

# XML-Pipeline

In der **XML-Pipeline** von Squeeze werden XML-Dokumente durch speziell angepasste Verarbeitungsstufen geführt, um den unterschiedlichen Anforderungen der Datenextraktion und Dokumentenkategorisierung gerecht zu werden.

Schritt	Aktuelle Pipeline
Einganskanäle	- Alle Eingangskanäle prüfen die Dokumente auf ihre Validität in Bezug auf die EN16931-Spezifikationen.
Initialisierungs-Schritt	- Hochgeladene XML-Dateien werden in ein internes Standard-XML-Format ( <code> intermediate.xml </code> ) überführt. Eine PDF wird aus der <code> intermediate.xml </code> erzeugt. - Je nach Kundenwunsch wird in diesem Schritt ein <a href="#">KoSIT-Prüfbericht</a> erstellt.
Barcode- Extraktions-Schritt	- Viewer-Bilder auf Basis der PDF werden erstellt, gilt auch für ZUGFeRD und XRechnung
OCR-Schritt	- PDF-Dateien mit eingebetteten XML werden normal verarbeitet, um ein OCR-Ergebnis zu erstellen. - Dieser Schritt wird für reine XML-Dateien übersprungen.
Klassifizierungs-Schritt	- Die Klassifizierung von XML-Dokumenten ist derzeit nicht möglich. Eine Fehler wird erzeugt, sobald das Dokument keine Dokumentenklasse nachweist
Extraktions-Schritt	- Übliche Extraktionsmechanismen werden ausgeführt oder übersprungen durch die Stapelklasseneigenschaft: <code> SkipXmlExtraction </code> . - <code> intermediate.xml </code> wird genutzt, um Kreditor und Mandant der XML zu ermitteln. Die eigentliche XML-Extraktion basiert auf der <code> intermediate.xml </code> .

## Extraktion von XMLs

In der **XML-Pipeline** von Squeeze erfolgt die XML-Extraktion als letzter Schritt vor der Autovalidierung/Validierung und hat dabei eine besondere Rolle: Sie überschreibt extrahierte Feldwerte aus den KI-Extraktionen und den Lokatoren-Ergebnissen. Die Extraktion von XML-Daten basiert auf dem Mapping, welches im Administrationsbereich angepasst werden kann. Alternativen

werden gemäß der Definition im Mapping verarbeitet.

# Rendering von XMLs

Das **XML-Rendering** in Squeeze bezieht sich auf den Prozess der Erstellung von PDF-Dokumenten aus XML-Dateien. Dieser Prozess unterscheidet sich je nach Art des XML-Dokuments und umfasst spezifische Anforderungen für standardisierte und benutzerdefinierte XML-Formate.

## Rendering von Spezifischen XML-Dokumenten

Für standardisierte XML-Formate wie XRechnung und ZUGFeRD, die der Norm EN16931 entsprechen, erfolgt das Rendering nach den folgenden Schritten:

1. **Erstellung des Zwischenformats:** Im InitStep wird aus der XML-Datei ein Zwischenformat erstellt, das als Basis für das PDF-Rendering dient. Dieses Zwischenformat wird aus der `intermediate.xml` generiert, die die für die PDF-Erstellung benötigten Daten enthält.
2. **PDF-Erzeugung:** Basierend auf dem Zwischenformat wird ein PDF-Dokument erstellt, das die strukturierten Daten aus der XML übersichtlich darstellt. Dies ermöglicht die Generierung von PDFs, die den Anforderungen der jeweiligen Spezifikation entsprechen und für den Austausch und die Archivierung verwendet werden können.

## Ausblick

In einer zukünftigen Version von Squeeze ist geplant, auch für XML-Dokumente, die nicht den EN16931-Spezifikationen entsprechen, eine vollständige Verarbeitung zu ermöglichen. Diese Weiterentwicklung wird es erlauben, alle Arten von XML-Dokumenten einheitlich und umfassend zu behandeln, wodurch die Integration und Handhabung von benutzerdefinierten XML-Formaten weiter verbessert wird.

---

Revision #28

Created 17 November 2022 13:51:04 by Fabian Terstegen

Updated 25 November 2024 13:31:07 by Vahdettin Balum