

expo.trafo

Datenverarbeitungen in der ExpoDB werden in deren zentralen Modul expo.trafo durchgeführt. Um die vielfältigen Bedarfe von Indexierung und Transformation von Sammlungsdaten in beliebige Formate zu bedienen, ist dieses Modul besonders leistungsfähig und flexibel konfigurierbar.

Das Modul expo.trafo baut auf der bewährten Softwareentwicklung [SWBtrafo](#) auf. SWBtrafo wird bereits seit längerem zur flexiblen Datenverarbeitung genutzt. Ausgangspunkt sind meistens XML-Dateien, z.B. XML-Exporte aus *imdas pro* oder andere XML-Quellen; Ergebnisse sind ebenfalls oft XML-Formate. Dazu werden über sog. Pipeline-Konfigurationen mehrstufige Verarbeitungen definiert, über die die Daten, insbesondere mittels [XSLT](#), prozessiert werden. Komponenten einer Verarbeitung können aber auch z.B. die Einbettung von extern in RTF-Dateien vorgehaltenen Inhalten in die XML-Daten durchführen. Oder der Transfer der Ergebnisse per SFTP oder Http an einen anderen Server.

Die Nutzung von XSLT erlaubt es, die Ausgangsdaten zu selektieren, umzustrukturieren und in beliebigen XML-, JSON- oder CSV-Formaten darzustellen. Bei Bedarf ist (mittels XSL-FO) auch eine Ausgabe in PDF möglich. XSLT ist eine ausgereifte Standardtechnologie, die eine große Verbreitung hat und für die daher breit Kompetenzen verfügbar sind.

Bei der Nutzung von expo.trafo in expo.digest, der Web-Schnittstelle der ExpoDB, wird zur Darstellung von Sammlungsdaten aus *imdas pro* vor allem JSON nachgefragt. Um eine automatisierter Konvertierung nach JSON zu ermöglichen, wird den Daten bereits beim Export aus *imdas pro* im XML eine spezielle Struktur gegeben, die über alle Schritte einer Verarbeitung beibehalten wird. Komplexe Datenelemente werden dazu mit dem type-Attribut "object" versehen. Eine Folge wiederholbarer Elemente wird in ein Element mit dem type-Attribut "array" eingebettet; die wiederholbaren Elemente selbst sind alle mit "_" benannt. Damit werden die JSON-Strukturelemente JSON-Element, JSON-Objekt und JSON-Array vorbereitet.

Expo.trafo bzw. SWBtrafo wird in etlichen Kontexten und eingebettet in etlichen Anwendungen verwendet:

- Aufbereitung von *imdas*-Exporten für die Indexierung in Solr.
- Transformation der Daten in der ExpoDB für die Schnittstelle expo.digest in JSON.
- Bereitstellung von Gesamtexporten als XML aus *imdas pro* via SFTP.
- Aufbereitung von Daten aus *imdas pro* für den Import in Cumulus als CSV-Datei.
- Datenlieferung an die Deutsche Digitale Bibliothek in DDB-Lido.
- Datenlieferung an [LEO-BW](#) über ein ZIP-File, in das auch die Bilder gepackt werden, mit angebundenem SFTP-Transfer an LEO-BW.
- Aufbereitung von Daten für den Transfer ins SAP-System beim Finanzministerium Baden-Württemberg als CSV-Datei.
- Bildimporte nach *imdas pro*
- Verarbeitung von Lieferungen von hochauflösenden Bildern für die Medienbereitstellung
- Aufbereitung von Metadaten aus Literaturverwaltungssystemen für den Import via SWORD-Schnittstelle nach OPUS4.
- Verarbeitung von Ablieferungen von E-Books von Verlagen an die Landesbibliothek im Rahmen der elektronischen Pflichtablieferung.
- u.v.a.