

Aufbereiten der hierarchischen Webpräsentation der SWA-Archivbestände

Technische Dokumentation / Andres von Arx

Inhalt

Sachdokumentation	2
Präsentation im Web.....	2
Generieren der Webseiten.....	2
Datenfluss.....	3
Firmen- und Verbanddokumentation	7
Präsentation im Web.....	7
Generieren der Webseiten.....	7
Datenfluss.....	8
Privatnachlässe.....	10
Präsentation im Web.....	10
Generieren der Webseiten.....	10
Systematische Gliederung der E-Zitungsausschnitte (E-ZAS).....	11
Präsentation im Web.....	11
Generieren der Webseiten.....	11
Datenfluss.....	12

Sachdokumentation

Die Deskriptoren der Sachdokumentation basieren auf einer lokalen, noch unverlinkten Version des „Standard-Thesaurus Wirtschaft“. Die Webseiten sind ziemlich stabil. Ein Wechsel auf den STW (linked Version) ist angedacht. Der aktuelle Thesaurus ist als Aleph Autoritätsdatei implementiert. Vom Einrichten einer neuen Aleph Autoritätsdatei für den STW wird abgeraten, da kaum nachhaltig.

Präsentation im Web

<http://www.ub.unibas.ch/ub-wirtschaft-swa/schweiz-wirtschaftsarchiv/sachdokumentation/>

Generieren der Webseiten

Anleitung:

http://ub-files.ub.unibas.ch/ordnungssystem/1_Eigene-Organisation/15_IT/153_Bibliothekssysteme/bibtools/hierarchie_toolbox/swa_sachdokumentation/ANLE_SWA_Sachdokumentation_aufbereiten_20140818_ava.html

Wie: Aleph GUI
Linux-Client mit Zugriff auf ub-files und SSH public key auf ub-webqm/ub-webvm

Wann: bei Bedarf, auf Bitte vom SWA

Laufzeit: Aufbereitung der Aleph-Daten: 5-10 Minuten
Konversionsskript: 3 Minuten

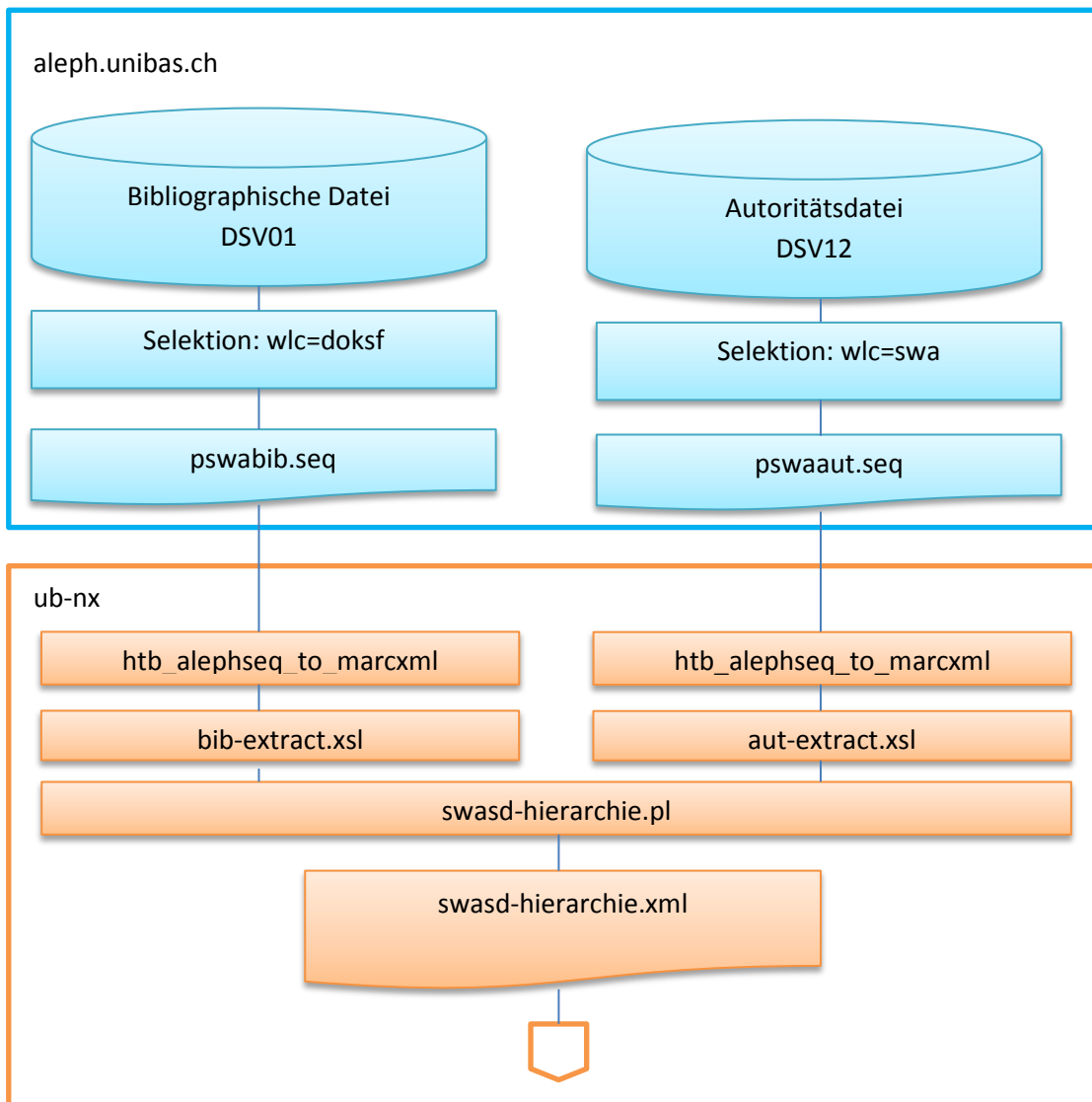
```
$ cd ~/intranet/ordnungssystem/1_Eigene-Organisation/15_IT/153_Bibliothekssysteme
$ cd bibtools/hierarchie_toolbox/swa_sachdokumentation
$ ./make-swa-sachdok.sh
```

Generiere Webseiten für die SWA Sachdokumentation

```
-----
* Dateien von Aleph herunterladen
  Anzahl Aleph Records:
  BIB: 1267
  AUT: 963
* flicke Zeichensatz
* konvertiere nach MARC21 XML
* extrahiere relevante AUT + BIB Daten
* kombiniere AUT + BIB zu Strukturdatei
* extrahiere Hierarchie als Textdatei
* extrahiere Hierarchie (Firmen+Verbände)
* dokumentiere Thesaurus
* generiere Listen für E-Zeitschriftenarchiv
* generiere Tree Menu
* generiere Detailseiten
* generiere Detailseiten (Firmen+Verbände, ohne Geographica)
* generiere A-Z-Liste
* generiere Wortindex für Suchmaschine
* generiere A-Z-Liste für Suchmaschine (Firmen+Verbände)
* generiere Wortindex für Suchmaschine (Firmen+Verbände)
* Outputdateien zippen und hochladen
fertig: 24. Juni 2016, 12: 25: 16
```

Datenfluss

(vereinfacht, Zwischendateien sind ausgelassen)



Die Autoritätsdatei enthält Aufnahmen für jeden Deskriptor, mit impliziter Hierarchie (übergeordnete und verwandte Begriffe, Synonyme). Die Bibliographische Datei enthält Aufnahmen für die einzelnen Dossiers mit den entsprechenden Deskriptoren; diese können ev. noch geographischen Unterteilungen enthalten.

Die Programme generieren aus diesen Daten eine hierarchisch gegliederte Datei **swasd-hierarchie.xml**, die alle Informationen zusammenfasst.

Diese Datei ist Ausgangspunkt der folgenden Produkte:



swasd-descriptors.xsl

../htb/generate_treemenu_pro.pl

output/tree_items.js

Sachdokumentation:
Steuerdatei von Tigra Tree Menu.
Javascript für Hierarchiebaum.



swasd-detailseiten.xsl

swasd-splitte-detailseiten.pl

output/details/did*.html

Sachdokumentation: rund 1300
Seiten für einzelne Deskriptoren, mit
Synonymen und verwandten
Begriffen, Vorkommen des
Deskriptors in anderem Kontext und
Deep-Linking zu Aleph.



swasd-indexterms.xsl

swasd-index-az.xsl

swasd-index-az.pl

output/index*.html

Sachdokumentation:
Indexseiten A-Z



swasd-indexterms.xsl

swasd-suchmaschine.xsl

swasd-suchmaschine.pl

tmp/suchmaschine.txt

Sachdokumentation:
Wortliste für eine einfache
sequentielle Volltextsuche mit
/cgi-bin/index/swasd-index.pl

Der Teilbaum „Wirtschaftssektoren“ der Sachdokumentation ist gleichzeitig die Systematische Gliederung der Firmen- und Verbanddokumentation (s.u.). Die dafür benötigten Dateien werden ebenfalls hier aufgebaut und für die Weiterbearbeitung in den Skripten für die F+V Doku bereitgestellt¹.



swafv-descriptors.xsl

swafv/swafv-descriptors.txt

Firmen- + Verbanddokumentation:
Grundgerüst des Hierarchiebaums.



swafv-detailseiten.xsl

swafv/detailseiten.html

Firmen- + Verbanddokumentation:
Texte zu den einzelnen Deskriptoren,
noch ohne konkrete Körperschaften,
ungesplittet.



swafv-indexterms.xsl

swasd-index-az.xsl

swafv/indexterms.txt

Firmen- + Verbanddokumentation:
Rohdaten für die A-Z Listen



swafv-indexterms.xsl

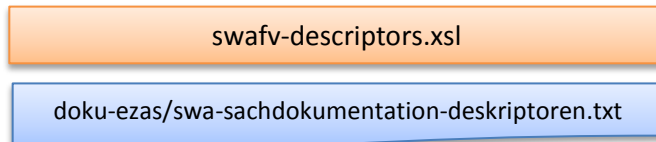
swasd-suchmaschine.xsl

swafv/suchmaschine_deskriptoren.tmp

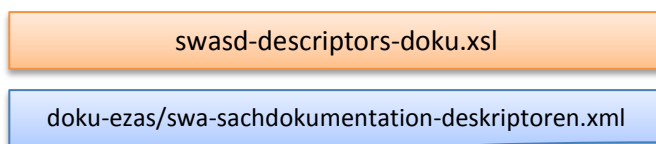
Firmen- + Verbanddokumentation:
Rohdaten für die Volltextsuche

¹ **Achtung:** diese beiden GEVER Verzeichnis sind ein und dasselbe (Unix hard link):
.../153_Bibliothekssysteme/bibtools/hierarchie_toolbox/swa_sachdokumentation/**swafv**
.../153_Bibliothekssysteme/bibtools/hierarchie_toolbox/swa_firmen_und_verbaende/**from_swasd**

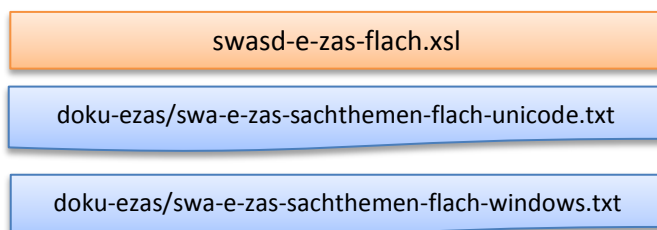
Die Deskriptoren der Sachdokumentation bilden auch einen Bestandteil des Thesaurus des Elektronischen Zeitschriften-Ausschnitt-Dienstes (E-ZAS, s.u.). Die dazu benötigten Dateien werden ebenfalls hier generiert.



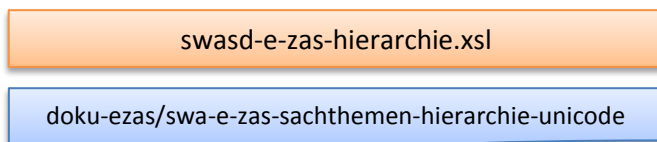
E-ZAS:
Baum der Sach-Deskriptoren (Text)



E-ZAS:
Baum der Sach-Deskriptoren (XML)



E-ZAS:
Baum der Sach-Deskriptoren (Text)
mit Geographica



E-ZAS:
Deskriptoren Sachdokumentation
mit Geographica, 5-stufige Ketten
für den Import in NewBase/E-ZAS

Firmen- und Verbanddokumentation

Präsentation im Web

<http://www.ub.unibas.ch/ub-wirtschaft-swa/schweiz-wirtschaftsarchiv/firmen-und-verbandsdokumentation/>

Generieren der Webseiten

Anleitung:

http://ub-files.ub.unibas.ch/ordnungssystem/1_Eigene-Organisation/15_IT/153_Bibliothekssysteme/bibtools/hierarchie_toolbox/swa_firmen_und_verbaende/ANLE_SWA_Firmen_und_Verbaende_aufbereiten_20140604_ava.html

Wie: Linux-Client mit Zugriff auf ub-files und SSH public key auf ub-webqm/ub-webvm

Wann: Jeden Freitag morgen

Laufzeit: 5-10 Minuten

```
$ cd ~/intranet/ordnungssystem/1_Eigene-Organisation/15_IT/153_Bibliothekssysteme
$ cd bibtools/hierarchie_toolbox/swa_firmen_und_verbaende
$ ./make-firmen-und-verbaende.sh
```

SWA Firmen und Verbände: Webseiten aufbauen

Fre Jun 24 15:00:57 CEST 2016

```
* Aleph Daten downloaden
  Daten vom 23 Jun 2016 22:05:04
* Qualitäts-Check
* Konkordanz Deskriptoren / Körperschaften
* Extrahiere Körperschaften
* Geographische Umschlagwörter extrahieren
* Branchen-Systematik mit Geographica anreichern
  [Anmerkung: nur Geographica, die in den doksf-Daten vorkommen.]
* Tree Menu generieren
* Detailseiten aufbereiten
  (bitte 5 Minuten Geduld...)
* generiere Wortindex (Deskriptoren)
* generiere Wortindex (Körperschaften)
* generiere A-Z Liste
* Outputdateien zippen und hochladen
* installiere Webseiten auf Webserver (TEST)
* installiere Wortindex (Deskriptoren) auf Webserver (TEST)
* installiere Wortindex (Körperschaften) auf Webserver (TEST)
* installiere Webseiten auf Webserver (PROD)
* installiere Wortindex (Deskriptoren) auf Webserver (PROD)
* installiere Wortindex (Körperschaften) auf Webserver (PROD)
* temporäre Dateien löschen
```

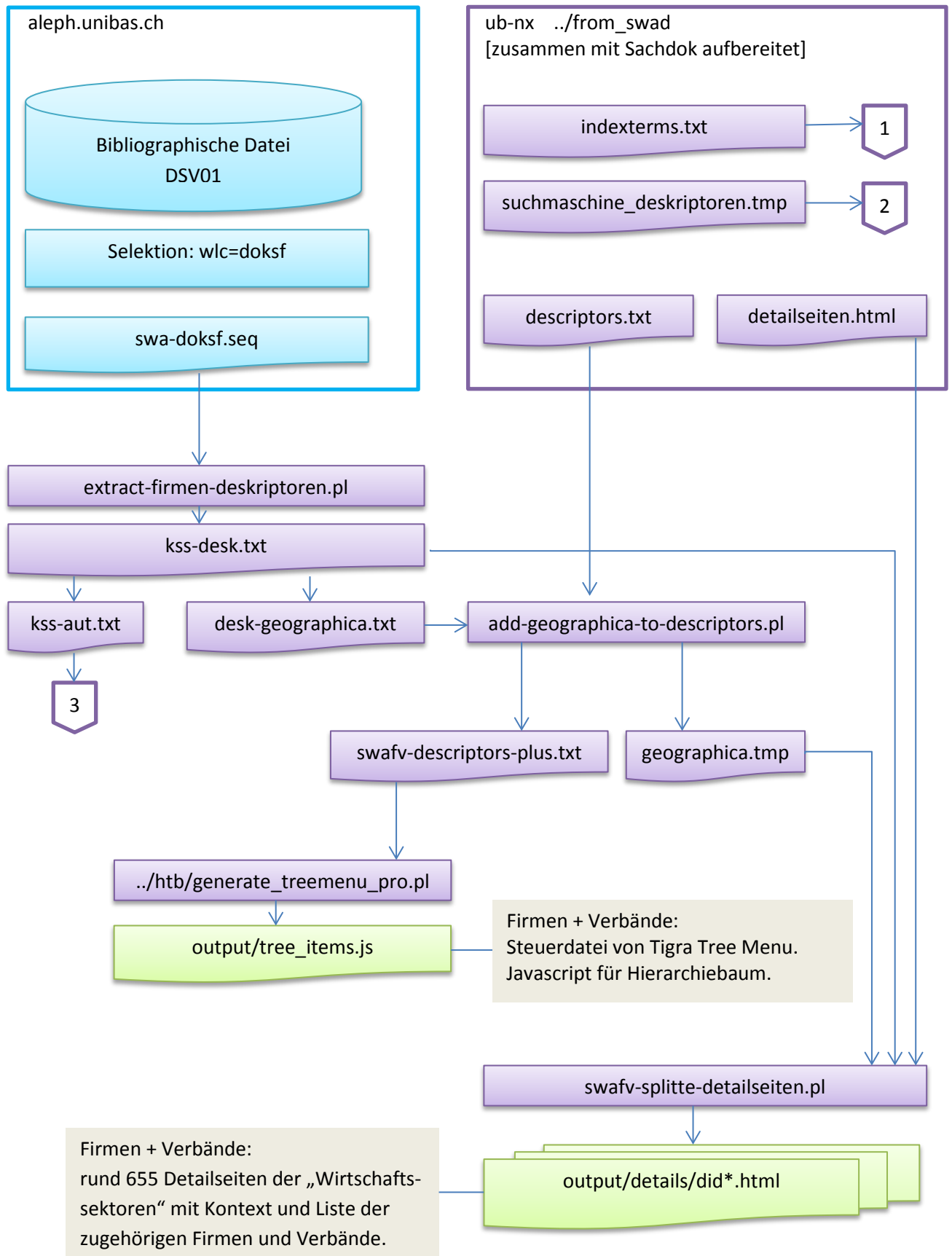
Alle Körperschaften sind GND-kontrolliert.

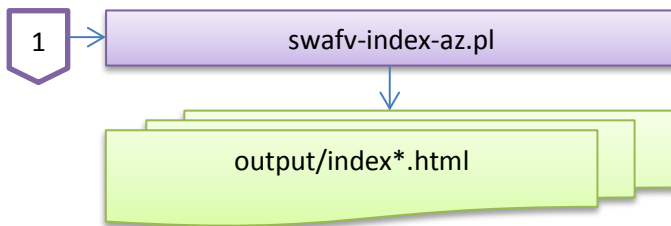
Fre Jun 24 15:02:08 CEST 2016

fertig.

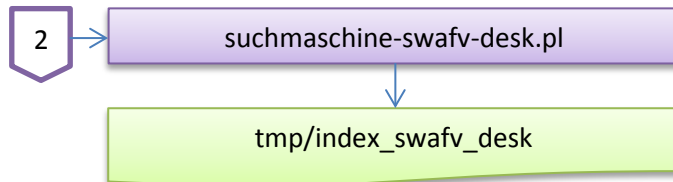
Datenfluss

(vereinfacht, Zwischendateien sind ausgelassen)

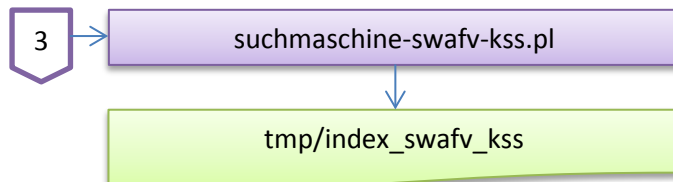




Firmen + Verbände:
Indexseiten (Deskriptoren) A-Z



Firmen + Verbände: Wortliste für
Volltextsuche (Deskriptoren)



Firmen + Verbände: Wortliste für
Volltextsuche (Körperschaften)

Privatnachlässe

Präsentation im Web

<http://www.ub.unibas.ch/ub-wirtschaft-swa/schweiz-wirtschaftsarchiv/privatarchive/>

Generieren der Webseiten

Anleitung:

http://ub-files.ub.unibas.ch/ordnungssystem/1_Eigene-Organisation/15_IT/153_Bibliothekssysteme/bibtools/hierarchie_toolbox/swa_privatarchive/ANL_SWA_Privatarchive_Daten_generieren_20130617_ava.html

Wie: Linux-Client mit Zugriff auf ub-files und SSH public key auf ub-webqm/ub-webvm

Wann: bei Bedarf, manuell

Laufzeit: 5 Minuten (Download von DSV05)
1 Minute (Laufzeit Skript)

```
$ cd ~/intranet/ordnungssystem/1_Eigene-Organisation/15_IT/153_Bibliothekssysteme
$ cd bibtools/hierarchie_toolbox/swa_privatarchive
$ ./make-swap.sh
```

SWA Privatarchive

```
-----
* extrahiere Daten aus Aleph Sequential (dsv05)
* konvertiere in MARC XML
* generiere Schlagwort-Struktur
* generiere Hierarchie
* generiere Tree Menu
* fixe Tree Menu
* baue Index für Suchmaschine
* installiere Treemenu auf Webserver
* installiere Wortindex auf Webserver
* cleanup tmp files
```

Systematische Gliederung der E-Zitungsausschnitte (E-ZAS)

Präsentation im Web

<http://ezas.wirtschaftsarchiv.ch/>

Generieren der Webseiten

Anleitung:

http://ub-files.ub.unibas.ch/ordnungssystem/1_Eigene-Organisation/15_IT/153_Bibliothekssysteme/E-ZAS/thesaurus/ANLE_ezas_thesaurus_aufbereiten_20140501_ava.html

```
$ cd ~/intranet/ordnungssystem/1_Eigene-Organisation/15_IT/153_Bibliothekssysteme
$ cd E-ZAS/thesaurus/bin
$ ./make-ezas-thesaurus.sh
```

Generiere E-ZAS Thesaurus

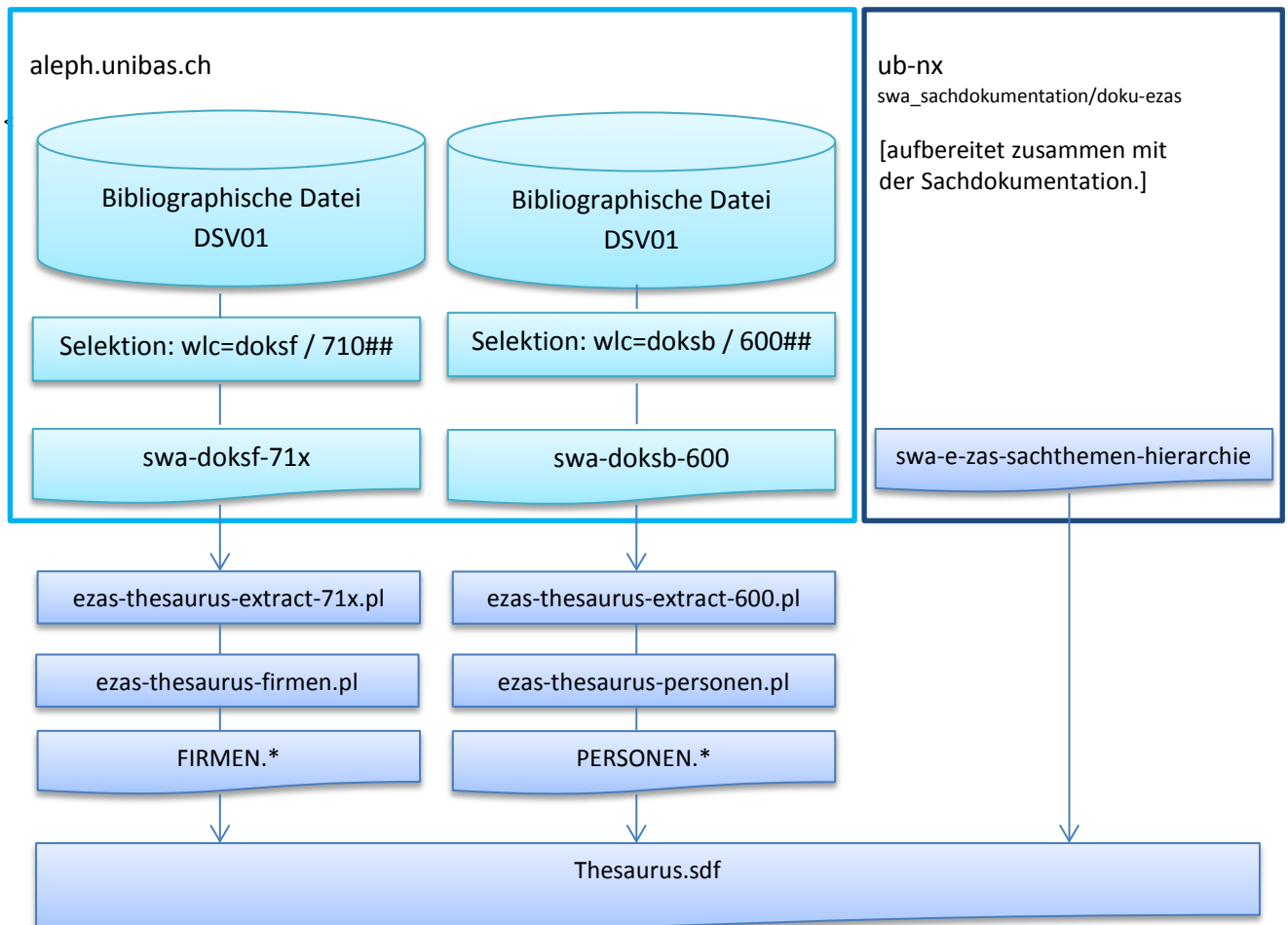
```
* download Aleph-Dateien
* prüfe Pfade und Dateien
* Sachthemen
* Firmen und Verbände
* Personen
* kombiniere Teile
  Thesaurus.sdf
  Thesaurus_kurz.sdf
* dokumentiere Unterschiede
  DOKU_Thesaurus_Aenderungen.html
  DOKU_Thesaurus_Aenderungen_komplett.html
* cleanup
done.
```

Datenfluss

(vereinfacht, ohne Zwischenschritte)

Der Systematische Gliederung in E-ZAS speist sich aus verschiedenen Quellen:

- Firmen und Verbände: Körperschaftsansetzungen aus Aleph Dossieraufnahmen („doks f“)
- Personen: Lokal- und GND-Personenschlagwörter aus Aleph Dossieraufnahmen („doks b“)
- Sachthemen: Deskriptoren aus der Sachdokumentation (s.o.)



Die Datei Thesaurus.sdf bildet die Systematische Gliederung innerhalb von E-ZAS.
Sie wird vom SWA direkt in E-ZAS eingespielt.

Die Programme generieren darüber hinaus eine Dokumentation der Änderungen in Thesaurus.sdf.
Mit Hilfe dieser Dokumentation können allenfalls die Deskriptoren in E-ZAS manuell geändert werden.

ACHTUNG: In der Nachfolge der Einführung von GND sind grössere Bereinigungsarbeiten an den Deskriptoren nötig. Das SWA (gfr) ist in Kontakt mit der Herstellerfirma Newbase, um sich Tools zur Batch-Mutation von Deskriptoren offerieren zu lassen. Eine Konkordanzliste alt/neu existiert. (Stand Juni 2016/ava)