

# Verändern oder Ergänzen des OAI-Harvesting

## Version: APS 2.8

Diese Angaben beziehen sich auf das Harvesting der Hauptlizenz (Erfassungsdatenbank) durch die online-Version von APS und durch dazu gehörende Unterlizenzen.

## Die Konfigurationsdatei *eje-oai.xml*

Die Rahmenbedingungen des Harvesting werden in der Datei *eje-oai.xml* definiert. Es gibt sie nur einmal, auch wenn es mehrere Datenbanken gibt, die harvesten. Sie ist zu finden im Verzeichnis *dbconf* der Erfassungsdatenbank (z. B. für Köln: *..KEK\_PRIVATE\_PRODUCTION\dbconf*). In dieser Datei befinden sich tags mit folgenden Namen:

```
<OAI>
<Repository>
<StateExtractor>
<Set>
<Include>
```

Der tag *StateExtractor* definiert die Dokumenttypen, die in den einzelnen *Sets* Verwendung finden sollen.

Hier können Bedingungen definiert werden, unter welchen die Dokumente den Status '*published*', '*notpublished*' oder '*draft*' erhalten (s. Beispiel 1). Ferner kann man einem Dokumenttyp auch grundsätzlich einen Zustand zuweisen (s. Beispiel 2).

Beispiel 1:

```
if (/obj/Status/text()[1]='draft') then 'draft'
else if (/obj/Status/text()[1]='published') then 'published'
else if (/obj/Status/text()[1]='check') then 'draft'
else if (/obj/Status/text()[1]='unpublished') then 'notpublished'
else 'draft'
```

Beispiel 2:

```
<StateExtractor doctype="kue">'published'</StateExtractor>
```

Über die weiteren tags wird noch im Einzelnen zu berichten sein.

Die URL der Quelldatenbank wird in den einzelnen manuell oder zeitgesteuert eingestellten [Jobs](#) in der jeweiligen Zieldatenbank definiert, z. B.:

```
<?xml version="1.0" encoding="UTF-8"?>
<harvest_job>
<baseURL>http://kek_private_production.verwaltung.stadtkoeln.de/db/apsisa.dl
l/oai-pmh</baseURL>
  <metadataPrefix>raw</metadataPrefix>
```

```
<set>wallraf</set>
<template>oai_xsls\raw</template>
<lookBehind>00:00:05</lookBehind>
</harvest_job>
```

## Aktivieren der neuen OAI

Nachdem die eje-oai.xml angepasst und in das vorgesehene Verzeichnis kopiert wurde, empfiehlt es sich zunächst das Harvesten in den betroffenen Datenbanken zu deaktivieren, damit es möglich ist eine „saubere“ Zieldatenbank aufzubauen.

Nun muss die Pipeline von APS neu gestartet werden. Dies geschieht, indem der Dienst mit folgenden Befehlen angehalten und neu gestartet wird. Der Befehl hierzu lautet:

*Net Stop ere*

*Net Start ere*

Um den neuen Harvesting-Vorgaben Geltung zu verschaffen, muss in der Quelldatenbank der Vorgang des **Morphens** gestartet werden. Hierdurch wird das Harvesting noch nicht durchgeführt (dieses erfolgt ja durch die Zieldatenbanken selbst), aber es werden im sogenannten **Repository** die Dokumente zum Harvesten bereitgestellt. Der entsprechende Job hat folgende Parameter:

```
<?xml version="1.0" encoding="UTF-8"?>
<docSelector_job>
  <DocSelection docType="obj"/>
  <DocSelection docType="kue"/>
  <DocSelection docType="lit"/>
  <DocSelection docType="soz"/>
  <DocSelection docType="per"/>
  <DocSelection docType="que"/>
  <DocSelection docType="ere"/>
  <DocSelection docType="wap"/>
  <DocSelection docType="aus"/>
  <DocSelection docType="ver"/>
</docSelector_job>
```

## Starten des Harvesting

Nach diesen Vorbereitungen kann man das Harvesting in den jeweiligen Zieldatenbanken (durch Einschalten der Zeitsteuerung oder durch direktes Starten des Jobs) starten. Es macht dabei ggf. Sinn die Dokumente der Datenbank durch entsprechende [Dokumentenverarbeitungen](#) löschen zu lassen, um einen sauberen Schnitt zu bekommen.



From:

<http://www.apswiki.de/> - **"Selbsthilfe"-APS-Handbuch**

Permanent link:

<http://www.apswiki.de/doku.php?id=aps:harvesting>

Last update: **2019/01/31 11:56**

