Pflichtenheft – BID

1. Zielbestimmung

Die zu entwickelnde Software soll es ermöglichen nach einer Suche nach einem Datensatz in einem Geodatenkatalog oder einem Repositorium, 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. 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 aus einer einzelnen Datei Daten mit Raumbezug zu extrahieren. Diese Daten können unterschiedlich detailliert sein (Punkt, Linie, Polygone, Bounding-Box).

**FE003:** Der Nutzer kann sich einen beliebigen räumlichen Zeitraum aus einer Datei extrahieren.

**FE004:** Der Nutzer kann raumbezogene und zeitliche (Erweiterungen) 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 Metadatensätze 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 Metadatendokumenten 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 ein Vergleich von zwei Datensätzen basierend auf einer Bounding-Box zur Verfügung mit Hilfe eines Ähnlichkeitswertes. Die Datensätze werden als ihr Verzeichnis spezifischer 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. hfffsjjjjfffskfd

**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**