

**wissens
management**
Das Magazin für Führungskräfte

wissens

Das Magazin für Führungskräfte

management

Titelthema

22

Kommunikation im Web 2.0: Pods, Blogs und RSS

Praxis Wissensmanagement

10

Der Heureka!-Effekt: Vom Wissen zur Innovation

Dokumentation + Kommunikation

34

Klimaforschung via Internet

Human Resources

41

Den ständigen Wandel leben

Trends

48

Durch vernetztes Wissen zum strategischen Erfolg

Wissen bewegt
Fachkongress mit Event-Charakter
Jetzt anmelden unter www.wissensmanagement-tage.de

Rolf Bastian

Klimaforschung via Internet

Daten der Klima- und Umweltforschung allgemein verfügbar zu machen, ist das Ziel des elektronischen Langzeitarchivs Pangaea – Publishing Network for Geoscientific and Environmental Data. Es wurde vom Alfred-Wegener-Institut für Polar- und Meeresforschung in Bremerhaven gemeinsam mit dem Zentrum für Marine Umweltwissenschaften (MARUM), Bremen, als Informationssystem für Mess- und Beobachtungswerte aufgebaut. In diesem Data Warehouse sind Milliarden exakter Messwerte aus der Erdsystemforschung gespeichert, auf die ansonsten kaum systematisch zugegriffen werden könnte. Sie stehen über Internet-Portale der Wissenschaft für weiterführende Analysen zur Verfügung. Das Bibliothekskonzept ermuntert immer mehr Wissenschaftler, ihre Forschungsdaten einzustellen, sodass der Bestand rapide wächst. Damit hat sich Pangaea zu einer zentralen Bibliothek für eine Vielzahl von geowissenschaftlichen Disziplinen entwickelt.

Wie werden Messwerte zugänglich?

Die Klima- und Umweltforschung ist nach dem Erscheinen des aktuellen IPCC-Reports einmal mehr in den Blickpunkt des öffentlichen Interesses geraten. Die Erkenntnis, dass der Menschheit offensichtlich eine hausgemachte Klimaveränderung bevorsteht, hat in Politik und Wissenschaft intensive Debatten angestoßen. Die Wissenschaft muss zudem weiterhin Prognosen verifizieren und Zusammenhänge noch detaillierter untersuchen; schließlich sind viele Prozesse des überaus komplexen Klimasystems nach wie vor erst ansatzweise bekannt. Deshalb entwickeln Forscher immer wieder neue Fragestellungen, um die Ver-

änderungen der Umwelt besser zu verstehen. Welchen Einfluss hat beispielsweise die Verteilung des Planktons in den Weltmeeren auf das Klima? Was bewirkt die Konzentration von Ozon und Partikeln in der Erdatmosphäre? Wie hat sich die Zusammensetzung der Erdatmosphäre in historischen Zeiträumen verändert? Zur Beantwortung gerade von Fragen mit globalem Ansatz ist es besonders wichtig, die Auswertung der vielen Milliarden Messwerte, die an verschiedenen Orten der Welt gewonnen werden, zu unterstützen.

Die Archivierung aller Daten in entsprechend leistungsfähigen Systemen mit standardisierten Beschreibungen ist jedoch in der Wissenschaft noch lange nicht selbstverständlich. Zwar steht im Prinzip eine permanent wachsende Menge an Daten zur Verfügung, die auch teilweise in zahlreichen Archiven weltweit gespeichert sind. Allein die Werkzeuge des Alfred-Wegener-Instituts, wie das Forschungsschiff „Polarstern“ oder die Observatorien der Neumayer-Station in der Antarktis, liefern auf jeder Expedition Millionen von Messwerten. Diese Forschungsarbeiten stellen wiederum nur einen Bruchteil der Projekte zur Erdsystemforschung auf nationaler, europäischer und internationaler Ebene dar.

Genau diese Fülle steht jedoch mittlerweile einer weiterführenden wissenschaftlichen Erkenntnis eher im Wege. „Daten werden mit viel Aufwand und hohen Kosten gewonnen, sind aber häufig nicht langfristig und allgemein verfügbar“, beschreibt Dr. Hannes Grobe, wissenschaftlicher Mitarbeiter und „Data Librarian“ von Pangaea am Alfred-Wegener-Institut das Problem. Ein Teil von ihnen wird den Publikationen in Tabellenform angehängt. Aber nur wenige sind in allgemein zugänglichen elektronischen Speichern abgelegt – und dann

Kurz gefasst:

- Das Alfred-Wegener Institut für Polar- und Meeresforschung führt wissenschaftliche Projekte in der Arktis, Antarktis und in Meeresgebieten der gemäßigten Breiten durch.
- Ziel ist es, die klimatischen Veränderungen zu erforschen.
- Alle evaluierten Daten werden in dem wissenschaftlichen Langzeitarchiv Pangaea dokumentiert und für die weitere Verwendung aufbereitet.

meist in sehr unterschiedlichen Strukturen und Formaten, die keine systematische Auswertung erlauben. Zwar gibt es diverse internationale Standardisierungsprojekte, die aber Zeit brauchen.

„Mit der zunehmenden Datenfülle seit den 80er Jahren wurden viele Daten überhaupt nicht mehr publiziert, geschweige denn abgedruckt. Da sie nicht zentral verfügbar sind, gehen sie der wissenschaftlichen Forschung verloren“, konstatiert Dr. Michael Diepenbroek, Geschäftsführer des World Data Center for Marine Environmental Sciences (WDC-MARE) in Bremen. Um die Wissenschaft optimal zu unterstützen, ist deshalb ein gezieltes Wissensmanagement unabdingbar. Diepenbroeks Schlussfolgerung: „Angesichts der rapide wachsenden Menge von wissenschaftlichen Publikationen und Daten ist mehr Prägnanz der Informationen notwendig – und am prägnantesten sind nun mal strukturierte Datenmodelle mit einem freien Zugang zu ihren Inhalten.“

Datenpublikation im Internet

Datenpublikation und zitierfähiger Datensatz heißen die Schlüsselbegriffe des Konzepts, das die Verfügbarkeit verbessern soll. Entwickelt wurde es am WDC-MARE, das als Publikationssystem die öffentlich zugängliche Datenbibliothek Pangaea verwendet. Bereits während der Konzeption hat das Team die Voraussetzungen geschaffen, um Pangaea für eine Veröffentlichung im bibliografischen Sinne verwenden zu können. Grundsätzlich werden die Messwerte mit den be-

schreibenden Metadaten (zum Beispiel: Wer hat die Daten erhoben? Um welchen Messwerttyp handelt es sich?) gemeinsam gespeichert. Alle Daten sind georeferenziert in Zeit und Raum (geografische Breite und Länge, Höhe bzw. Tiefe, Datum/Uhrzeit oder auch geologisches Alter). Über die Metadaten wird außerdem erkennbar, wann, von wem, wie und in welchem Medium (Wasser, Luft, Eis, Sediment etc.) ein Wert gemessen wurde.

Neue Messgrößen sind leicht zu integrieren, denn das Datenmodell kann jederzeit erweitert werden. Damit lässt es sich neuen wissenschaftlichen Entwicklungen anpassen. Soll etwa eine Planktonart als zusätzlicher Parameter aufgenommen werden, muss der Anwender in der Tabelle keine weitere Spalte anlegen (was die Struktur verändern würde). Er definiert einfach in einer neuen Zeile, wie dieser Typ zu interpretieren ist. Über einen internen Namen wird die Referenz auf die große Messwert-Tabelle hergestellt, die den Wert selbst enthält. Damit ist die Tabelle immer gleich aufgebaut. Gerade dies ist eine wesentliche Voraussetzung für vielfältige Anfragen an ein wissenschaftliches Data Warehouse, welches die unterschiedlichsten Messwert-Typen mit einbeziehen soll.

Neben Datensätzen und Tabellen sind in Pangaea auch binäre Objekte wie Bilder oder Modelle gespeichert. Auch sie sind georeferenziert und in der Datenbank recherchierbar. Eine URL verweist auf das Objekt selbst. „Mit diesem Konzept erlaubt Pangaea die Erfassung nahezu aller in der naturwissenschaftlichen Grundlagenforschung anfallenden geografisch und zeitlich einzuordnenden Daten“, betont Grobe. Die technische Grundlage dieses Konzepts ist eine spezielle analytische Datenbank (Sybase IQ). In ihr werden die validierten, publizierten und danach unveränderlichen Daten gespeichert. Dass Pangaea eine Vielzahl von Attributen und Messwerten in einer einzigen großen Tabelle zur Verfügung stellen kann, basiert nicht zuletzt auf der hohen Komprimierungsfähigkeit sowie dem Indizierungsprinzip dieser Datenbank: Da vorab keine Schlüsselbegriffe definiert werden müssen, sondern jedes Feld automatisch als Index dienen kann, sind Recherchen sehr flexibel. Egal

Neue wissenschaftliche Erkenntnisse in der Klimaforschung

Nährstoffe im Ozean beeinflussen das Plankton-Wachstum, das wiederum in Wechselwirkung mit dem Klima steht – somit ist die Planktonverteilung in den Weltmeeren zum Beispiel ein wichtiger Parameter in der Klimaforschung. Um festzustellen, wie sich etwa der Anteil von Nitrat und Silikat in den 1980er Jahren verändert hat, wählte Dr. Rainer Sieger, wissenschaftlicher Mitarbeiter und Datenkurator beim Alfred-Wegener-Institut für Polar- und Meeresforschung, am Bildschirm aus 46.000 Parametern diese beiden Typen aus. Als Zeitraum legt er 1. Januar 1980 bis 31. Dezember 1989 fest, der Vergleichbarkeit halber nur Messwerte aus einer Tiefe von 45 bis 55 Metern unter dem Meeresspiegel. Ein Knopfdruck – und Sekunden später ist aus rund 1,6 Milliarden Daten im Informationssystem Pangaea die entsprechende „Datenfläche“ des Ozeans selektiert. Die gewonnene Tabelle wird mit einer speziellen Grafikanwendung aufbereitet und Sieger kann auf einer Weltkarte die Konzentration der gesuchten Nährstoffe in den Ozeanen studieren.

Ein weiteres Beispiel der Analysefähigkeiten von Pangaea demonstriert Dr. Christian Schäfer-Neth, wissen-

schaftlicher Angestellter im Rechenzentrum des Alfred-Wegener-Instituts, mit der Auswertung von Daten der Neumayer-Station in der Antarktis. Das luftchemische Observatorium kann in der sehr sauberen Luft der Polarregion unbeeinflusst von lokalen menschlichen Einflüssen langfristige Entwicklungen besonders genau verfolgen. Gemessen werden in bestimmten Abständen (Minuten und Stunden) unter anderem Ozon, Partikel, Treibhausgase, Spurengase, Aerosole und Trübung. Neben den aktuellen Luftwerten lassen sich durch Bohrungen im Eis Entwicklungen in der Erdatmosphäre mehrerer hunderttausend Jahre zurückverfolgen.

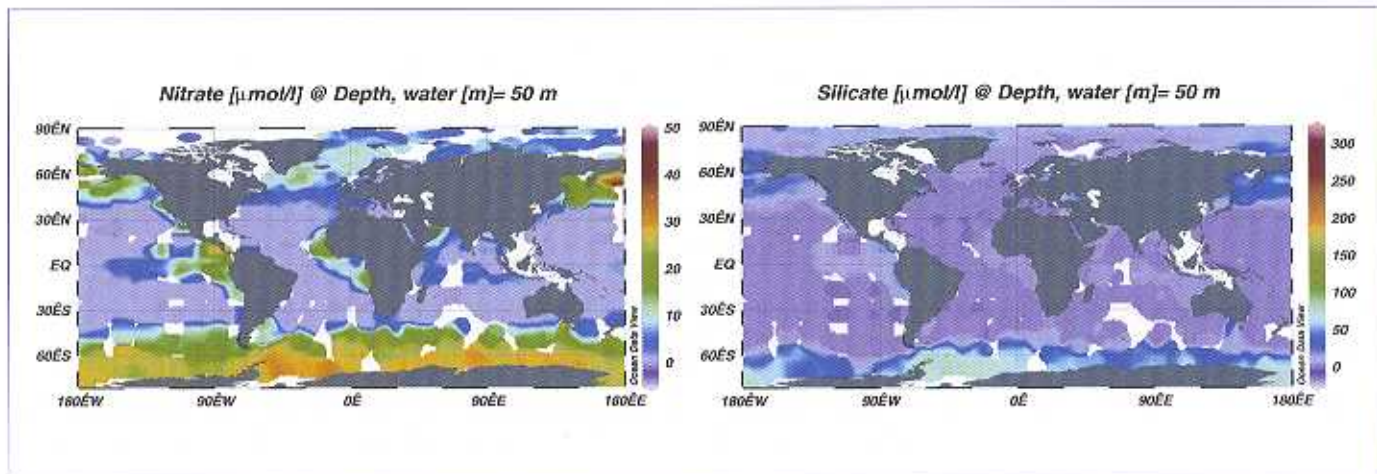
Alle gewonnenen Werte werden in Pangaea archiviert und erlauben so die Verschneidung auch extrem langer Zeitreihenbetrachtungen aus historischen und aktuellen Werten. Um beispielsweise die Ozon- und Partikelkonzentration in den 90er Jahren festzustellen, gibt Schäfer-Neth diese beiden Parameter zusammen mit der Zeitdefinition (1/1990 bis 12/1999) ein – und erhält in Sekundenschnelle aus rund elf Millionen Luftchemie-Daten ein Diagramm mit den gewünschten Mittelwerten.

welche Fragestellung kommt – die Datenbank bietet immer Zugriffshilfen. Zudem sind Abfragen auch bei extrem großen Datenmengen jederzeit und sehr schnell möglich.

Bibliothekskonzept genutzt

Um Wissenschaftler zukünftig zu motivieren, ihre Daten über entsprechende Systeme verfügbar zu machen, müssen die Daten im bibliografischen Sinne zitierfähig sein. Dazu bildet Pangaea den etablierten Arbeitsfluss Autor – Verlag – Bibliothek ab. Um eine zuverlässige Langzeit-Verfügbarkeit zu garantieren, wurde ein Code, der seit einigen Jahren auch

für die Kennzeichnung wissenschaftlicher Publikationen verwendet wird, auf die Datenwelt übertragen. Seit 2004 erhält jeder veröffentlichte Datensatz einen Digital Object Identifier (DOI). Dieser Schlüssel wird von der TIB Hannover (der größten wissenschaftlich-technischen Bibliothek der Welt) vergeben. Damit ist der Datensatz zuverlässig auffindbar und im bibliografischen Sinn zitierfähig. Pangaea ist mit weiteren Partnern Initiator der Verwendung des DOI-Systems für die Publikation von Daten. Der Forscher kann somit in wissenschaftlichen Publikationen sachgerecht zitiert werden – und der Nutzer der Daten hat die Gewissheit, dass sie vor der Veröffentlichung validiert wurden. Mit DOI



Weltkarte mit der Konzentration von Nitrat und Silikat in den Ozeanen auf Basis einer Pangaea-Analyse

als Teil des Zitats können die Leser den Datensatz jederzeit über das Internet direkt aufrufen.

Die Daten in Pangaea werden in einem Datenbank-Server verwaltet. Interessenten können darauf über Suchmaschinen oder Internet-Portale zugreifen. Darüber hinaus entwickeln Diepenbroek und der Diplomphysiker Uwe Schindler von der Universität Bremen derzeit ein universelles Zugriffs-Werkzeug, mit dem beliebige Teilmengen aus dem Datenpool extrahiert werden können. „Damit haben Nutzer, die eine Vielzahl von Messgrößen einzeln oder in beliebiger Kombination benötigen, die Möglichkeit, diese in jeder raumzeitlichen Menge mit der gewünschten Filterung herausziehen – mit Antwortzeiten im Sekundenbereich“, beschreibt Diepenbroek den entscheidenden Fortschritt. Wissenschaftler können auf diese Weise individuell zusammengestellte Teilmengen zum Beispiel von Zeitreihen betrachten. Auch dafür bietet die analytische Datenbank Sybase IQ durch eine effiziente Möglichkeit zur Indizierung von Zeit-Informationen eine technische Unterstützung.

Zentralarchiv für wissenschaftliche Daten

Zeitreihen, ozeanografische und seismische Profile, Sedimentprofile, geologische Karten – die Daten aus Pangaea lassen sich nahezu unbegrenzt auswerten. Und Fülle wie Vielfalt wachsen ständig. Hat Pangaea ursprünglich mit Paläoklimadaten des Alfred-Wegener-Instituts

begonnen, hat es sich mittlerweile zu einem zentralen Archiv für nahezu alle wissenschaftlichen Geodaten entwickelt. Es entfaltet eine Sogwirkung auf externe Datensammlungen, die teilweise Fördermittel erhalten, um ihre Daten in Pangaea zu archivieren. Beispielsweise wurde inzwischen eines der weltweit umfangreichsten Archive für Baumringe eingebracht – ein wichtiger Beitrag für die Erforschung des historischen Klimas. Bei zahlreichen Projekten auf nationaler, europäischer und internationaler Ebene wird Pangaea von vornherein mit dem Datenmanagement betraut.

Der Inhalt des Systems ist in den vergangenen Jahren exponentiell gewachsen: Seit 1996 hat sich die Datenmenge jedes Jahr verdoppelt. Im April 2007 waren 138 Forschungsprojekte mit 530.000 Datensätzen und 1,7 Milliarden Datenpunkten aufgenommen. Dadurch wurde ein Datenvolumen von 1,2 Terabyte erreicht. Und die Datenmenge wächst permanent weiter. So liefern einige Beobachtungsstationen neue Werte im Minutentakt. Die im Aufbau befindlichen Sensornetze aus Ozeanografie, Meteorologie und Geophysik werden den Datenfluss nochmals um ein bis zwei Größenordnungen erhöhen.

„Indem wir eine solche Menge an Daten zusammen mit den Metadaten in einem zentralen System mit einem hohen Ordnungsgrad speichern, haben wir eine enorme Effizienzsteigerung erreicht“, resümiert Grobe. „Es gibt kein anderes wissenschaftliches Datenmodell, das mit solchen Volumina bei gleichzeitig hoher

Komplexität klarkommt. Da unterschiedlichste Aspekte in einer einzigen Abfrage verknüpft werden können, sind damit ganz neue wissenschaftliche Fragestellungen möglich. Wir können so die Erforschung unserer Erde mit einem neuen effizienten Werkzeug zur Archivierung und Verarbeitung der ständig größer werdenden Datenmengen unterstützen.“

Literatur:

- [1] www.pangaea.de
- [2] www.awi.de
- [3] www.wdc-mare.org
- [4] www.rcm.marum.de



Der Autor:

Rolf Bastian ist Partner in der Unternehmensberatung KD+M in Montabaur. Seine Schwerpunkte liegen in den Bereichen Kommunikation für High-Tech-Themen und Demografieberatung.

bastian@wissensmanagement.net