



Treffen des Arbeitskreises Fernerkundung am 06. und 07. Oktober 2022 in Halle (Saale)

Christian Beilschmidt (Geo Engine GmbH), Johannes Dröner (Geo Engine GmbH), Michael Mattig (Geo Engine GmbH), Philip Schweitzer (Geo Engine GmbH), Bernhard Seeger (Geo Engine GmbH, Philipps-Universität Marburg)

Geo Engine: Ein dynamischer Data Cube für interaktive Analysen und Verarbeitungs-Pipelines

Die Geo Engine ist eine Cloud-fähige Geodaten-Analyseplattform, die einen einfachen Zugang zu Geodaten, Verarbeitung, Schnittstellen und Visualisierung bietet. Nutzer können mit der Geo Engine interaktive Analysen durchführen und Verarbeitungs-Pipelines in Form von Workflows erstellen. Hierfür gibt es eine browserbasierte Benutzeroberfläche sowie einen API-Zugang für z.B. Jupyter-Notebooks. In diesem Vortrag zeigen wir die Grundlagen des Systems und seine Eigenschaften. Darüber hinaus geben wir einen Ausblick auf die Integration von Deep Learning für Fernerkundungsdaten.

Der aktuelle Trend der Data Cubes hat ein grundlegendes Ziel: Ein harmonisierter Zugriff auf raum-zeitliche Daten aus verschiedenen Quellen. Sollen mehrere Datensätze in Analysen einbezogen werden gibt es eine große Herausforderung: Unterschiedliche Datenformate, Auflösungen von Sensoren, Koordinatenreferenzsysteme und Datentypen müssen kombiniert werden. Data Cubes lösen dieses Problem, indem sie alle Daten in ein einheitliches Format und eine einheitliche Auflösung konvertieren. Der Nachteil dieses Ansatzes ist jedoch, dass Ergebnisse nur für konkrete Anwendungsfällen geeignet ist. Wenn beispielsweise eine nur etwas anderen Auflösung benötigt wird muss der Data Cube neu erstellt werden.

Der Ansatz der Geo Engine hingegen besteht darin, die Daten ad-hoc innerhalb eines Workflows zu harmonisieren, anstatt sie im Voraus zu bearbeiten und neu zu speichern. Der Vorteil ist, dass stets die Original-Daten angesprochen werden können und kein Informations- oder Präzisionsverlust entsteht. Die Geo Engine ist aus Anwendersicht ein virtueller, dynamischer Data Cube. Der Anwender spezifiziert eine Anfrage mit gewünschtem Gebiet, Zeitraum und Auflösung und die Geo Engine kombiniert und homogenisiert automatisch alle Datenquellen sowohl räumlich als auch zeitlich. Durch Caching-Strategien und das Wiederverwenden von Teilergebnissen können wir in der Praxis eine ähnliche Performance wie statische Data Cubes bieten bei gleichzeitig viel mehr Flexibilität.

Wir stellen eine erweiterbare Verarbeitungs-Toolbox zur Verfügung, die z.B. Operatoren für Filter und Datenkombination enthält. Durch parametrisierte und wiederverwendbare Workflows können Analysen direkt in den Produktivbetrieb überführt werden. Ist ein Workflow einmal definiert, kann er für verschiedene räumliche Regionen oder verschiedene Zeitpunkte wiederverwendet werden. Workflows sind sowohl für kurz- als auch für langlaufende Aufgaben geeignet und skalieren für beliebig große Datensätze. Um dies zu realisieren, nutzt die Geo Engine asynchrone und chunk-basierte Berechnungen, z.B. auf Raster-Kacheln. Dies ermöglicht zum Beispiel KI-Workflows, bei denen die Daten von der Vorverarbeitung in das Modelltraining fließen. Die daraus resultierenden maschinellen Lernmodelle können anschließend als Operatoren in Workflows wiederverwendet werden.

Die Geo Engine GmbH ist ein Start-up von Informatik- und Geographie-Forschern an der Universität Marburg. Sie entwickelt die Geo Engine als Open-Source-Projekt und bietet einige Enterprise-Features unter einer kommerziellen Lizenz an. Die Geo Engine wird in verschiedenen Forschungs- und Infrastrukturprojekten (z.B. NFDI4BioDiversity) eingesetzt und läuft bereits in kommerziellen Anwendungen.