

# Pflichtenheft – BID

## 1. Zielbestimmung

Die zu entwickelnde Software soll es ermöglichen nach einer Suche nach einem Datensatz in einem Geodatenkatalog, dem Nutzer ähnliche Datensätze vorzuschlagen. Des Weiteren soll die Möglichkeit bestehen eigene Datensätze hochzuladen. Das Projekt soll eine „Free and Open Source Software“ sein.

### 1.1 Muss-Kriterien:

- Suchen von Geodatenätzen
- Einfügen/Hochladen von Geodatenätzen
- Vorschlagen ähnlicher Geodatenätze
- Verwalten von Geodatenätzen
- Extraktion von Metadaten

### 1.2 Wunschkriterien:

- Erstellen von Geodatenätzen
- Erleichterung der Suche nach anderen passenden Daten
- Unterstützung des Austausches von Forschungsdaten

### 1.3 Abgrenzungskriterien:

- es werden nur Geodatenkataloge, Metadaten und Datensätze genutzt

## 2. Produkteinsatz

Das Produkt dient zur Suche von Geodatenätzen. Zudem sollen ähnliche Datensätze zu dem gesuchten Datensatz angezeigt werden.

### 2.1 Anwendungsbereiche

Das Produkt wird als FOSS (Free and Open Software Source) zur Verfügung gestellt werden.

### 2.2 Zielgruppen

- Wissenschaftler, die nach ähnlichen oder verwandten Arbeiten suchen (während der Recherche, beim Schreiben von Manuskripten oder beim Auswerten von Beiträgen)
- Betreiber von Repositorien und Katalogdiensten, die Ähnlichkeitsmessungen leicht integrieren wollen

- Datenerhalter, die den Zugang zu Suchergebnissen (Daten, Artikeln) verbessern

## 2.3 Betriebsbedingungen

- Das Produkt wird über einen Desktop-PC mit Internetzugang ausgeführt.

## 3. Funktionale Anforderungen

### Extraction

**FE001:** Das Programm unterstützt die folgenden (Meta-)Daten Formate: GeoPackage, NetCDF, GeoJSON, Shapefile, CSV on the Web, ISO 19xxx, GeoTIFF

**FE002:** Ein CLI Tool, welches es dem Nutzer ermöglicht den räumlichen Umfang der Daten zu extrahieren in verschiedenen Detailleveln (Punkt, Linie, Bounding-Box).

**FE003:** Ein CLI Tool zur Extraktion des Umfangs der zeitlichen Messungen aus einer einzelnen Datei.

**FE004:** Der Nutzer kann den räumlichen und zeitlichen Umfang aus einem Datenverzeichnis extrahieren.

**FE005:** Ein Nutzer kann für eine Datei über eine API die Extraktion der Metadaten anfordern, damit diese für angemeldete Nutzer zur Verfügung stehen.

**FE006:** Die Metadatenextraktion wird automatisch für neue Daten, die der Basissoftware hinzugefügt werden, ausgeführt.

**FE007:** Die Extraktion neuer Metadatenätze während der Erstellung eines neuen Datensatzes erfolgt in einem unabhängigen Prozess.

### API

**FA001:** Alle Nutzerbezogene Funktionalitäten sind über RESTful http API „endpoints“ verfügbar.

**FA002:** API endpoints geben gültige JSON als Antwort zurück, inklusive der Errors.

**FA003:** API endpoints nutzen angemessene HTTP Statuscodes

**FA004:** Raumbezogene Daten aus der API sind in GeoJSON (RFC 7946) kodiert.

**FA005:** Erweiterte Metadaten, die Beispielsweise extrahierte raumbezogene und zeitliche Informationen beinhalten, sind in den regulären Metadaten-Dokumenten mit inbegriffen.

**FA006:** Mit dem hinzugefügten Parameter „similar=n“ zu einer Anfrage um Datensätze zu lesen, bekommt der Nutzer die IDs und einen Ähnlichkeitswert für n ähnliche Datensätze.

## Ähnlichkeitsberechnung

**FS001:** Ein API endpoint stellt mit Hilfe eines Ähnlichkeitswertes einen Vergleich von zwei Datensätzen, basierend auf einer Bounding-Box, zur Verfügung. Die Datensätze werden mit ihrer Verzeichnis spezifischen ID bereitgestellt.

**FS002:** Ein API endpoint stellt eine sortierte Liste mit ähnlichen Datensätzen für eine Repositorium spezifische Datensatz ID. Die Länge der zurückgegebenen Liste kann vom Benutzer definiert werden, eine maximale Länge kann Serverseitig konfiguriert werden.

**FS004:** Der Ähnlichkeitswert ist im Intervall  $[0,1[$  normalisiert.

**FS005:** Der Input Datensatz wird nie in der Ähnlichkeitsliste mit aufgeführt.

**FS006:** Der Ähnlichkeitswert ist höher, wenn der gleiche Datentyp vorliegt.

## UI

**FU001:** Eine konfigurierbare Anzahl an ähnlichen Datensätzen wird auf der Seite angezeigt um einen einzelnen Datensatz anzeigen zu können (darf nicht in die bereits existierende Benutzeroberfläche der Basissoftware eingebunden werden).

## Konfiguration

**FC001:** Alle Konfigurationen von zusätzlicher Funktionalität ist möglich über Textdateien z.B. YAML Format und idealerweise mit Konfigurationsmechanismus von der Basissoftware.

**FC002:** Die Konfiguration ist spätestens nach dem Neustart des Dienstes aktiv.

## Bonus Eigenschaft

Geospatial search and map-based browsing of records.

## 4. Produktleistungen

**FL001:** Die API-Aufruf Antwortzeit eines Aufrufes an ein Repository mit Anzeigen von ähnlichen Datensätzen soll höchstens doppelt so lange dauern wie die Antwortzeit eines Aufrufes ohne Anzeigen ähnlicher Datensätze

**FL002:** Das Aufrufen einer Seite soll in unter fünf Sekunden umgesetzt werden können

**FL003:** Komplexe Seiteninhalte können unabhängig voneinander geladen werden

**FL004:** Interaktive Visualisierungen sollen in unter drei Sekunden reagieren können

**FL005:**

Wartungsfreundlichkeit: Das Projekt läuft unter den gegebenen Lizenzanforderungen der genutzten Software. Es ist ein Open Source Projekt und ist dementsprechend dokumentiert.

## 5. Testdaten

Es gibt eine entsprechende Anzahl von Testdatensätzen. Pro unterstützten Dateiformat gibt es mindestens einen Datensatz. Die Testdaten werden unter einer offenen Lizenz veröffentlicht und ihre Herkunft wird klar dokumentiert. Test- und Demo-Daten werden aus vorhandenen Online-Repositorien entnommen.

Datenquellen:

- "Open" "Nc" on Zenodo
- "geotiff" on <https://pangaea.de>
- "shapefile" on B2SHARE
- "geopackage" on Figshare

## 6. Qualitätsanforderungen

Produktqualität	Sehr Gut	Gut	Normal	Nicht Relevant
<b>Funktionalität</b>				
Angemessenheit			x	
Richtigkeit		x		
Interoperabilität		x		
Ordnungsmäßigkeit		x		
Sicherheit		x		
<b>Zuverlässigkeit</b>				
Reife			x	
Fehlertoleranz		x		
Wiederherstellbarkeit		x		
<b>Benutzbarkeit</b>				
Verständlichkeit	x			
Erlernbarkeit		x		
Bedienbarkeit	x			
<b>Effizienz</b>				
Zeitverhalten		x		
Verbrauchsverhalten			x	
<b>Änderbarkeit</b>				
Analysierbarkeit		x		
Modifizierbarkeit		x		
Stabilität		x		
Prüfbarkeit		x		

## **7. Benutzeroberfläche**

In Abhängigkeit von der Bonusaufgabe wird eine Kartenansicht erstellt, um Daten über die Karte suchen zu können.

## **8. Nicht-funktionale Anforderungen**

keine

## **9. Technische Produktumgebung**

### **9.1 Software**

- Betriebssysteme: Windows (8, 10), MacOS
- Basisprogramm: PyCSW

### **9.2 Hardware**

- Laptop
- PC

### **9.3 Produkt-Schnittstellen**

- RESTful HTTP API endpoints

## Zeitplan ab dem 31.10.18

Aufgabe	Zeitraum
- PyCSW installieren und Daten hinzufügen - Batch und Python lernen	31.10.2018 - 14.11.2018
Gruppenaufteilung: - API - CLI Tool, Extrahierung von Daten	14.11.2018 - 28.11.2018
Programmierung - Ähnlichkeit und Ähnlichkeits-Algorithmus	28.11.2018 - 12.12.2018
Programmierung - Bonus Feature , Demo	12.12.2018 - 26.12.2018
Programmierung - Absprache offener Projektteile, die noch erledigt werden müssen	26.12.2018 - 09.01.2019
Fertigstellung, Testen, Dokumentation fertig schreiben	09.01.2019 - 23.01.2019