

OpenDataHub HSR – Teilen, Beziehen und Konvertieren von Daten

Bachelorarbeit Frühlingssemester 2015, Studiengang Informatik

Aufgabenstellung

Der Austausch von strukturierten Daten - d.h. der Datenaustausch – zwischen unabhängigen Organisationen ist nicht einfach. In einer idealen Welt einigt man sich auf einen gemeinsamen Datenaustausch-Mechanismus, beschreibt Datenmodelle für verschiedene Anwendungsbereiche und jedes beteiligte System ist dafür besorgt, dass die Daten gemäss diesem Modell in einem einheitlichen Format importiert, bearbeitet und wieder exportiert werden können.

Die Realität scheint sich immer weiter von dieser Vision zu entfernen. Die marktbeherrschenden Softwarehersteller überlassen das Problem den Nutzern. Die Datenlieferanten und Systembetreiber sind überfordert und einigen sich auf spezifische oder proprietäre Lösungen - oder sie begnügen sich mit 1:1-Konvertern (vgl. OGR und geoconverter.hsr.ch etc.) oder analogen Lösungen (Papier).

Ein Lösungsansatz könnte ein zentrales Portal (= "OpenDataHub") sein, über dieses sich Daten offen teilen, beziehen und konvertieren lassen wobei sich die Beteiligten nicht um die Formatkonversion und Schematransformation kümmern müssen: Die Datenlieferanten liefern ihre Daten in ihrem Format und Schema und die Nutzer beziehen es ihrerseits, wie sie wollen - vorausgesetzt, die entsprechende Formatkonversion und Schematransformation existiert. Darum kümmern sich Spezialisten.

Typische Rollen dieser Architektur sind:

- Datenlieferanten
- Nutzer
- Spezialisten für Schematransformationen

Typische Komponenten dieser Architektur sind:

- Lese- und Schreib-Komponenten
- Komponenten für typische Konversionsaufgaben
- Fertige "Transformationen" (vgl. Use Cases)

Use Cases

1. Postadressen (CSV)
2. Verkehrshindernisse-Datenbank
3. Gebäudeadressen (MOPublic)

Deliverables

- Überblick und Evaluation ähnlicher Projekte, insbesondere: „Project dat“, CKAN (geopol.ch).
- Erwähnte Use Cases inkl. Beschreibung, (Demo-)Daten und Schema-Transformationen etc.
- Überlegungen und Evaluation zum Schema-Mapping: Sprache und Tools (Pentaho Kettle, ogrtools, FME, SQL, Eigene?)
- Die Implementation, die u.a. umfasst:
 - Listen-Anzeige bzw. einfaches Suchen/Filter von Daten
 - Publizieren von Daten
 - Storage (File oder Database, relational/NoSQL)
 - Beziehen von Daten

- Lesen/Schreiben ausgewählter Formate, insbesondere CSV, Shapefile, GeoPackage, GML, Interlis etc. (z.B. mittels OGR bzw. ogrtools)
 - Schema-Mapping-Software
- Webseite mit (Cloud/Web-)Services mit Demonstration, mind. enthaltend die erwähnten Use Cases
- Angebot von vordefinierten (parametrisierbaren) Schema-Transformationen.

Vorgaben/Rahmenbedingungen

- Die Software soll Open Source und auf gängigen Browsern und Linux-Servern lauffähig sein.
- Die SW-Engineering-Methode und Meilensteine werden mit dem Betreuer vereinbart.
- Sourcecode und Software-Dokumentation sind Englisch (inkl. Installation, keine Benutzerdokumentation, höchstens eine Online-Kurzhilfe).
- Das Software-Frontend ist Deutsch (optional auch Englisch).
- Die Projekt-Dokumentation und -Präsentation sind auf Deutsch.
- Die Nutzungsrechte an der Arbeit bleiben bei den Autoren und gehen auch an die HSR/Betreuer über. Die Softwarelizenz ist „MIT“.
- (Ein Video - gemäss den Vorgaben des Studiengangs – ist optional).
- Ansonsten gelten die Rahmenbedingungen, Vorgaben und Termine des Studiengangs Informatik bzw. der HSR.

Inhalt der Dokumentation

- Die Projektdokumentation (Prosa) und die Benutzerschnittstelle sind in Deutsch. Der Code, die Kommentare und die Versionsverwaltung sind in Englisch.
- Die fertige Arbeit muss folgende Inhalte haben:
 1. Abstract, Management Summary, Aufgabenstellung
 2. Technischer Bericht
 3. Projektdokumentation
 4. Anhänge (Literaturverzeichnis, Glossar, CD-Inhalt)
- Die Abgabe ist so zu gliedern, dass die obigen Inhalte klar erkenntlich und auffindbar sind.
- Zitate sind zu kennzeichnen, die Quelle ist anzugeben.
- Verwendete Dokumente und Literatur sind in einem Literaturverzeichnis aufzuführen.
- Dokumentation des Projektverlaufes, Planung etc.
- Weitere Dokumente (z.B. Kurzbeschreibung, Poster) gemäss www.hsr.ch und Absprache mit dem Betreuer.

Form der Dokumentation

- Bericht (Struktur gemäss Beschreibung) gebunden (2 Exemplare) und in Ordner (1 Exemplar „kopierfähig“ in losen, gelochten Blättern).
- Alle Dokumente und Quellen der erstellten Software auf sauber angeschriebenen CD (3 Ex.).

Bewertungsschema

Es gelten die üblichen Regelungen zum Ablauf und zur Bewertung der Arbeit (6 Aspekte) des Studiengangs Informatik der HSR jedoch mit besonderem Gewicht auf moderne Softwareentwicklung (Tests, Continuous Integration, einfach installierbar, funktionsfähig).

Beteiligte

Diplomanden

Fabio Scala

Christoph Hüsler

Remo Liebi

Industriepartner

- (Open Community / Open Source).

Betreuung HSR

Verantwortlicher Dozent: Prof. Stefan Keller (sfkeller@hsr.ch), Geometa Lab am IFS der HSR