stored in a database. Often the full document cannot technically be reconstructed on the basis of the segments alone. The question is: are the separate individual segments in that database still subject to copyright claims?

The starting point here is that copyrights are not tied to a particular form or shape; they can just as well apply to one chapter in a 1,000-page book as to the entire book. The lyrics of a pop-song may consist of only 10 lines of text, but clearly, they can be copyrighted. For segments of a larger text, the relevant criterion is whether a given segment qualifies as the author's "*own intellectual creation*."

Case law from the European Court of Justice has confirmed that a sequence of eleven words can be a copyrighted 'work', provided that from those 11 words one can still distinguish 'the hand of the author'. So copyright protection applies to that segment only if it is recognizable as the creation of the author. If the segment does not carry that signature, it is not protected by copyrights. Under US copyright law, a broadly similar originality test is used to draw the line. For a text to be copyrightable it must be "original to the author" and "possess more than a *de minimis* quantum of creativity". Short and commonplace phrases are not protectable, but phrases that do bear originality may be protected by copyrights[1].

Let's look at the lyrics of the famous song "Miss American Pie" to provide some examples. The lyrics were written by Don McLean and it is full of unusual choices of words, and hence a clear example of a text which is the author's "own intellectual creation". If we break down the song text into segments, it gets more difficult. The single line "*Drove my Chevy to the levee, but the levee was dry*" is clearly recognizable as Don McLean's creation, so and copying (or translating)

1 The debate around the protection of short and common phrases under US Copyright recently revived as a result of the court case regarding Taylor Swift's "shake it off" - see for example: http://www.lawjournalnewsletters.com/2020/01/01/when-are-short-phrases-in-songs-protectable/?slreturn=20200031165256

these eleven words will probably have copyright relevance. However, the hand of the master is less easy to recognize in the fifteen words of the line: "*And I asked her for some happy news, but she just smiled and turned away*". So this segment will, in isolation, probably not qualify as copyrighted work.

In practice many segments are far less poetic and will not be recognizable as an author's intellectual creation. Think of sentences like: "Please type in your password", or "Cookies are text files placed on your computer to collect standard internet log information and visitor behavior information." These can hardly be considered an original piece of art and will not pass the creativity threshold set forth in European copyright law.

2.3 Possible Exceptions for Machine Translation

As a rule, translations of a work are considered a "modified copy" of that work, or a "derivative work". This means that the act of translating a copyrighted text is also an act of infringement, unless an exception applies.

In the US, those who crawl the public Internet for language pairs that they can 'feed' to their MT algorithm, may sometimes be able to rely on the 'implied license' doctrine, or the 'fair use' exception[2]. To what extent this will succeed, is highly context-dependent. In any event, the Translator will have to state and prove that their use of the original text satisfied the criteria of the relevant exception. Theoretically, therefore, the MT business seems to be a risky one. In practice, however, we are not aware of significant damages being awarded by courts or of landmark cases which indicate that MT are illegal - despite the fact that MT has been on the rise for many years.

In Europe, the act of translating is almost by definition an act of infringement, provided that the text that was translated was indeed a copyrighted work. Copyright law in the EU is less flexible than US copyright laws, as the European doctrine is based on a limited set of exceptions, which are defined in the law. There is no general exception for infringements that do not cause harm and are commonly considered appropriate, or any other mechanism comparable to the US doctrine of 'fair use'. This leaves us with an even bigger gap between the rule of the law and everyday practice. As in the US, it seems that copyright holders are not very interested in suing those who (machine) translated their online publications, and even if they do sue, the courts appear to be unwilling to award damages. In other words, we are not aware of case law in which a Translator has been found liable for the single act of running a (machine) translation engine on a work that was publicly available on the internet.

The new EU Directive on Copyright in the Digital Single Market introduced an exception for 'text and data mining' (TDM)[3]. This can be explained as follows: "Text and data and data analytics methods extract data from existing electronic information to establish new facts and relationships, building new scientific findings from prior research. These new methods involve copying of prior works as part of the process to extract data"[4]. The introduction of this exception confirms that, in principle, TDM is a copyright-relevant act. In addition to the exception for research organizations, universities and cultural heritage applications, an

2 See for a more extensive analysis : and Prof. S. Yanisky-Ravid and C. Martens: "From the Myth of Babel to Google Translate: Confronting Malicious Use of Artificial Intelligence— Copyright and Algorithmic Biases in Online Translation Systems" (see; https://digitalcommons.law.seattleu.edu/cgi/viewcontent.cgi?article=2632&context=sulr), 2019; and E. Ketzan: "Rebuilding Babel: Copyright and the Future of Online Machine Translation", 2007, https://journals.tulane.edu/TIP/article/view/2529

3 The Directive on Copyright in the Digital Single Market entered into force on 6 June 2019.

4 De Wolf & Partners, Study of the legal framework of text and data mining, March 2014, p. 6 (available on https://publications.europa.eu/en/publication-detail/-/publication/074ddf78-01e9-4a1d-9895-65290705e2a5/language-en).

exception is also introduced for the TDM of legally accessible works owned by other users than research institutions, going beyond the purely research domain[5]. This exception should help contribute to the development of data analytics and artificial intelligence in the EU[6]. In some Member States such an exception has already been adopted, but only within the scope of the permissible research exception[7]. This exception for TDM for purposes other than scientific research, could possibly apply to a broader array of (machine) translation operations, including web-crawling for data to train translation algorithms.

2.4 Copyright in Practice

For a Translator or Agency it is hard to tell whether a segment is copyrighted. Most segments - in isolation - do not fall within the scope of European copyright law concepts, either because they do not meet the creativity threshold or because the work it was derived from is public domain in the first place.

If a segment *does* fall under copyright, the Translator needs to have permission from the copyright owner. This process is often referred to as 'copyright clearance'. Typically, the Translator and the Agency will, in their respective terms of service, ask that the Customer takes care of any copyright clearance.

Should a copyright owner find out that a Translator has copied and translated their work and consider this problematic, they could demand that the language data is deleted. In theory, this owner could also claim damages, but in practice courts (at least in Europe) are unwilling to award financial damages if the copyright owner cannot demonstrate that they have actually suffered economic loss from the unauthorized use of their work.

This paradox is a feature of the digital area: the Internet is full of audiovisual and other content that has been copied from other sources, and in many cases no consent was sought from the original rights holders. Famous examples include Google Books, Facebook and Flickr, which contain millions of text documents, photos and videos that are clearly proprietary, and yet are still published online without a paper trail of the consent of the rights holders. In the vast majority of all these cases nothing happens. And if and when a rights holder has a problem with their work being copied and published, this is generally solved through a "notice and take-down" process: the platform providing access to the works verifies the claim and subsequently ensures the work is no longer accessible.

2.5 Data Protection Law Considerations for Translators

With the rise of modern data protection laws, spearheaded by Europe's new General Data Protection Regulation (GDPR), launched in Europe in 2018, data-driven businesses need to understand what sort of privacy laws their activities are subject to and whether they comply with these.

We will summarize a number of key data protection law concepts, and then briefly plot these on the two types of data that we have distinguished in this paper - "language data" and "translation data".

5 Article 3 and 4 Directive on Copyright in the Digital Single Market.

6 European Commission, Press release: Digital Single Market: EU negotiators reach a breakthrough to modernise copyright rules, 13 February 2019.

7 ELRA, ELRC Report on legal issues in web crawling, 22 March 2018, p. 19 & 29.

2.5.1 Scope of privacy laws: "PII" or "personal data"

Data protection laws are about data relating to private individuals. The GDPR speaks of "personal data", whereas privacy laws in the US typically concern "personally identifiable information", or "PII". The GDPR concept of "personal data" is broader than that of PII, as it is defined as *any information that relates to a directly or indirectly identifiable natural person*.

This also includes types of data that are not in the scope of the "PII" definition, such as, for example, information relating to an individual in their capacity as the representative or owner of a company, as well as data that only indirectly says something about a living individual, such as the IP address used, or their license plate number or job title.

2.5.2 'Lawful basis' / Processing Ground

How does the GDPR apply to our crucial concepts of *language data* and *translation data*? The GDPR applies primarily to translation data and specifically to the names and other identifiable data of the translators, reviewers, project managers and clients insofar as this PII is collected as part of the translation data. All of these persons must, according to the GDPR, be made aware of the recording and storage of their personal data by the technology or platform owner who is collecting the data. In addition, the party that is collecting the data must also ensure that in doing so, it can rely on one of the "processing grounds" (i.e. legal reasons) in the GDPR.

Consent given by the individual is the best-known processing ground. However, other important processing grounds include the fact that the processing is necessary for the performance of a contract or for the compliance with statutory laws. Finally, many personal data processing operations are based on the legal ground that the processing is in the *legitimate interest* of the data controller, provided that that interest is not overridden by the privacy interests of the individual. This would, for instance, apply to everyday processing operations, such as the storage of someone's IP address or cookie data for the shopping cart, or the language preferences of a website.

2.5.3 Transparency

Individuals whose data are being processed by others have the right to know this fact, and for that reason the GDPR requires that those who process personal data disclose this. They should be transparent about what data are being used, for what purposes, and how long they are stored. They should also inform these individuals about their rights. It is not necessary to inform others of processing activities that are already known to the individual. Clearly, if you receive an email from someone, you do not have to actively tell this person that you will store their message on your computer. However, if you would do things with the email that the sender could not reasonably have expected, then you should inform them.

Many websites and online companies try to satisfy this requirement by publishing a "privacy policy" or "privacy statement". This would be a sensible step for most players in the translation ecosystem as well. Actively informing the individuals whose data are being translated will be difficult for the Translator, as they will typically not have the means to identify who the individuals are, nor the contact details to reach them. Where this is the case, the online privacy notice is probably the best alternative solution. It seems fair to expect that courts and regulators will understand this.

2.5.4 Data Minimization, data retention and 'privacy by design'

Two other key concepts of data protection law are "data minimization" and "privacy by design". The data minimization principle essentially means that an organization must always ask the question as to whether collecting and keeping all data points is really necessary. In other words,

if the purpose can also be achieved while using 50% of the data points, only that 50% should be collected. Similarly, once particular records are no longer necessary, they should be disposed of.

A simple example is the common situation where an app wishes to ensure that its service is only used by individuals aged 18 and over. On sign-up the app could ask each user to fill in their birthday. However, the age verification could also take place by simply asking the user how old they are, or in what year they were born. If the app wishes to be sure that users who just turned 18 are not excluded, it could ask for the full day of birth, but only use these data points to verify the age limit. Once the age is calculated, the full birthday information is deleted or replaced by a less-detailed data point, such as "18 - 25 y". This would also be an appropriate data minimization measure.

The privacy by design principle takes this one step further and requires that organizations consider the privacy impact when changing or implementing new systems and/or processes. This principle is relevant for those who use or implement translation software and SaaS services. They should assess how the system can be configured in the most 'privacy friendly' manner, e.g. avoiding the collection and/or sharing of data points if not strictly required.

2.5.5 Translation Data

Translation data typically contains considerable information about the individual who created or curated the translation. The Translator is the "data subject" and the Customer, as well as any platform operators and translation agencies, can only process such data in compliance with the GDPR. The owner of the technology or the platform that collects and hosts language data is specifically responsible for compliance with GDPR insofar as it concerns the personal data pertaining to the users of this platform or the technology that is collected when operating the platform. The Customer and the Agency will have similar obligations that apply to the data they collect and keep on record.

This means that all these actors must, in respect of their own personal data processing, *inter alia*, ensure they have a proper processing ground, provide information about how the information is used and how long it is stored, and have processes in place to deal with requests from data subjects to access or delete their personal data.

2.5.6 Language Data

Language data can contain personal data, but whether this is the case will be hard to tell for the Translator. Is "Harry Potter" the name of a living individual, an invented or joke name, or the name of a deceased person? In the first case, the GDPR would apply, whilst in the other two cases the processing of this data element will remain out of scope for the GDPR.

A Translator will mainly work for and on behalf of their Customer and expect that the Customer has a legal basis for the processing of any personal data in the source text. That said, the Translator may also want to keep the source and target text files in their own database, for their own business purposes. For this - secondary - processing, the Translator will be personally responsible for GDPR compliance.

2.6 Mind the Regional Differences

Please note that data protection laws differ from region to region. The above analysis is based on the GDPR, i.e. the privacy laws applicable in the EU[8]. The GDPR applies to organizations that are located in the EU, but also to non-European organizations to the extent they are processing personal data of individuals that are located in the EU, when this operation takes place in the course of a contractual relationship, or entails the systematic tracking of their behavior. This may mean that the Agencies and Customers that use translators who work out of Europe come within the scope of the GDPR.

Many other counties, from Japan to Brazil, and from New Zealand to Canada have privacy laws that are based on the same principles. The scope of federal privacy laws in the US is somewhat more limited, but US based organizations may also be subject to the data privacy laws of their home state. The Californian Consumer Protection Act ("CCPA"), for instance, has introduced concepts which are similar to the GDPR, but enforced in a different way.

8 For more information on the general requirements of the GDPR and guidance on how to cope with these, please refer to: "The Guide to the General Data Protection Regulation (GDPR)" of the ICO, available at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/

3. Recommendations

Translation inherently triggers questions about intellectual property rights and data protection laws. As the translation ecosystem is complex, it is not easy to draw simple conclusions on who is responsible for what and which use cases are legitimate or not. Expert advice is required.

But, at the end of the day when the lawyers have gone home, we as professionals in the translation industry have to use our own common sense and do what's right. We have to ask ourselves very practical questions and follow a set of simple rules to reduce regulatory risk and enhance our compliance. Here are the most typical cases:

Q1: Wouldn't our customers expect us to leverage our knowledge in order to deliver the best possible service at competitive prices? And surely that knowledge includes our experience and data stored both in our brains and in our computer systems?

A: Of course. The translation situation is not so different from other professions (lawyers, accountants, and consultants) that almost certainly store documents and reuse parts wherever it makes most sense. So make an inventory of the type of data that is being processed by your company and work out for yourself which data elements are being processed for your own purposes, and which processing is done solely for someone else.

Example: Let's assume that a Customer asks its Global Vendor to translate a group policy document. The Global Vendor forwards the document to a Regional Vendor, who in turn forwards the document to a Translator, who runs a machine translation, and then reviews, edits and curates the document. The Original and the Translated document are sent back via the same route. Obviously, the Translator is allowed to copy the document for the creation of a translated version. This is what he was instructed to do.

The Translator may want to keep the file for future reference, benchmarking and other analytics. This may not be something the Customer has instructed the Translator to do, but it makes perfect sense for the Translator to do this for her/his own purposes. Whether this is legitimate depends on what the parties have agreed and on the Translator's own data protection controls. Generally speaking it is hard to argue that the Translator cannot use his/her translations for the training of engines, as long as the Translator takes care that personal information is removed.

Q2: Is there a precedent of a translator or a translation agency being penalized or imprisoned for using data for which they perhaps did not own the copyright?

A: Most unlikely but possible. From an IP and a Data Protection perspective, always follow up on any complaints in a responsible and respectful way to reduce the risk of legal issues. Check to see if there are any cases within your community so that you are fully prepared if such a question arises.

Q3: How exactly do I use "my" language data? Does my method damage the interests of the customer in any way? Or could it possibly damage the interests of persons whose names may appear in the language data?

A: We suggest making an inventory of the types of data that are being processed and work out for yourself which data elements are being processed for your own purposes, and which are processed solely for someone else. In respect of the data you are processing for your own purposes, you will have to walk the extra mile to ensure you comply with data protection laws.

The data you process solely for the Customer is much less a concern, as it is typically the Customer's responsibility to ensure the translation can be made without infringing on the rights of others.

Q4: When we come across personal information in language data, what should we do?

A: Check your policies and legal notices to see how transparent your organization is about how personal data is processed and what the policies say about respecting intellectual property rights. Assess whether a given case really concerns sensitive personal information as opposed to information that can be generally known. In the latter case, there is not much to worry about. If you are using personal data that was already available online, the GDPR may still apply, but it is safe to assume that having such data in a language data set is not in breach of the GDPR if you have taken the measures that are expected from any other organization, such as data minimization, transparency, and measures to enable individuals to make inquiries, lodge a complaint or exercise other rights conferred on them.

Q5: Who and what are the GDPR and other data protection laws intended to protect?

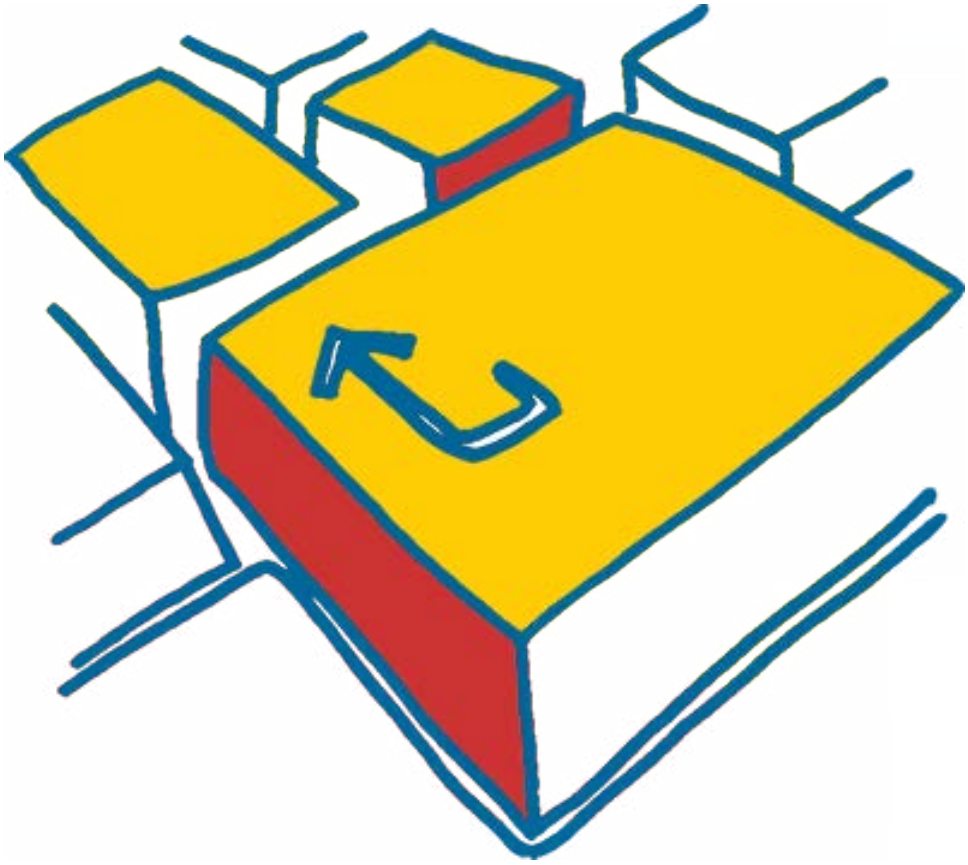
A: Data Protection laws are essentially designed to protect citizens against governments and organizations that want to use their data. The GDPR is a clear example of a law that aims to "empower" individuals. The thinking is that stronger rules on data protection mean that people have more control over their personal data and businesses benefit from a level playing field. So the law is not against the use of personal data. Instead, it seeks to promote the processing of personal data in a more transparent and responsible manner - and in particular to give individuals more say in and control over how their data is used.

Conclusion

We have tried to apply the essentials of Europe's GDPR to the translation ecosystem and to touch on some intellectual property issues as well. Although we have tried to be comprehensive, the key takeaway is that both areas of law do not provide black and white answers to questions that appear to be simple and straightforward. Seeking clarity around what data you process and for whom is an important first step for any organization. Common sense, in combination with some essential rules of thumb will help getting grips on legal compliance, whilst IP and data protection laws gain relevance.

The best way forward is to assign someone to monitor legal developments and best practices and inform your workforce and other stakeholders clearly about data ownership issues as they evolve.

While it is impossible for most organizations to remain completely outside the scope of the laws, there do not seem to be legal precedents that pose a serious threat to the general translation business. Moreover, the typical translation business model is not incompatible with most IP and data protection laws - most processes can be designed or adapted to enable all actors to fully comply with the law. Organizations should therefore pay attention to their processes and compliance controls, but overall, the industry can sleep fairly soundly on this issue!



Authors

Wouter Seinen is a partner and the head of the IP/IT & Commercial Practice Group at Baker McKenzie in Amsterdam. He has significant experience in assisting national and international clients with respect to issues concerning ownership and protection of electronic data. A significant part of his practice involves data protection law and associated compliance matters, such as data security and data breaches. GDPR projects ran by Wouter's team often have a significant international angle and / or deal with innovative products or services.

About Baker McKenzie

With 77 offices in 46 countries and over 6,000 lawyers worldwide, Baker McKenzie is one of the largest law firms in the world by headcount and revenue. As a truly global full service firm, Baker McKenzie has always has a strong focus on technology, media and communications law. One of its objectives is to stay 'ahead of the curve' and this is why Baker McKenzie has a particular interest in teaming with organizations and institutions that operate at the fore front of innovation. The lawyers and advisers at Baker McKenzie have a deep understanding of the culture of business the world over and are able to bring the talent and experience needed to navigate complexity across practices and borders with ease. Baker McKenzie is trying to be different from other law firms by combining an instinctively global perspective with a genuinely multicultural approach, enabled by collaborative relationships and yielding practical, innovative advice.

Jaap van der Meer was the founder and CEO of some of the largest global translation and localization service companies in the 1980s and 1990s. In 2005 he founded the Translation Automation User Society (TAUS). TAUS is an innovation think tank and platform for industry-shared services for the global translation and localization sector. Many of the largest IT companies, government translation bodies and their suppliers of translation and localization services and technologies are members of TAUS. TAUS offers among others a platform for translation quality evaluation and benchmarking and a platform for pooling and sharing of translation memory data. Jaap van der Meer has written many articles over the years about the translation industry.

TAUS, the language data network, is an independent and neutral industry organization. We develop communities through a program of events and online user groups and by sharing knowledge, metrics and data that help all stakeholders in the translation industry develop a better service. We provide data services to buyers and providers of language and translation services.

The shared knowledge and data help TAUS members decide on effective localization strategies. The metrics support more efficient processes and the normalization of quality evaluation. The data lead to improved translation automation.

TAUS develops APIs that give members access to services like DQF, the DQF Dashboard and the TAUS Data Market through their own translation platforms and tools. TAUS metrics and data are already built in to most of the major translation technologies.

For more information about TAUS, please visit: https://www.taus.net

TAUS
the language data network