

Titel: Automatisierte, strukturierte Datenextraktion aus geologischen Informationen (Raster PDF) als Text, Zahlen, Linien und Muster



VE/Amt: Bundesamt für Landestopografie swisstopo



Kontaktperson:



Geschäftsprozess & -kontext:



Beschreibung der Challenge:



Die Digitalisierung schreitet stetig voran, jedoch bestehen viele Information nur als Raster PDF ohne strukturieren Inhalt.

Über die letzten Jahre und Jahrzehnte haben sich sehr viele PDF angesammelt, deren Inhalt zeitaufwändig, durch manuelle Erfassung, in strukturiere Datensätze umgewandelt werden. Desweitern übernimmt die swisstopo durch die Digitalisierung der Kantonsarchive ca. 100'000 digitalisierte Bohrprofile und ca. 1'000 digitalisierte geologische Karten. Diese Information strukturiert in unsere Datenbanken zu erfassen, und mit den richtigen Metadaten zu versehen, braucht ca. 300 Mannjahre und ist somit, weder finanzierbar, noch schaffbar.

Würden diese Information jedoch als digitale, strukturierte Daten vorliegen, würde ein massiver Mehrwert generiert, welcher nicht nur der swisstopo, sondern auch der ganzen Schweizer Geologie Szene für das Verständnis des Untergrund hilft.

Aus digitalisierten, als Raster PDF vorliegenden Dokumenten (geologische Schichtprofile, Bohrungskorrelationen und geologische Karten) sollen mit Machine Learning Algorithmen der Informationsgehalt wie Text (Lithologie, Lithostratigraphie, Chronostratigraphie, geologische Einheiten, Ablagerungsräume), Zahlen (Tiefen, Koordinaten, Georeferenzierung), Linien (Polylines und Polygone) und Muster (unter anderem geologische Patterns) erkannt, und in strukturierte, gemäss von der swisstopo vorgegebenen Datenmodellen, in die dafür vorgesehen Datenbanken abgefüllt werden.

Die Herausforderung besteht darin, die verschiedenen Elemente richtig zu erkennen, zu extrahieren und anschließend auch den richtigen Attributen der Datenbank zuzuweisen. Desweitern wird, speziell in der Geologie, oft nicht eine einheitliche Nomenklatur benutzt, womit mit gezielten Alias-Libraries gearbeitet werden muss.

Speziell bei älteren Dokumenten stellt zusätzlich auch die Scan-Qualität eine große Herausforderung dar.

Entwicklungsstand der Challenge (nicht zutreffenden Text streichen):



• Challenge mit der Führungslinie abgestimmt (keine Ressourcen für Folgeauftrag verfügbar)



# Erläuterungen und Hinweise zum Challenge Canvas

## Geschäftsprozess und -kontext:

dient dazu, den Start-Ups zu vermitteln, **in welchem Kontext bzw. Anwendungsfall** eine Herausforderung oder ein Bedarf auftritt, welche durch eine innovative Lösung zu beheben ist. Wichtig ist es dabei, dies so zu beschreiben, dass es **auch für Aussenstehende verständlich** ist. Insbesondere sollte zudem auch beschrieben werden, was passieren würde, wenn der **Bedarf (weiter) nicht adressiert** würde.

#### **Entwicklungsstand der Challenge:**

informiert Start-Ups darüber, was nach einem Gewinn der Challenge folgen könnte. Für Challenges, für die bereits Ressourcen vorhanden sind, könnte ein Proof of Concept umgesetzt werden. In einem solchen Fall ist es auch wichtig, auszuführen, inwieweit es in der relevanten Verwaltungseinheit adäquate Ressourcen und Strukturen, um einen möglichen Proof of Concept zu begleiten, gibt. Für Challenges, für die (bisher) keine Ressourcen vorgehalten sind, folgen nach Gewinn der Challenge vorerst keine weiteren Arbeiten.

#### Beschreibung der Challenge:

spezifiziert den zu **erfüllenden Bedarf**. Zudem soll auch dargestellt werden, was die Zielsetzung der Lösung ist. An dieser Stelle können auch **Nebenbedingungen oder bestehende Annahmen** für eine mögliche Lösung aufgeführt werden.

### Folgende Fragen können Ihnen helfen, die Challenge zu beschreiben:

- Welches Geschäftsanliegen fordert (neue) Lösungen?
- Wer soll von der neuen Lösung profitieren? Was soll mit der Erfüllung des Bedarfs erreicht werden?
- Gibt es bereits Vorstellungen über mögliche geeignete Technologien oder Lösungsansätze? Falls nein, welche konkrete Herausforderung ist zu lösen bzw. welcher Bedarf ist zu erfüllen?
- Was wollen wir lernen? Was sind Mögliche nächste Schritte?

#### Folgende fiktive Beispiele für Challenges können Orientierung bieten:

- Stark wachsendes Volumen von Anträgen/Anfragen
- Gesuche/Anträge nach Komplexität sortieren
- Prozessintegration mit Partnern
- Intelligenter Assistent zum Erstellen von Pflichtenheften

# Fragen?

<u>alexander.kaempfer@bk.admin.ch</u> uwe.heck@bk.admin.ch