

***TM Marketplace White Paper***  
***Sharing Translation Memory Data***  
***Aligned from Third-Party Documents:***  
***Legal Considerations***

## **The Process**

TM Marketplace accesses pre-existing, third-party documents in multiple languages, segments the documents, and matches the segments in the different languages (alignment) with the goal of creating a bilingual or multilingual database of matching segments (translation memory). The length of the segments is restricted to ten (10) words (or any equivalent thereof) in the source portion, and the resulting database is in alphabetical or another order that bears no similarity to the source text or any non-trivial component of it. Consequently, it is virtually impossible to use the database to generate anything substantially similar to the source documents or any substantial component of it. The resulting database is commercially distributed for use in developing translations of other pre-existing, independently generated texts.

## **The Legal Situation**

### **Copyright**

While US copyright law provides that the owner of the copyrights in a document generally has the exclusive right to generate and distribute copies or derivations of a copyrighted document, the copy or derivation must be at least substantially similar to a substantial portion of a protectable expression in the original work or, in the case of a translation, to a translation of the original work.<sup>1</sup> In this regard, facts recited in an original work are not protectable in themselves; rather, it is only an author's original analysis, the structuring of material, and the marshaling of facts that can constitute protectable expression of facts.

Similarly, short phrases and slogans are not copyrightable in themselves because individually they do not provide sufficient originality to be protected by copyright. In addition, phrases and expressions conveying an idea typically expressed in a limited number of stereotyped fashions are not subject to copyright protection.

Since the distributed database is comprised of short phrases of no more than 10 words of factual or technical information, each such phrase is not copyrightable because it is insufficiently original and protectable under the copyright laws. The resulting compilation of such phrases is re-ordered in the translation memory database, and the

---

<sup>1</sup> For supporting court decisions on this and all other facts cited in the White Paper, please contact [info@tmmarketplace.com](mailto:info@tmmarketplace.com).

translation memory database is not provided for the purpose or with the practical result of generating any substantial portion of the original text; instead, the purpose is to generate a translation of other, independently generated text. Therefore, the translation memory database, and any translations generated with it are not likely to ever be substantially similar to any protectable expression in the source documents. As a result, the translation memory database and its intended use will not constitute or result in any infringement of copyright in any source documents.

### **Fair Use**

The downloading of the source documents, the use of them in an intermediate database-generation step, and the generating of the downloaded copy are likely to be authorized and thus would not constitute a copyright infringement. The same is also true of the use of a given downloaded copy and the use of the segments from the copy in the intermediate database-generation step.

Assuming, however, that the downloaded copy and the intermediate database-generation step are not authorized by the copyright owner, they still constitute fair use under the copyright law since the entire purpose is to generate an instructional, factual translation memory database that will not provide any meaningful amount of protectable expression in the database or through use of the database.

The purpose of the downloaded copy and the intermediate database-generation step is solely to generate works – the translation memory database and works to be generated with it – that do not supersede the object of the original creation. Instead, they provide a new work with a wholly different purpose and character. This factor supports a finding of fair use.

### **The Nature of the Source Texts**

In regard to the nature of the copyrighted work, it is generally recognized that there is a greater need to disseminate factual works than works of fiction or fantasy. Here, the source documents are factual and technical. As a result, this factor also constitutes a finding of fair use.

### **The Market Value**

In consideration of the use in the potential market for the value of the copyrighted work, there is no harm to any market for the original works in issue or protectable derivatives of them that would flow from the proposed use of these documents. For example, most if not all of the source documents will be instruction materials that are downloadable on the web. Consequently, the proposed use and resulting internal document copy and databases would not damage any market or reduce any revenue to the copyright owner for such source documents. Thus, this factor too supports the fairness of the proposed uses of the source documents.

## **Conclusion**

The above-described process of downloading documents, aligning them into translation memory databases, and their subsequent sale

- does not infringe any copyright since the original text cannot be recreated;
- does not violate the principles of fair use since the purpose of the translation databases is completely different from the original document; and
- does not compete with their use and their market value.