MRS Evidence Matters

Report

TRANSLATION AND
TRANSCRIPTION
PROCUREMENT:
DATA PRIVACY AND
COMPLIANCE

October 2024

Introduction

This document is designed to create a shared starting point between research practitioners and the language service community for buying translation and transcription services.

It covers vocabulary and core areas to clarify communications in the glossary of terms and looks in depth at some of the more common areas to consider during the translation and transcription procurement process such as data privacy and translation memory.

We wish to express our thanks to Derek Jones from Acuigen and Ruth Partington from EMPOWER, for their drafting of this document, and Raisa McNab, CEO of the Association of Translation Companies for her input to this report.

Please see the MRS/ATC Checklist for Buying Translation Services for detailed guidance about the use of translation agencies.

Please note: **this guidance is provided for information. It is not legal advice and cannot be relied upon as such. Legal advice should be taken in relation to any specific legal problems or matters.**

Glossary of terms

Term	Description
AI	Artificial Intelligence. Specifically in this context, AI is applied much like MT (see below).
Client	Any individual, organisation, department or division, including any belonging to the same organisation as an MRS Member, which is responsible for commissioning or applying the results from a multilingual research project *(source: MRS Code of Conduct).*
Confidential Information	Information that is not intended to be made available or disclosed to unauthorised individuals *(source: ISO 27001:2022)*
Computer-aided translation tool (CAT tool)	Software often used during the translation process. *Note:* CAT tools incorporate the use of Machine Translation, AI and Translation Memories in the translation process.
Controller(s)	The natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data *(source: ICO).*
Data Processing Agreements (DPA)	A written agreement between Controllers and Processors to ensure they understand their obligations, responsibilities and liabilities. Contracts also help them comply with the UK GDPR and assist Controllers in demonstrating to individuals and regulators their compliance as required by the accountability principle.
Data Subject	The identified or identifiable living individual to whom personal data relates *(source: ICO).*

IPR	Intellectual Property Rights.
Language Services Provider (LSP)	Any individual, organisation, department or division that provides multilingual translation or transcription services.
Large Language Models (LLM)	An AI model used for general-purpose language generation.
Machine Translation (MT)	Automated translation of text or speech from one language to another. *Note:* Machine Translation (MT) solutions range from public, generic solutions to customised solutions for specific subject matters and language pairs. MT can be applied where there is no previously translated content. *Note 2:* Common MT solutions are DeepL, Google Translate, and Microsoft Bing.
Non-Disclosure Agreement (NDA)	A legally binding contract that establishes a confidential relationship between parties signing the agreement. An NDA can be agreed between 1. Client and the Research Agency, or 2. Research Agency and the LSP, or 3. Client, Research Agency and LSP (three parties agreeing to a single NDA). Further information about NDAs is available via the MRS Guidance Note about NDAs.

Post-Editing, Machine Translation Post-Editing	A reprocess of editing and correcting output automatically generated by a machine translation engine.
(PEMT or MTPE)	*Note:* MTPE is typically used to refer to a process which does not include the use of previously translated content from a Translation Memory. A process which includes both previously translated content from a TM and machine translation output carried out in a CAT tool can be described as a multi-modal translation environment.
	Note 2: AI can equally be used in this context for AI post-editing.
Processor(s)	A natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller (source: ICO).
Research Agency	Organisation that undertakes a research project.
	Note: a Research Agency can also be an in-house researcher or an independent consultant.
Source Document	A file in the original language.
	Note: Source documents may be in any format; are typically Word or Excel documents or exported program files such as XML.
Target Document	Translated or transcribed file.
Transcription	A set of processes to render content from audio-visual or spoken format into written format.
	Note: Transcription is typically used to denote a monolingual process, e.g., transcribing audio in English into written format in English. A transcription process may also incorporate a translation process, e.g. transcribing and translating audio in French into written format in English.

Translation	A set of processes to render content from one language (source) into another (target) in written form.
Translation Memory (TM)	A database of previously translated source and target language segments, presented to translator in a CAT tool. *Note:* CAT tools typically split the source content into sentences or phrases based on grammatical punctuation rules. The CAT tool stores the source text and translation into translation units or 'segments' in the TM for reuse in future translation projects. *Note2:* The most common CAT Tools are SDL Trados, MemoQ, and Phrase.

Executive Summary

1. Personal data can be processed (translated) and stored using CAT tool by Processors (LSPs) provided that this is permitted by Data Processing Agreements, the resulting Translation Memory is suitably protected, content does not leak to third parties, data subjects have consented to its use and the content can be deleted after the data has been processed for the stated purpose of the research.

2. Ownership and the Intellectual Property Right (IPR) of content protected by data protection legislation (e.g., Source Documents, Target Documents) rests ordinarily with the Controller(s) of the content; or as may be specified in another written agreement.

3. Personal data cannot be held in a Translation Memory by a Processor (LSP) and repurposed or sold to unrelated third parties except with the explicit permission of the Controller and the data subjects. Note that an LSP's lack of permission to use automated processes (Translation Memory or a CAT tools) will impact the LSP's ability to automate processes, manage quality and costs.

4. Confidential Information subject to NDA can be translated in a CAT tool with the approval of content owners, provided that there are suitable contracts/technical measures in place to protect the content from leakage/inappropriate disclosure.

5. Confidential Information subject to NDA should not be used to enhance or train Translation Memories that will be accessed by third parties, unless the owner (of the confidential Source Document) has agreed that it can be used for this purpose.

6. Confidential Information subject to an NDA may not be put into a Translation Memory and shared with third parties who are not subject

to the NDA without the consent of owner of the confidential Source Document.

7. Commercial contracts are likely to determine when payment takes place for any services for translation/transcription. Only Controllers, data subjects (i.e. the participant) or the owners of Confidential Information can consent to how data will be processed, not the Processors, who act only upon the instruction of the Controllers or owners.

8. Where an LSP uses AI, MT or LLM translation or transcription services provided by third party solutions such as DeepL, Google Translate or Microsoft Bing, they must ensure that any data usage or sharing complies with relevant agreements with the Controller and ensures the protection of Confidential or Sensitive Information, and where necessary seek the Controller's approval for data usage or sharing. Any participant personal data processed in third party solutions should be reflected in relevant participant project privacy notices.

Personal data privacy assessment & content profiling

In all research projects, keeping personal data safe and secure is paramount to the integrity of projects. Assessing personal data privacy requirements and profiling content ahead of commissioning a translation or transcription process ensures that data is processed in accordance with the GDPR throughout the supply chain.

Important:
The assessment below applies equally to translation AND transcription procurement.

When a Research Agency commissions an LSP to translate or transcribe content, the Source Documents or audio-visual content will usually contain one of the below three distinct types of information:

- Non-confidential/non-commercially sensitive content (e.g. sales literature, brochures) that does not contain personal information about data subjects and therefore data protection legislation does not apply, nor is confidential material;

- Confidential or commercial content (e.g. confidential questionnaires or confidential discussion guides, commercial contract, IP designs, stimulus material) that will often be the subject of Non-Disclosure Agreements (NDA) that is unlikely to contain personal information about data subjects, and therefore data protection legislation would not apply; or

- Content that contains personal information (e.g. customer feedback, personal opinions, private statements, focus group content, interview transcriptions). Some of this content may also be sensitive content, e.g. private political views or private medical records, is subject to local privacy laws and may also be subject to an NDA. Within the translation community 'sensitive information' should also carry a higher level of protection.

When Source Documents or audio-visual content that contains personal data is sent for translation or transcription to an LSP, the content will always have at least one Controller (who is also typically the owner of the content) and all parties involved in the translation or transcription (who are not Controllers) will be acting as Processors (or sub-processors) acting upon the instruction of the Controller(s).

Note: A Processor cannot make decisions on the usage of the data, nor can a Processor use the content for purposes that have not been approved by the Controller/s (and implicitly the originating Data Subject, where applicable).

Important note: A Research Agency handling personal data must have a Data Processing Agreement (or data processing clauses as part of another agreement) in place with any sub-processors (including any LSPs) which includes appropriate consent to process the data. Some LSPs may have their own DPA/privacy policy. In all circumstances, Research Agencies will need to agree DPAs with the LSPs they commission.

The following table summarises the challenges in managing content for translation or transcription.

Potential Usage Scenarios

Scenario	Data Protection implications	Can the Source Documents and translations be stored in the Translation Memory for reuse [after the project ends]?	Can the Translation Memory be used by other language service providers in the future?	Data Retention Policy
Research Agency passes content that is not Confidential to an LSP for translation, transcription, or analysis	Not applicable	Yes, subject to the commercial terms and conditions agreed with the parties.		Subject to the commercial terms and conditions agreed with the parties.
Research Agency passes Confidential Information to an LSP for translation, transcription, or analysis. There will be a written confidentiality clause in the contract(s) between the Client, Research Agency and LSP; this may also be documented in an NDA.	Not applicable	Data leakage must be avoided. This means that the Source Document, Target Document and resulting Translation Memory need to be protected from data leakage. If a contract/NDA has been agreed, its scope should determine who is allowed access to the Source and Target Documents and any resulting Translation Memory holding. The contract/NDA may also determine who owns the resulting Translation Memory, whether it can be reused and in what context, and who is permitted access to it. There should be clarity in the clauses of the contract/NDA about: - Retention period for the Source and Target Documents and any resulting Translation Memory. - Whether the Translation Memory containing the		Subject to the commercial terms and conditions agreed with the parties.

Scenario	Data Protection implications	Can the Source Documents and translations be stored in the Translation Memory for reuse [after the project ends]?	Can the Translation Memory be used by other language service providers in the future?	Data Retention Policy
		Confidential Information can be shared with third parties, during or after the project ends. - Where the Translation Memory is stored.		
Research Agency passes content that is identified as containing personal data to an LSP for translation, transcription, or analysis	The source and target documents, and any resulting Translation Memories are processed subject to local data privacy laws compliance (GDPR etc.).	Possibly, if the personal data can be anonymised or pseudonymised. Retention of personal data without anonymising or pseudonymising it would require the permission of the Data Controller and the permission of the data subject as permitted use of the data. Refer also to the notes above about the contract/NDA – the rules are the same.	Only with the consent of the Data Controller for the purposes of the project. Protected Content should not be disclosed to unauthorised parties. Refer also to the notes above about the contract/NDA – the rules are the same.	As determined by the Data Controller for the purposes of the project which will be documented in a data processing agreement/ contract.