

## Datenimport MARC21 MARCXML

In diesem Kapitel nutzen wir exemplarisch Daten aus dem TIBKAT der Technischen Informationsbibliothek. Diese stehen als sogenannte Open Data zur Verfügung.

### 1 Download

Öffnen Sie die Webseite des Open Data Service der TIB:

<http://www.tib-hannover.de/de/dienstleistungen/opendata/download/4/>

Lesen Sie die Lizenzbedingungen und bestätigen Sie diese:



Wählen Sie „OK“ aus.

Auf der nächsten Seite werden eine Beispieldatei und die Daten selbst zum Download angeboten:



**Die Metadaten aus dem TIB-Katalog zum Herunterladen**  
Stand: 05.03.2015  
Größe der entpackten Archivdateien: ca. 10 GB  
Format der Archivdateien: tgz (tar/gzip)  
Format der Metadaten: [MARCXML](#)  
Umfang der Metadaten: ca. 2,7 Mio. Datensätze

- [tibkat\\_example\\_20150305.tgz](#) (0.57 MB)
- [tibkat\\_full\\_20150305\\_1\\_of\\_4.tgz](#) (195 MB)
- [tibkat\\_full\\_20150305\\_2\\_of\\_4.tgz](#) (166 MB)
- [tibkat\\_full\\_20150305\\_3\\_of\\_4.tgz](#) (160 MB)
- [tibkat\\_full\\_20150305\\_4\\_of\\_4.tgz](#) (136 MB)

Laden Sie die erste Datei << tibkat\_full\_20121030\_1\_of\_4.tgz >> in das Verzeichnis << Downloads >> in Ihrem Homeverzeichnis herunter.

## 2 Entpacken

Entpacken Sie die heruntergeladene Datei in das Verzeichnis << /usr/local/vufind2/local/harvest/ >>. Wählen Sie dazu die Datei im Dateimanager aus und klicken Sie dann im Kontextmenü „Entpacken nach...“ (ohne Abbildung) an. Die Archivdatei enthält ein Verzeichnis namens << tibkat\_full\_20150305\_1\_of\_4 >>. In diesem Verzeichnis befinden sich die Daten in einzelnen MARCXML-Dateien. Jede Datei enthält 2.000 Datensätze.

Da wir uns im Rahmen des Tutorials mit dem Import vertraut machen wollen, werden wir nur insgesamt 50.000 Datensätze importieren.

Erstellen Sie im Verzeichnis << /usr/local/vufind2/local/harvest/ >> ein neues Verzeichnis namens << TIBKAT >>. Kopieren Sie die ersten 25 XML-Dateien aus dem Verzeichnis << tibkat\_full\_20150305\_1\_of\_4 >> in den Verzeichnis << TIBKAT >>.

## 3 Import vorbereiten

Öffnen Sie mit Mousepad die Datei << marc\_local.properties >> im Verzeichnis << /usr/local/vufind2/import/ >>. Ändern Sie die beim Testimport editierten Werte „collection“ und „institution“ wie folgt:

```
collection = "TIBKAT"  
institution = "TIB"
```

## 4 Import durchführen

Starten Sie VuFind.

Führen Sie im Terminal die folgenden Befehle aus:

```
cd /usr/local/vufind2/harvest/  
./batch-import-marc.sh TIBKAT
```

Der zweite Befehl startet den Import. Dieser verarbeitet nacheinander die einzelnen XML-Dateien und meldet den Import jedes Datensatzes im Terminal:

```
INFO [main] (MarcImporter.java:317) - Added record 1999 read from file: 312383037
INFO [main] (MarcImporter.java:317) - Added record 2000 read from file: 312390998
INFO [main] (MarcImporter.java:615) - Adding 2000 of 2000 documents to index
INFO [main] (MarcImporter.java:616) - Deleting 0 documents from index
INFO [main] (MarcImporter.java:490) - Calling commit (with optimize set to false)
INFO [main] (MarcImporter.java:502) - Done with the commit, closing Solr
INFO [main] (MarcImporter.java:505) - Setting Solr closed flag
INFO [main] (MarcImporter.java:625) - Finished indexing in 0:40,00
INFO [main] (MarcImporter.java:634) - Indexed 2000 at a rate of about 49.0 per sec
INFO [main] (MarcImporter.java:635) - Deleted 0 records
INFO [Thread-2] (MarcImporter.java:565) - Starting Shutdown hook
INFO [Thread-2] (MarcImporter.java:584) - Finished Shutdown hook
stefan@stefan-VirtualBox:/usr/local/vufind2/harvest$
```

Während des Importes werden die importierten Dateien in ein Verzeichnis namens << processed >> verschoben. Wird der Import nach einer Unterbrechung fortgesetzt, werden do die Dateien im Ordner << processed >> nicht nochmals importiert.

## 5 VuFind neu starten und Index optimieren

Starten Sie VuFind neu.

Optimieren Sie den Index.

## 6 Import kontrollieren

Führen Sie eine leere Suche in VuFind durch:

The screenshot shows the VuFind search interface. At the top left is the 'ma lis' logo. In the top right, it says '0 in der Auswahl' and 'Login Sprache'. Below the logo is a search bar with a dropdown menu set to 'Alle Felder' and a green 'Suchen' button. To the right of the search bar is a link 'Erweitert'. Below the search bar, it says 'Suche:'. Underneath, there are several facets for 'Ähnliche Schlagwörter... innerhalb Ihrer Suche.' including 'Congresses 1.259', 'Hochschulschrift 1.055', 'Zeitschrift 332', 'Kongress 844', and 'Maschinenbau: Allgemeines 347'. To the right of these facets is a 'Suche einschränken' section with two expandable panels: 'Bestand' and 'Format'. The 'Bestand' panel shows 'TIBKAT' with 50.000 items and 'DNB Reihe A' with 2.519 items. The 'Format' panel shows 'Buch' with 48.009 items, 'Mikrofilm' with 2.469 items, and 'Buchkapitel' with 1.227 items. Below the facets, it says 'Treffer 1 - 20 von 52.519 für Suche: "", Suchdauer: 4,81s' and 'Sortieren Relevanz'. There are buttons for 'E-Mail', 'Export', 'Drucken', 'Speichern', and 'In die Merkliste'. The first search result is 'Zur Hydroformylierung von Epoxiden' by Weber, Robert, published in 1998. It has a 'Volltext' link and a 'Buch' button. The title is 'ZUR HYDROFORMYLIERUNG VON EPOXIDEN' and the author is 'Weber, Robert'.

Anhand der Facette „Bestand“ können Sie sehen, dass zu den 2519 Datensätzen aus der DNB-Reihe A nun 50.000 Datensätze aus dem TIBKAT hinzugekommen sind.

## Nächstes Kapitel

Konfiguration der Suche, Teil II

## Quellen

MARC Records. VuFind Documentation. [https://vufind.org/wiki/importing\\_records](https://vufind.org/wiki/importing_records)