# SmartDataPorter NetCDF Parser

Der NetCDF Parser vom SmartDataPorter ermöglicht den Import von Daten aus NetCDF Dateien in von SmartData unterstützte Datenbanken.

Das NetCDF Format ist ein Datenformat für den Austausch von großen Mengen an wissenschaftlichen Daten. Die Datenstruktur, in welcher die Daten organisiert sind, ist n-Dimensional.

Wenn die Daten in der NetCDF geordnet abgespeichert wurden, ist ein Zugriff auf Ausschnitte der Daten möglich. Die Daten können jedoch prinzipiell auch unsortiert gespeichert worden sein, dann ist mit dem vorliegenden NetCDF Parser kein Import von konkreten Datenteilen möglich.

## Beispiel für NetCDF4 Daten

Ein Beispiel für NetCDF4 Daten sind die Leaf Area Index Daten des Copernicus Programms der ESA. Diese Dateien enthalten einen Index der die Ausgeprägtheit des Bewuchses in einem 300m x 300m Raster über die gesamte Erde angibt. (Wasserflächen sind ausgeschlossen und durch den Wert -1 präsentiert) Die Werte für den LAI liegen zwischen 0 (niedrigster) und 210 (höchster) und geben die stärke des Bewuchses an.

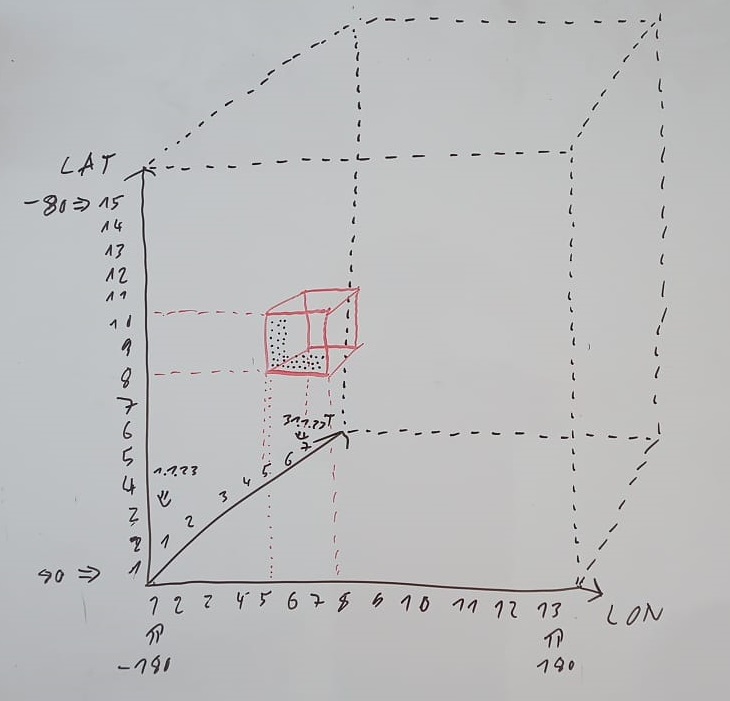
Obwohl die LAI Dateien immer nur Werte für einen Zeitpunkt beinhalten (alle 10 Tage eine kummulierte Messung) enthält die Datenstruktur als Dimension 1 die Zeit. Diese besitzt einen ungültigen Wert, der mit dem Wert der Messung (dieser ist nur aus dem Dateinamen ersichtlich) überschrieben werden muss.

Die nächste Dimension ist die Latitude (LAT) welche in den Daten in **absteigender** Reihenfolge gespeichert sind.

Die dritte Dimension ist die Longitude (LON) die die **aufsteigender** Reihenfolge gespeichert sind.

Da die Messungen auf ein 300m x 300m Raster berechnet werden, sind die in WGS84 angegebenen Koordinaten in LAT und LON nicht äquidistant.

Die vierte Dimension enthält den LAI Indexwert neben ein paar anderen Werten zur Qualitätsbestimmung und Nachvollziehbarkeit der Berechnung des LAI.



Quele und weitere Informationen: [Leaf Area Index | Copernicus Global Land Service](https://land.copernicus.eu/global/products/LAI)

## Erläuterung der Funktionsweise

Der Parser liest zunächst jeder der Dimensionen T,LAT,LON (in dieser Reihenfolge) ein und erstellt einen Index. Während der Erstellung des Indexes werden außerdem die Positionen der Minima und Maxima des gewünschten zu importierenden Bereiches für jede Dimension ermittelt.

Beispel:

Importiert werden sollen von LAT 12.5 bis LAT 13:

Es werden die LAT in der Dimension LAT gesucht:

* Minimum: Der LAT Wert, der am nächsten an 12.5 dran ist und über 12.5 ist.

{i: 100, v: 12.51231}, {i: 101, v: 12.50102}, {i: 102, v:**12.5001}**, {i: 102: v;12.49978}, {i: 103, v:12.48123}

* Maximum: Der LAT Wert, der am nächsten an 13 dran ist und unter 13 ist.

{i: 50, v: 13.10232}, {i:51, v: 13.01313}, {i: 52, v: **12.9923}**, {i: 53, v: 12.94132}, {i: 54, v: 12.82332}

Die Indexwerte von Minima und Maxima werden erfasst und in eine „Spec“ gechrieben, welche den Wahlzugriff in die Datei steuert.

Beispiel:

Spec: :,52:102,100:150

Der erste Doppelpunkt gibt den Zugriff auf die erste Dimension (T) an. Hier werden alle Daten der Dimension gelesen (da im Beispiel der LAI Dateien nur ein Wert existiert)

Die Angabe 52:102 gibt den Zugriff auf die zweite Dimension (LAT) an. Hier werden nur die Daten (LAI Werte) genommen die sich in einer Dimension unterhalb der Dimensionen 52-102 von LAT befinden.

Die Angabe 100:150 gibt den Zugriff auf die dritte Dimension (LON) an. Hier werden nur die Daten (LAI Werte) genommen, die sich in einer Dimension unterhalb der Dimensionen 100-150 befinden.

Die Spec ist verkettet zu lesen, also: „Alle Datenwerte (LAI) die sich im Bereich 100-150 der LON befinden, die sich wiederrum im Bereich 52-102 der LAT befinden, die sich wiederrum in egal welchem T befinden.

Mit dieser Spec wird über die netcdfall Bibliothek der gewünschte Datenbereich aus der Datei ausgelesen. Die Ausgelesenen Daten werden in einem n-Dimensionalen Array zurückgeliefert, dabei enthält die unterste Ebene den gewünschten LAI-Wert. Damit die anderen Werte (T, LAT, LON) wieder zugeordnet und ein Datensatz gebildet werden kann, muss dieses Array rekursiv durchlaufen werden.

Anhand des Index im Array und der Verschachtelung werden die Indizes der einzelnen Dimensionen mit Hilfe des zuvor gebildeten Werteindex in die zugehörigen Werte übersetzt.

Anschließend werden die so zusammengebauten Datensätze in die Datenbank gechrieben.