



XLIFF vs. IDML

Vor- und Nachteile dieser Formate in der Übersetzung

Einführung

Die Übersetzung von InDesign-Dokumenten ist in dem proprietären Format .indd nicht möglich.

Es muss also in jedem Fall ein Zwischenschritt erfolgen – entweder

1. Export der Inhalte als Text oder
2. Export des Dokuments als IDML oder
3. Export der Inhalte als XLIFF (ex TranslationFilter) oder
4. Nutzung eines Plug-Ins zur direkten Integration von InDesign in ein TMS

Die Vor- und Nachteile der einzelnen Optionen wollen wir nachfolgend näher betrachten.

Textexport

In InDesign befindet sich der Text immer in Textrahmen bzw. Textketten (mehrere miteinander verlinkte Rahmen), die miteinander oder mit anderen Objekten gruppiert sein, beliebig auf den Seiten angeordnet sein oder sich auf verschiedenen Ebenen befinden können. Da es in InDesign per se keinen Textexport für alle Inhalte gibt, müssen die einzelnen Rahmen bzw. Textketten jeweils als einzelne Datei manuell exportiert werden. Das lässt sich im Prinzip zwar durch Skripte oder 3rd-Party Plug-Ins etwas automatisieren, aber letztlich muss hier ein erheblicher Aufwand bzgl. Übersetzung getrieben werden, da ja die übersetzten Texte genauso manuell wieder zurückgeführt werden müssen. Je nachdem, ob man nur reinen Text oder formatierten Text (RTF oder ID-Tags) benutzt gehen in der Regel auch noch die Formatierungen verloren und müssen manuell wieder erzeugt werden.

IDML

Adobe hat 2008 mit InDesign CS4 die „InDesign Markup Language“ (IDML) als **Datenaustauschformat** eingeführt. Eine IDML-Datei beschreibt ein komplettes InDesign-Dokument, d.h. den Inhalt, die Geometrien, die Definitionen (Farben, Formate, Typographie, Vorgaben usw.). Durch die Nutzung dieses Austauschformats kann bspw. ein Dokument aus einer höheren Version von InDesign in einer älteren Version geöffnet werden. Dabei gehen aber bewusst die Daten verloren, die von der älteren Version



nicht erkannt werden (z.B. neue Features etc.). Dieses Format ist von Adobe für den Datenaustausch zwischen verschiedenen InDesign-Versionen konzipiert und nicht als Format für die Integration von Drittsystemen.

Einige (bei weitem nicht alle!) TMS-Systeme haben dieses Austauschformat eingebunden – mit eigenen und damit unterschiedlichen Routinen zur Interpretation und Manipulation der Inhalte. Daraus resultiert, dass sich InDesign-Dokumente durchaus unterschiedlich in Abhängigkeit des TMS verhalten und damit übersetzen lassen.

Letztlich handelt es sich um Manipulationen an einer „toten“ Datei, die außerhalb des ursprünglich gedachten Einsatzzwecks verändert wird. Dadurch kommt es immer wieder vor, dass sich die übersetzten IDML-Dateien erst gar nicht öffnen lassen (weil korrupt) oder Seiteneffekte auftreten, weil unbeabsichtigt die Struktur des InDesign-Dokuments verändert wurde (an Stellen die man gar nicht vermuten würde). Es gibt keinerlei Schutzmechanismen für die Integrität des Dokuments. Sehr häufig gehen bei der Übersetzung Formatierungen verloren.

Ein weiterer Nachteil dieses Formats ist es, dass eine IDML-Datei immer ein komplettes InDesign-Dokument beschreibt – völlig egal, ob dieses Dokument z.B. einzelne Artikel beinhaltet, die jeweils inhaltlich zusammengehören und deshalb auch in diesem Kontext übersetzt werden sollten.

Auch die Reihenfolge der Texte einer IDML-Datei folgt nicht den inhaltlichen Strukturen, die der Leser erwartet, sondern ist technisch bedingt. So werden dabei auch Kopf- und Fußzeilen-Elemente (wie z.B. Seitenzahlen) mit den Inhalten gemischt und Strukturen wie Fußnoten, Verweise, Index-Einträge usw. geben dem Übersetzer Rätsel auf.

IDML-Dateien enthalten keine Bilder, sind aber trotzdem relativ groß, was bei Online-TMS dann eine entsprechende Bandbreite zur Übertragung erfordert.

Aus einer IDML-Datei entsteht zwangsläufig nach der Übersetzung ein neues InDesign-Dokument – ohne „Wissen“ des Ursprungdokumentes. D.h. bei Verwendung aus einem Redaktionssystem heraus muss das übersetzte Dokument neu eingesehen und alle Metadaten usw. dazu neu erfasst, Artikel neu erstellt und Bilder neu verlinkt werden.

XLIFF

Das „XML Localization Interchange File Format“ (engl. für „XML-Dateiformat zum Austausch von Lokalisierungsdaten“), abgekürzt XLIFF, ist ein XML-Format zur Darstellung hierarchisch strukturierter Inhalts-Daten in CAT-Werkzeugen.

Das XLIFF Format ist speziell für den verlustarmen Austausch von Übersetzungsdaten und den dazugehörigen Kontextinformationen gedacht. Es ist XML-basiert und damit erweiterbar.

Praktisch jedes TMS „versteh“ dieses Format, weil es der Standard in der Übersetzungsindustrie ist.



Eine XLIFF-Datei enthält neben dem zu übersetzenden Inhalt auch einige Metadaten wie z.B.

- Die Quell- und Zielsprache der Übersetzung
- Die Version für die eindeutige Beschreibung der verfügbaren Optionen
- Das verwendete Tool für die Erstellung / Verarbeitung dieser Datei
- zahlreiche weitere Informationen

Eine XLIFF-Datei enthält den Inhalt immer in sogenannten „Segmenten“ deren Inhalt von der Konfiguration des Tools abhängt. Ein Segment kann z.B. nur ein Wort beinhalten, aber auch einen Satz (oder Teilsatz) oder einen ganzen Absatz – aber niemals ein Return.

Da Segmente grundsätzlich die Basis eines „Translation-Memory“ sind, und damit für die Wiederverwendung von bereits übersetzten Inhalten steht, ist die Segmentierung ein wichtiger Faktor im Datenaustausch von / zum TMS.

Der **ex TranslationFilter** exportiert eine XLIFF-Datei auf Basis einer detaillierten Konfiguration, um die Reproduzierbarkeit zu optimieren und damit die Kosten der Übersetzung zu minimieren. Diese Konfiguration wird verschlüsselt mit in die XLIFF-Datei gegeben und überlebt damit den „Round Trip“ im Übersetzungsprozess. Die Segmentierung kann durch ein umfangreiches Regelwerk optimiert werden. Dabei werden auch sogenannte Inline-Tags verwendet, die z.B. innerhalb eines Segmentes Formatierungen beschreiben.

Damit ist der **ex TranslationFilter** das einzige Werkzeug, welches aus einer „lebenden“ Umgebung (dem laufenden InDesign-Prozess mit allen seinen Schnittstellen für die Plug-Ins) die Inhalte in der gewünschten Form / Struktur exportiert und nach der Übersetzung wieder importiert – und dabei sowohl die Dokumentenstruktur schützt als auch alle Formatierungen beibehält und dadurch eine manuelle Nacharbeit praktisch nicht mehr erforderlich ist.

Dabei werden auch notwendige Funktionen unterstützt wie z.B.

- Änderungen der Schreibrichtung (links-nach-rechts versus rechts-nach-links für Sprachen wie z.B. arabisch, hebräisch usw.) – einschl. der Umstellung des Dokuments
- Automatisches Ersetzen von Schriften in Abhängigkeit der Zielsprache
- automatische Anpassung der Texte um Übersatz zu vermeiden
- Umstellung der Textsprache im Layout, um sofort die richtige Silbentrennung und Rechtschreibung aktiv zu haben
- Filter für den Export / Import von Ebenen
- Erstellung von zusätzlichen sprachabhängigen Ebenen (mehrsprachige Dokumente)
- Das automatische Erkennen und Auslassen von Seitenzahlen etc.
- Unterstützung von Index-Verzeichnissen



- automatische Entfernung von doppelten Leerzeichen, Soft>Returns, Unterschneidungen usw. welche nur aus „Layout-technischen“ Gründen eingefügt worden sind und in der Übersetzung keinen Sinn machen

ist ein Alleinstellungsmerkmal des **ex TranslationFilters**.

Bei Einbettung in Redaktionssysteme wie z.B. WoodWing oder Vjooon durch „Smart Translation Solution“ werden Artikelstrukturen erkannt und unterstützt, d.h. ein Artikel-basiertes Arbeiten ist möglich – ohne, dass (wie bei IDML) die externen Verknüpfungen zu Bildern und Artikeln manuell wiederhergestellt werden müssen. Dabei können u.a. Bilder als sprachabhängig oder sprachneutral behandelt werden.

Vergleichstabelle

	IDML	XLIFF
Inhalte werden segmentiert übergeben	✘	✓
Segmentierungsregeln einstellbar	✘	✓
Unterstützung des SRX-Formats	✘	✓
Dokumentenstruktur / -integrität geschützt	✘	✓
Mit allen TMS kompatibel	✘	✓
Kleine Dateigröße	✘	✓
Artikelbasiertes Arbeiten möglich	✘	✓
Ebenen basiertes Arbeiten möglich	✘	✓
Unterstützung der ISO-Codes für die Sprachen	✘	✓
Übergabe von Metadaten an das TMS	✘	✓
Änderung der Schreibrichtung bei RTL-Sprachen	✘	✓
Automatisches Ersetzen von Schriften in der Zielsprache	✘	✓
Automatische Umstellung der Textsprache im Layout	✘	✓
Unterstützung von Indizes	✘	✓
Automatischer Textfilter für doppelte Leerzeichen etc.	✘	✓
Automatisches Entfernen von Unterschneidungen etc.	✘	✓
Automatische Erstellung eines PDFs für den Übersetzer	✘	✓
Unterstützung von Varianten (WoodWing)	✘	✓