XLIFF vs. IDML

Advantages and disadvantages of these formats in translation

Introduction

Translation of InDesign documents is not possible in the proprietary .indd format.

In any case, an intermediate step must be carried out - either

1. Export of the contents as text or
2. Export the document as IDML or
3. Export of the contents as XLIFF (ex TranslationFilter) or
4. Use of a plug-in for direct integration of InDesign into a TMS

The advantages and disadvantages of the individual options are described below.

Textexport

In InDesign, the text is always in text frames or text chains (several linked frames) that are grouped together or grouped with other objects, arranged anywhere on the pages or on different layers. Since there is no text export per se for all text content in InDesign, the single frames or text chains must be exported manually each as a single file. In principle, this can be automated by scripts or 3rd party plug-ins, but in the end a significant effort has to be made regarding to translation, because the translated texts have to be re-imported just as manually. Depending on whether you only use plain text or formatted text (RTF or ID tags), the formatting is usually lost and has to be manually generated again.

IDML

In the year 2008 Adobe introduced the InDesign Markup Language (IDML) as the data exchange format in InDesign CS4. An IDML file describes a complete InDesign document, i. e. its contents, geometries, definitions (colors, formats, typography, specifications, etc.). By using this exchange format, for example, you can open a document from a higher version of InDesign in an older version. However, the data that is not recognized by the older version (e. g. new features, etc.) are deliberately lost.

This format is designed by Adobe for data exchange between different InDesign versions and not as a format for integrating third-party systems.

Some (by far not all!) TMS systems have integrated this exchange format - with their own and thus different routines for interpreting and manipulating the contents. As a result, InDesign documents behave differently depending on the TMS and can therefore be translated.

Finally, they are manipulations of a "dead" file, which is modified outside the original intended purpose. As a result, it happens again and again that the translated IDML files cannot be opened at all (because corrupt) or side effects occur because the structure of the InDesign document has been unintentionally changed (in places you wouldn't suspect). There are no protection mechanisms for the integrity of the document. Formatting is very often lost during translation.

Another disadvantage of this format is that an IDML file always describes a complete InDesign document - regardless of whether this document contains individual articles that belong together in terms of content and should therefore also be translated in this context.

Also the order of the texts in an IDML file does not follow the content structures that the reader expects, but is technically conditioned. Headers and footer elements (such as page numbers) are mixed with the content and structures such as footnotes, links, index entries, etc. riddled the translator.

IDML files do not contain any images, but are still relatively large, which requires a corresponding bandwidth for transmission with online TMS.

An IDML file inevitably becomes a new InDesign document after translation - without "knowledge" of the original document. This means that when using the translated document from an editorial system, the translated document must be checked in again and all metadata, etc. must be re-entered, articles re-created and images re-linked.

XLIFF

The "XML Localization Interchange File Format", abbreviated XLIFF, is an XML format for displaying hierarchically structured content data in CAT tools.
The XLIFF format is specially designed for the low-loss exchange of translation data and associated context information. It is XML-based and therefore extensible.
Virtually every TMS "understands" this format because it is the standard in the translation industry.

In addition to the content to be translated, an XLIFF file also contains some metadata such as

- The source and target language of the translation
- The version for a clear description of the available options
- The tool used to create / process this file
- A lot of other information

_____

An XLIFF file always contains the content in so-called "segments" whose content depends on the configuration of the tool. For example, a segment can contain only one word, but also a sentence (or partial sentence) or an entire paragraph - but never a return.

Since segments are basically the basis of a "translation memory" and therefore represent the reuse of already translated content, segmentation is an important factor in the data exchange from / to the TMS.

The ex TranslationFilter exports an XLIFF file based on a detailed configuration in order to optimize the reproducibility and thus minimize the costs of the translation. This configuration is encoded into the XLIFF file and thus survives the "round trip" in the translation process. Segmentation can be optimised by means of a comprehensive set of rules. So-called inline tags are also used to describe formatting within a segment, for example.

This makes **ex TranslationFilter** the only tool that exports the content from a "living" environment (the running InDesign process with all its interfaces for the plug-ins) in the desired form / structure and imports it again after translation - while protecting the document structure and maintaining all formatting, making manual reworking virtually unnecessary.

Necessary functions are also supported, such as

- Changes in the writing direction (left-to-right versus right-to-left for languages such as Arabic, Hebrew, etc.) - including the conversion of the document.
- Automatic font substitution depending on the target language
- automatic adaptation of the texts to avoid overset
- Conversion of the text language in the layout to have the correct hyphenation and spelling active immediately
- Filter for exporting / importing layers
- Creation of additional language-dependent layers (multilingual documents)
- Automatic recognition and omission of page numbers etc.
- Index directory support
- automatic removal of duplicate blanks, soft returns, spacing, etc. which have only been inserted for "layout-technical" reasons and make no sense in the translation

These are unique selling points of the **ex TranslationFilter**.

When embedding in editorial systems such as WoodWing or Vjoon with "Smart Translation Solution", article structures are recognized and supported, i. e. article-based work is possible - without having to manually restore the external links to images and articles (as with IDML). Images can be treated as language-dependent or language-neutral.

Comparison Table

	IDML	XLIFF
Content is transferred segmented	✘	✓
Segmentation rules adjustable	✘	✓
SRX format support	✘	✓
Document structure / integrity protected	✘	✓
Compatible with all TMS	✘	✓
Small file size	✘	✓
Article-based working possible	✘	✓
Layer-based working possible	✘	✓
Support of ISO codes for languages	✘	✓
Transfer of metadata to the TMS	✘	✓
Changing the writing direction for RTL languages	✘	✓
Automatic replacement of fonts in the target language	✘	✓
Automatic conversion of the text language in the layout	✘	✓
Index support	✘	✓
Automatic text filter for double spaces etc.	✘	✓
Automatic removal of kerning, tracking etc.	✘	✓
Automatic creation of a PDF for the translator	✘	✓
Support of variants (WoodWing)	✘	✓