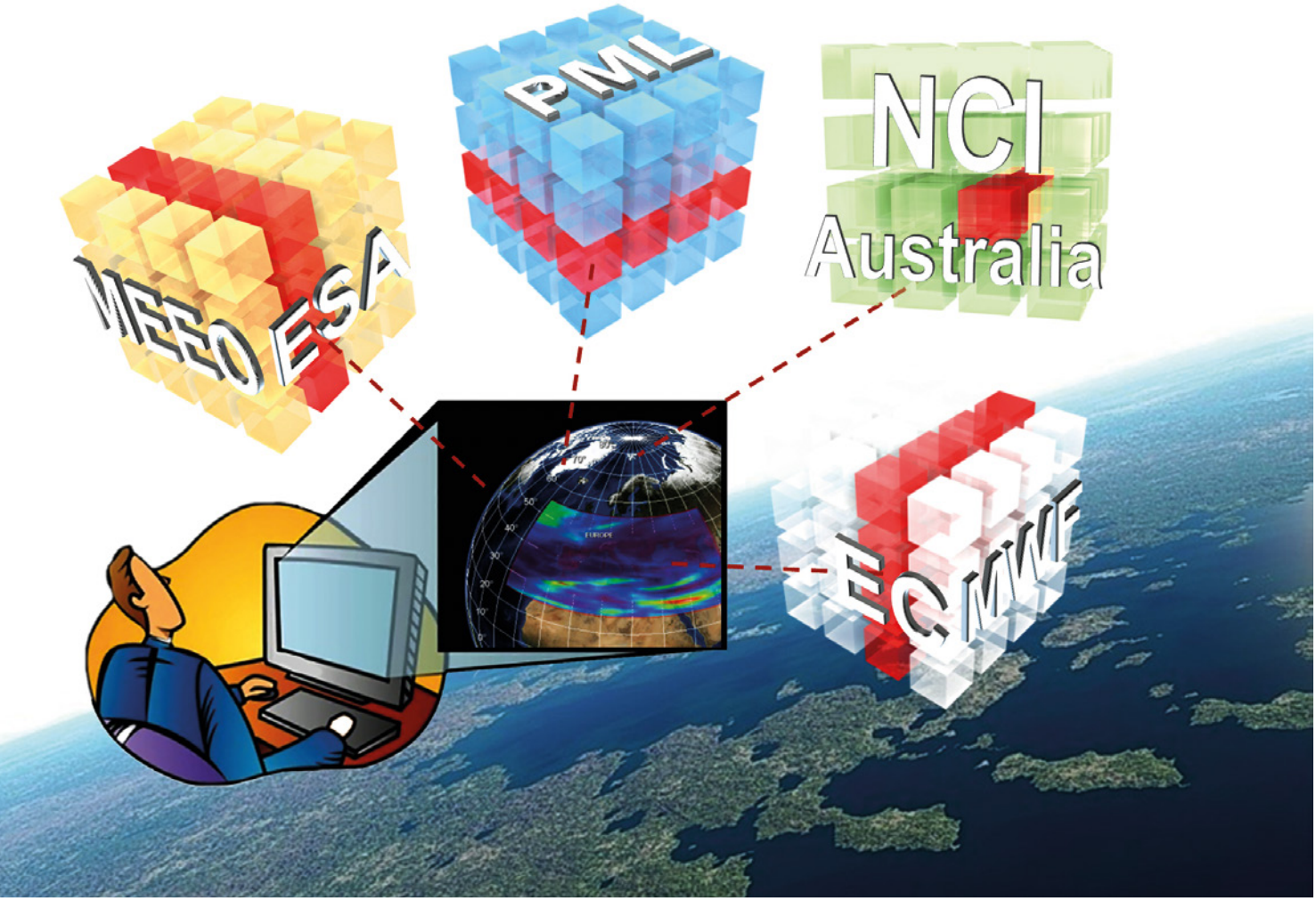

Quelle: Rasdaman

Big Data per Mausklick – Erfolgsstory EarthServer

Die internationale EarthServer-Initiative geht in ihre zweite Phase und hinterlässt bereits heute ihre Spuren in der Geodatenwelt. Bereits im ersten Teil setzte das Projekt neue Maßstäbe bei der Verarbeitung von Hunderten von Terabyte an Erdbeobachtungsdaten. In der zweiten Phase wird es Wissenschaftlern durch den Einsatz einer innovativen Rasterdaten-Technologie erstmals möglich, Petabyte-Datenwürfel aus einer Vielzahl von Quellen zu kombinieren und zu analysieren.

Autorin: **Heike Hoenig**

Erfolgreiche Forschungsprojekte gibt es einige, doch nur wenige schaffen den Quantensprung, zukunftsträchtige Veränderungen hervorzubringen und die Datenwelt nachhaltig zu revolutionieren. Die global tätige EarthServer-Initiative ist solch ein Projekt. Begonnen hatte EarthServer im Herbst 2011, gefördert von der Europäischen Kommission im 7. Forschungs-Rahmenprogramm. Bei der Evaluierung dieser Phase 1 vergaben die internationalen Gutachter Ende vergangenen Jahres Bestnoten: Die EarthServer-Plattformtechnologie, das Array-Datenbanksystem Rasdaman, würde „Datenzugriff und -nutzung signifikant transformieren" und das Projekt habe „durch seinen Einfluss auf die Standardisierung zweifelsohne die Big-Earth-Data-Landschaft geformt".

Auf dem Erfolg der ersten Projektphase aufbauend, bewarb sich EarthServer erfolgreich um die Genehmigung der nächsten Phase. EarthServer-2 begann im Mai 2015. Der Schwerpunkt des Projekts liegt weiterhin auf dem Datenwürfel-Paradigma und der Verwaltung noch größerer Datenmengen. Gemäß dem Motto: „Ein Datenwürfel sagt mehr als eine Million Bilder" stehen nun multidimensionale Analysen mit Datenwürfeln in Petabyte-Größenordnungen auf der Agenda.

Vision war und ist es, mit nur wenigen Mausklicks direkt auf den riesigen „Big Earth Data"-Archiven zu arbeiten. Generell können die Datenwürfel unter anderem aus Sensordaten, Bildern, Simulationsergebnissen und Statistiken bestehen, oft mit einer zusätzlichen zeitlichen Dimension. Außerdem können die Daten sowohl regelmäßige als auch unregelmäßige raumzeitliche Gitter bilden. Alle Daten werden einheitlich als OGC Coverages modelliert und über die OGC-Dienste Web Coverage Service (WCS) und Web Coverage Processing Service (WCPS) bereitgestellt [1]. Zur effizienten, performanten Bereitstellung dieser Datenwürfel benutzt EarthServer die Array-Datenbank [2] Rasdaman [3]. Unter Ausnutzung der Server-Kapazität kann damit über beliebige Endgeräte mit standardkonformen Clients auf nahezu unbegrenzten Datenvolumina gearbeitet werden. Neben der höheren Benutzerfreundlichkeit ermöglichen diese Datenwürfel auch die Integration von Daten aus verschiedenen Disziplinen, sodass Wissenschaftler zum Beispiel ihre Messdaten mit den aus Simulationen gewonnenen Informationen kombinieren können.

Den Hintergrund zur Projektidee bildeten zwei nach wie vor aktuelle Faktoren: Zum einen herrscht ein großer Bedarf an neuen Verwaltungskonzepten für die Datenflut, die uns überrollt. Zum anderen ist mit der Rasterserver-Technologie Rasdaman eine Datenwürfel-Lösung verfügbar, die die notwendigen Funktionen für solche Auswertungen mitbringt. So werden in EarthServer weiterentwickelte Datenwürfel und spezielle Web-Portale aufgebaut, mit denen Benutzer geowissenschaftliche Informationen in vielfältiger Weise, etwa als eindimensionale Diagramme, zweidimensionale Karten, dreidimensionale Bild-Zeitreihen und 4D/5D-Klimadatenwürfel, extrahieren und visualisieren können. Dabei setzt das Projekt konsequent auf offene Standards, um flexible, skalierbare Dienstleistungen auf multidimensionalen Daten sämtlicher Erdwissenschaften anzubieten.

Projektpartner der Phase 2 sind:

- Plymouth Marine Laboratory, Großbritannien (Analyse ozeanographischer Daten),
- Europäisches Zentrum für mittelfristige Wettervorhersage (ECMWF), Großbritannien (Analysen mit 4D/5D-Klimadaten),
- MEEO s. r. l., KMU, Italien (Analyse von ESA-Satellitenbilddaten),
- National Computational Infrastructure (NCI), Australien (Analyse von Satellitenbilddaten),
- Jacobs University, Deutschland (Bereitstellung und Weiterentwicklung der Plattformtechnologie sowie Analysen mit planetaren Daten),
- Rasdaman GmbH, Deutschland (Bereitstellung und Weiterentwicklung der Plattformtechnologie Rasdaman),
- CITE S. A., KMU, Griechenland (Metadaten-Integration) sowie
- Nasa (Datenvisualisierung mit dem virtuellen Globus Nasa WorldWind).

Um die großen Erddatenmengen effizient zu verwalten, wurden in EarthServer die bestehenden Technologien und Standards weiterentwickelt, insbesondere im Bereich Array-Datenbanken. Im Prinzip wurde dazu die Datenbanksprache SQL mit der Signal- und Bildverarbeitung zusammengeführt, um das Konzept „jede Frage, jederzeit" zu ermöglichen. Mit dieser neuen Herangehensweise entsteht mittlerweile eine neue ISO-Norm, SQL/MDA, welche SQL um „Multi-Dimensional Arrays" erweitert. Parallel dazu wurden verschiedene OGC-Standards weiterentwickelt, insbesondere die OGC-Coverage-Definition, der OGC WCS sowie die OGC-Raster-Anfragesprache WCPS. Auch die Initiative zur Vereinheitlichung europäischer Geodateninfrastrukturen, Inspire, wird durch die neu entstandenen Spezifikationen nachhaltig beeinflusst: Derzeit wird WCS – unter Mitwirkung von EarthServer – als Inspire Coverage Download Service etabliert.

Hinter diesen standardbasierten Schnittstellen leistet die Rasdaman-Technologie die Hauptarbeit. Bereits heute sind Rasdaman-Datenbanken mit weit über 100 Terabyte im Einsatz und ankommende Datenwürfel-Anfragen wurden auf über 10 000 Cloud-Knoten parallelisiert. Die verschiedenen EarthServer-Portale sind miteinander vernetzt und können Daten austauschen, wenn Anfragen – beispielsweise zur Kombination zweier Datenwürfel – dies erfordern. Dabei wird jede ankommende Anfrage automatisch analysiert, um den effizientesten Datenstrom und damit die schnellste Beantwortung zu garantieren. So entsteht ein weltweites Geflecht, eine Föderation von Rasdaman-Installationen, welche dem Nutzer zusammen einen einzigen vereinheitlichten Datenraum anbieten.

Viele Institutionen und Forschungszweige können künftig von diesen neuen Geodiensten profitieren, denn unsere Möglichkeiten, die Erde zu beobachten, steigen damit gewaltig. Wichtige Institutionen wie der Deutsche Wetterdienst sind interessiert, der EarthServer-Datenföderation beizutreten. „Big Data per Mausklick" lautet die Vision der EarthServer-Initiative – und diese Vorstellung liegt zum Greifen nah.

Quellen

[1] https://de.wikipedia.org/wiki/Web_Coverage_Service

[2] https://en.wikipedia.org/wiki/Array_DBMS

[3] http://rasdaman.org

Autorin:

Heike Hoenig

Wissenschaftsjournalistin

I: www.rasdaman.de

E: hoenig@rasdaman.com