

(Geo)Kettle – Ein unvollständiger, subjektiver Erlebnisbericht

Stefan Ziegler

Amt für Geoinformation
Rötistrasse 4
4500 Solothurn

xx. Januar 2014

Inhalt

- Was ist Kettle? Was ist PDI? Was ist GeoKettle?
- Geo ohne GeoKettle?
- Demo
- Inspiration

Kettle/PDI

- ETL-Tool (Extract, Transform, Load)
- Kettle = PDI (Pentaho Data Integration)
- Dual-licensing

Funktionsweise

- Hübsche Oberfläche
- Relativ intuitiv...
- Prozesse gespeichert in einer XML-Datei (in DB-Repo oder filebasiert)
- Kettle *interpretiert* XML-Datei
- Prozess steuerbar mit Parameter und Variablen.
- Prozesse werden gestartet:
 - aus der Oberfläche *Spoon*
 - von der Kommandozeile *Pan* resp. *Kitchen*
 - via Webserver *Carte* (auf entferntem Rechner)
 - oder ganz abgefahren in Cluster.

Weitere ETL-Tools

- Talend Open Studio (TOS):
 - komplizierter?
 - Codegenerator: Erzeugt Java-Code.
- FME:
 - Kenne ich nicht.
 - Arbeitsplatzwechsel resp. Arbeitsmittelwechsel notwendig ☹
 - Interlis ☺

GeoKettle

- Kettle um Geoprozesse erweitert.
- «Geo» *nicht* als Plugin integriert.
- GeoKettle basiert auf Kettle 3 (aktuell Kettle 5).
- aber...

Geo ohne GeoKettle

- Sämtliche Geoprozesse werden in der Datenbank gemacht.
- Funktioniert tadellos aber einschränkend, da z. B. kein Import/Export von Shape-Dateien etc. möglich ist.
- Beispiele:
 - Bodenbedeckungsarten pro Liegenschaft (im öffentlichen Eigentum)
 - AV-Fileverifikation

Inspiration

- `samples/-Verzeichnis`
- `http://www.slideshare.com`
- `http://forum.pentaho.com`
- `http://forum.spatialytics.com`
- `http://diethardsteiner.blogspot.ch`

Persönliches Fazit

- Keine eierlegende Wollmilchsau.
- Oftmals reicht wahrscheinlich auch ein Skript.
- GeoKettle, quo vadis?
- Flexibel: läuft z. B. auch als Cronjob.
- Übersichtliches GUI: einzelne Prozessschritte sind klar sichtbar.
- Tiefe Hemmschwelle: man programmiert nicht.
- Grosse Auswahl an Werkzeugen: E-Mail-Versand, Logging, Access/Excel In-/Output, SAP-Input (?) etc.