

Metadatenpflege: Informationen aus XML-Dateien extrahieren und in unsere Datenbank schreiben.

Aufgabe:

- Das DKRZ betreibt als Teil des IPCC Data Distribution Centres (<http://ipcc-data.org>) das Referenzdatenarchiv für die globalen Klimamodellrechnungen.
- Die Informationen zu den verwendeten gekoppelten Klimamodellen in AR5 wurden als XML files in der Datenbank abgelegt. Deren Inhalte sollen für Datennutzer einfacher zugänglich gemacht werden. XML-Bsp.: http://cera-www.dkrz.de/WDCC/CMIP5/downloadAddInfo?acronym=QC_add_info_2283414&id=2
- Dazu müssen die XML files analysiert werden, um daraus Modellname, Modellkurzbeschreibung, Modellkomponenten, Modellauflösung, Referenzen zu den Modellen usw. zu extrahieren.

Stelle:

- Wir brauchen eine Studenten mit geowissenschaftlichem Hintergrund.
- Die Aufgabe ist abgeschlossen. Wir haben aber weitere Aufgaben im Bereich Metadatenpflege, so dass eine längere Beschäftigung möglich ist.
- Ansprechpartner: Martina Stockhause (stockhause@dkrz.de)