

**KLIIN**

31. August 1998

The cover image shows a weather station with various sensors and a wind vane against a blue sky with scattered white clouds. In the background, a cityscape is visible under a hazy sky.

# Abschlußbericht Multimediale Klimainformation im Internet

**Sabine Reimers, Gerrit Henken**

Deutsches Klimarechenzentrum GmbH

Bundesstraße 55

D-20146 Hamburg

## Inhalt

1 Einführung	1
2 Projektbeschreibung: Aufgaben und Ziele	2
2.1 Offener Zugang zu Klima- und Global Change-Informationen	2
2.1.1 Bestandsaufnahme der verfügbaren Klimainformationen	2
2.1.2 Datenbank-WWW Kopplung	3
2.1.3 Strukturierte Daten und Volltextinformationen aus dem WWW	3
2.2 Visualisierung	4
2.2.1 Klimainformationen mit Echtzeit Video/Audio im WWW	4
2.2.2 Visualisierung von Klimadaten	5
3 Projektverlauf: Arbeiten und Ergebnisse	6
3.1 Offener Zugang zu Klimadaten und -informationen	6
3.1.1 Verfügbare Klima- und Global Change-Informationen	6
3.1.2 Datenbank-WWW Kopplung der CERA-Datenbank	7
3.1.3 Strukturierte Daten und Volltextinformationen mit Harvest	10
3.2 Visualisierung	16
3.2.1 Echtzeit Video/Audio im WWW	16
3.2.2 Visualisierung von Klimadaten	28
3.3 Präsentationen und Veröffentlichungen	33
4 Zusammenfassung	35

## Anhang

- A Akronyme
- B BLOB-Daten-Bereitstellung über WWW/AFTP
- C Filter und Regeln eines Gatherers
- D Konfigurationen eines Brokers
- E Struktur eines MPEG-Streams
- F Zuordnung von Farbpaletten und Variablen

# 1 Einführung

Sowohl das Informationsangebot als auch die Nachfrage nach relevanten Daten und Informationen im Internet (speziell im World Wide Web) nehmen immer mehr zu. Suchmaschinen wie Lycos, Infoseek und Yahoo, die weltweit die WWW-Server absuchen, sind häufig nicht die geeigneten Werkzeuge zum Auffinden der gewünschten Information. Sie berücksichtigen keine Fragestellungen, die über eine Stichwortsuche hinausgehen und liefern häufig Scheinergebnisse oder veraltete Links. Andererseits sind Anfragen, bezogen auf ein Fachgebiet, nur begrenzt möglich.

Mit zunehmender Informationsflut im Internet werden daher übergeordnete Instanzen wichtiger, die einen strukturierten Zugang zu bestimmten Themengebieten für eine definierte Zielgruppe anbieten. In diesem Kontext sollte im DKRZ ein Klimainformationssystem aufgebaut werden, das neben Informationsangeboten verschiedener Einrichtungen vor allem unterschiedliche Informationsstrukturen berücksichtigt. Zusätzlich sollte die Attraktivität des Angebotes durch multimediale Anwendungen wie echtzeitfähiges Video/Audio und Visualisierungskomponenten bereichert werden.

Der Bericht ist folgendermaßen gegliedert: Im Anschluß an diese Einführung werden in Kap. 2 das Umfeld, die Aufgaben und Ziele des Projektes vorgestellt.

In Kap. 3 werden der Projektverlauf, die durchgeführten Arbeiten und die im Projekt erzielten Ergebnisse dargelegt. Dabei haben sich im Laufe des Projektes neue Aspekte ergeben, die in die hier aufgeführte Aufgabenstellung eingeflossen sind.

Ein Beispiel ist die Untersuchung verschiedener Echtzeit Video/Audio-Systeme, die zu einer Auswahl von zwei unterschiedlichen Produkten geführt hat. Eine weitere Veränderung betrifft die Visualisierung von Klimadaten mit VRML. Eine im gegenseitigen Einvernehmen vereinbarte vorzeitige Beendigung des Projektes ließ hier nur Zeit für eine konzeptionelle Behandlung.

Abschließend liefert Kap. 4 eine Zusammenfassung über das im Projekt entstandene Klimainformationssystem.

Anhänge ergänzen den Bericht mit Detailinformation, wobei an dieser Stelle nur auf Anhang A „Akronyme“ hingewiesen werden soll, der eine Liste der im Bericht verwendeten Abkürzungen enthält.

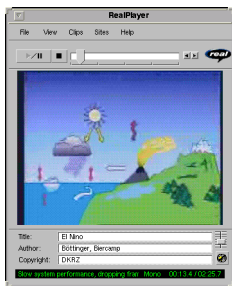
## 2 Projektbeschreibung: Aufgaben und Ziele

Ein wesentliches Ziel des Projektes war es, einen möglichst umfassenden Informationsdienst für den Klimabereich zu etablieren. Zielgruppe dieses Angebotes sollten neben Meteorologen und Klimaforschern vor allem „interessierte“ (naturwissenschaftlich gebildete) Laien sein, bei denen kein hohes fachspezifisches Vorwissen vorausgesetzt werden kann.

Das Projekt gliederte sich in mehrere Phasen. Am Anfang stand eine Bestandsaufnahme der verfügbaren Klimainformationsquellen. Ferner sollte im ersten Realisierungsschritt der Zugang zur Klimadatenbank am DKRZ umgesetzt werden.

Darauf folgte eine Phase der Konzepterstellung, in der Informationsangebote verschiedener Einrichtungen und vor allem unterschiedliche Informationsstrukturen berücksichtigt werden sollten. Hier sollte auch die Entscheidung für die einzusetzende Software, beispielsweise WAIS (Wide Area Information Servers) oder Harvest zur Suche und Erzeugung von Indizes, fallen.

Nach der Implementation des Klimainformationssystems sollten zunächst die am Projekt beteiligten Einrichtungen auf das Angebot zugreifen. In einer zweiten Stufe sollte das System optimiert und erweitert und der offene WWW-Zugang zu den Klimadaten über einen ausgewählten Pilotnutzerkreis getestet werden, bevor der Zugang allgemein angeboten werden sollte.



Ein weiterer Schwerpunkt lag in der Integration multimedialer Angebote zur Präsentation von Klimainformationen. Dies sollte mit Echtzeit Video/Audio realisiert werden. Dafür war die an der „University of Illinois at Urbana, Champaign“ entwickelte Client/Server-Software „Vosaic“ vorgesehen. Sie realisiert ein Echtzeit-Protokoll – Video Datagram Protocol (VDP) genannt – um Daten in „Echtzeit“ zu übertragen, Wartezeiten beim Transfer zu reduzieren, Verluste zu ignorieren und Abstände zwischen Frames bei der Darstellung zu minimieren.

### 2.1 Offener Zugang zu Klima- und Global Change-Informationen

#### 2.1.1 Bestandsaufnahme der verfügbaren Klimainformationen

Es war geplant, ausgehend vom vorhandenen Klimadatenkatalog des DKRZ WWW-Servers eine Bestandsaufnahme der verfügbaren Klimainformationsquellen vorzunehmen, um möglichst umfassend die Angebote in Deutschland berücksichtigen zu können. Dabei wurde in Betracht gezogen, das eventuell Absprachen mit den Anbietern zu treffen waren, falls Indizes direkt bei den Anbietern aufgebaut werden sollten.

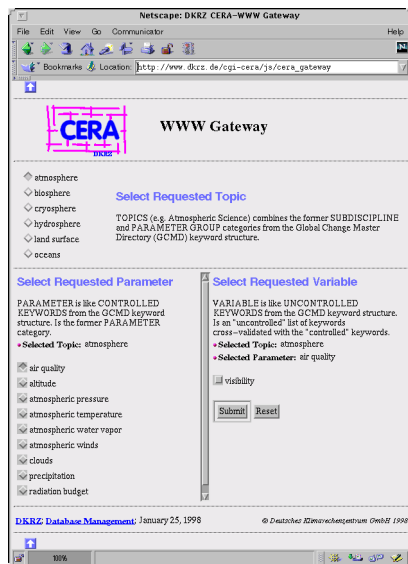
Der Klimadatenkatalog des DKRZ war und ist eine Zusammenfassung der Klima-orientierten Datenzentren (Deutschland, Europa, Welt) mit Links zu diesen WWW-Servern. Der Katalog wird manuell gepflegt, erfaßt natürlich nicht alle Anbieter und enthält eine grobe Kurzbeschreibung über die angebotenen Informationen der Ein-

richtungen. Es war Ziel des Projektes, diesen ersten Ansatz der globalen Klimainformationsbereitstellung durch einen strukturierten und komfortablen Zugang zu den gewünschten Informationen abzulösen.

## 2.1.2 Datenbank-WWW Kopplung

Im ersten Realisierungsschritt sollte der WWW-Zugang zur Klimadatenbank am DKRZ nutzerfreundlich gestaltet werden. Diese Klimadatenbank stellt den Zugang zu einer Datenbasis von Beobachtungs- und Klimamodelldaten dar und verwaltet einen Datenbestand von derzeit ca. 15 TeraByte.

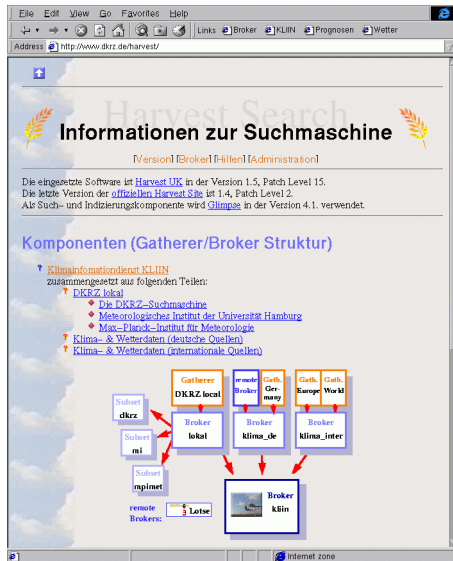
Es sollte ein „offener“ (aber kontrollierbarer) Zugang zum Lesen und Suchen in der Klimadatenbank mittels WWW geschaffen werden. Dabei waren verschiedene Navigationsmöglichkeiten (z.B. Search, Browse) vorgesehen und es sollten verschiedene Zugriffsverfahren auf die Datenbank untersucht werden, wie Datenbank-WWW-Kopplung über die CGI-Schnittstelle und auch andere Konzepte, die das dynamische Generieren von HTML-Dokumenten ermöglichen (z.B. Einsatz von Java).



Als Antworten auf die Anfragen werden von der Klimadatenbank neben Klimamodelldaten im Normalfall Metadaten zurückgeliefert, die Informationen über die eigentlichen Daten wie Name des Experimentes, Name des Datensatzes, Name des Datenzentrum, Name der verantwortlichen Person, Format usw. enthalten. Die Metadaten werden durch das CERA-Datenmodell (Climate and Environmental data Retrieval and Archive) festgelegt. Es sollte sich bei den bereitgestellten Daten aber nicht nur um Metadaten handeln, sondern es war auch vorgesehen eine Auswahl von vorprozessierten Daten, z.B. Einzelbilder und Videosequenzen anzubieten.

## 2.1.3 Strukturierte Daten und Volltextinformationen aus dem WWW

Es sollten die Informationsangebote verschiedener Einrichtungen und vor allem unterschiedliche Informationsstrukturen berücksichtigt werden. Dabei waren Daten im CERA-Format (z.B. beim DKRZ und AWI, umgesetzt in einer Datenbank), Informationen im sogenannten Data Interchange Format (DIF) wie auch unstrukturierte HTML-Dokumente in das Konzept einzubeziehen. Besonders bei den letztgenannten Informationen waren Verfahren zu entwickeln, die einen effizienten und schnellen Zugriff auf die Informationen ermöglichen. Hier lag der Einsatz von WAIS (Wide Area Information Servers) oder Harvest nahe.



Der anfragende Nutzer sollte mit einer möglichst einheitlichen Oberfläche dezentrale Informationen zum Klimabereich abfragen können. Das System sollte daraus Anfragen an Server generieren und absetzen. Weitere Komponenten des Systems sollten Aufschluß über die Anfragen in Form von Statistiken bzw. Logbüchern geben, aus denen dann ggf. neue Regeln oder Konfigurationen abgeleitet werden sollten.

## 2.2 Visualisierung

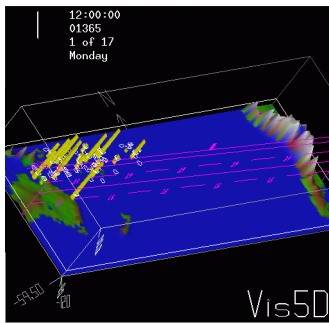
### 2.2.1 Klimainformationen mit Echtzeit Video/Audio im WWW

In diesem Projektbereich war geplant, Echtzeit Video/Audio-Präsentationen mithilfe der Client/Server-Software Vosaic (der Name entstand aus „Video Mosaic“), die ursprünglich an der „University of Illinois in Champaign-Urbana“ entwickelt worden ist, bereitzustellen. Im Gegensatz zu herkömmlichen Browsern, die Dokumente in ihrer Gesamtheit übertragen, um sie anschließend mittels „external applications“ anzuzeigen, beginnt der Vosaic-Client mit der Darstellung bereits beim Übertragen der Daten. Dazu wird das Real Time Protocol VDP verwendet, das in der Lage ist, Datenraten adaptiv an die verfügbare Bandbreite anzupassen, Verluste zu ignorieren und die Darstellung zu optimieren.



Zunächst sollte die Software installiert und getestet werden. In der ersten Stufe waren dafür Dokumente der Entwickler vorgesehen, anschließend sollten eigene Ausschnitte von vorhandenen DKRZ-Präsentationsfilmen über den A/V-Server bereitgestellt werden. Dazu waren Umsetzungen in die durch die Software unterstützten Formate erforderlich. Nutzer sollen gezielt auf Sequenzen dieser A/V-Dokumente zugreifen können, d.h. es waren Annotationen der Animationen zu erstellen.

## 2.2.2 Visualisierung von Klimadaten



Es sollten aktuelle Grafikformate (VRML) bzw. Visualisierungssoftware (Vis5D) im Informationsdienst unterstützt werden. Das WWW-Angebot sollte außerdem um Ergebnisse aus Simulationen am DKRZ (wiss. Animationen) und entsprechende Annotationen stufenweise ergänzt werden.

## 3 Projektverlauf: Arbeiten und Ergebnisse

Die Laufzeit des Projektes betrug 20 Monate. Wegen des vorzeitigen Ausstiegs der Projektmitarbeiterin, der im gegenseitigen Einvernehmen stattfand, wurden keine VRML-Darstellungen von Klimadaten erzeugt. Ebenso mußte auf Grund der gekürzten Projektlaufzeit der Pilotbetrieb mit einem größerem Nutzerkreis inklusive Bewertung reduziert werden. Alle anderen Teilgebiete konnten im vorgesehenen Umfang bearbeitet werden.

### 3.1 Offener Zugang zu Klimadaten und -informationen

#### 3.1.1 Verfügbare Klima- und Global Change-Informationen

In der ersten Phase des Projektes erfolgte zunächst eine Bestandsaufnahme vorhandener Klimainformationsangebote im WWW. Dazu gehören neben klimarelevanten Datenquellen und wichtigen Daten- und Metadatenformaten auch Visualisierungs- und Suchwerkzeuge. Einen Überblick über diese Thematik lieferte der „Earth Observation (EO)/GEO WWW-Workshop“ Anfang Februar '97 in Washington, auf dem das KLIIN-Projekt mit einem eigenen Vortrag vertreten war.

Auf dem Workshop trafen sich Entwickler Web-basierter Systeme aus den Bereichen „Earth Observation“ (EO), „Global Change“ (GC) und „Geographische Informationssysteme“ (GIS). Schwerpunktthemen waren Datenquellen, Kataloge und Formate, (geographische) Suchverfahren und Datenzugang sowie die Visualisierung von Daten. Anhand der vorgestellten Projekte konnte eine Einordnung und eine Abgrenzung des KLIIN-Projektes zu laufenden, zum Teil mit großem Personalaufwand und weltweiten Kooperationen durchgeführten Aktivitäten vorgenommen werden. Als großes internationales Projekt ist z.B. das EOSDIS/IMS (Earth Observing System Data and Information System/Information Management System) unter der Leitung der NASA (National Aeronautics and Space Administration) hervorzuheben, welches Kooperationspartner und Datenlieferanten in Kanada, USA, Europa und Australien vereint.

#### Metadaten

Zu den wichtigsten internationalen Entwicklungen von Metadaten-Standards aus dem „Global Change“- und „Earth Observation“-Bereich gehören das Directory Interchange Format (DIF) der Global Change Master Directory (GCMD), der „Content Standard for Digital Geospatial Metadata“ des Federal Geographic Data Committee (FGDC) und eine ISO-Norm (ISO/TC 211 Geographic Information/Geomatics). Das FGDC hat die zweite Version des Standards verabschiedet, dessen wichtige Neuerungen die Unterteilung in einen Kernbereich, Standardelemente und nutzerdefinierbare Metadatenelemente sind. Durch diese Neuerungen läßt sich der Standard einfacher auch für komplexe Datenbeschreibungen, die besonders bei der Archivierung von Meßdaten aber auch von Klimamodelldaten anfallen, anwenden.

Andere komplexe Formate/Protokolle (z.B. Z39.50), aus denen in der Regel definierte Untermengen (Profile, z.B. GILS – Government Information Locator Service, MARC – Machine-Readable Cataloging) zum Einsatz kommen, stammen ursprünglich aus dem Bibliothekenwesen.

Die wichtigsten Metadatenformate lassen sich aufeinander abbilden. Es gibt Werkzeuge zur Erstellung und Konvertierung von Formaten (z.B. Metadata Parser „mp“ von Peter Schweitzer).

Korrespondierend zur Entwicklung von Standard-Formaten sind internationale, über das WWW abfragbare Sammelstellen von Metainformationen entstanden. Ein Beispiel hierfür ist die GCMD, die mit der Archivierung von Metainformationen über Satellitenbeobachtungen begann und deren Struktur/Metadatenformat für die Beschreibung von Modell- und Meßdaten erweitert wurde. Im nordamerikanischen Raum wird ferner der „Government Information Locator Service“ als verteilte Datensammlung betrieben.

## Daten

Neben elektronischen Bibliotheken (z.B. ADL – Alexandria Digital Library, MEL – Master Environmental Library) gibt es im World Wide Web unter anderem Zugänge zu Satelliten- (z.B. ISIS – Interactive Satellite Image Server), Umwelt- (z.B. SFEP – South Florida Ecosystem Program, Datenbank mit Metadaten) und Klimadaten (z.B. NODC Distributed Databases – U.S. National Oceanographic Data Center). Diese Zugänge sind entweder über ein (eingeschränktes) Datenbankinterface zu Daten oder Metadaten, einen Katalog oder einen (verteilten) Index realisiert.

Ein sehr aufwendiges und komplexes Web-Gateway stellt das Earth Observing System Data and Information System (EOSDIS) in der Version 0 dar. In einer Kooperation zwischen weltweit verteilten Datenzentren (DAAC – Distributed Active Archive Center) werden durch Such- und Browsemechanismen deren (häufig frei verfügbare) Daten oder Datenquellen aus dem Bereich „globale Erdwissenschaften“ zugänglich gemacht. Dieses Gateway ist für die Zielgruppe „Wissenschaftler aus dem EO (Earth Observation) Bereich“ konzipiert, die beispielsweise mit verschiedenen Datenformaten oder Sensortypen vertraut sind.

Allgemein zugängliche Datenquellen sind vor allem in der USA beheimatet, der öffentliche Zugang zu den meteorologischen Daten im europäischen Raum ist viel stärker eingeschränkt. Am DKRZ wird mit dem Klimadatenkatalog (<http://www.dkrz.de/forschung/project/klimadatenkatalog.html>) eine Übersicht der relevanten Datenzugänge gepflegt.

### 3.1.2 Datenbank-WWW Kopplung der CERA-Datenbank

Bei der Anbindung von Datenbanken an das WWW und beim Design der Nutzeroberfläche stellt sich sofort die Frage nach der Zielgruppe und deren Erwartungen. Wir wollten sowohl Klimaforscher (Experten) wie auch interessierte Laien unterstützen. Dazu wurden zwei unterschiedliche WWW-Zugänge zur Klimadatenbank des DKRZ für unterschiedliche Anwendungen implementiert und bereitgestellt.

## Wetterdaten

Seit September '96 wird am DKRZ in Zusammenarbeit mit dem Meteorologischen Institut der Universität Hamburg eine Wetterstation betrieben. Die auf dem Dach gemessenen Daten werden „online“ im WWW veröffentlicht und in Dateien gespeichert.

Im Rahmen des KLIIN-Projektes wurde die Archivierung der Daten in der Klimadatenbank des DKRZ realisiert. Diese Daten werden nun täglich ASCII-formatiert in der Klimadatenbank des DKRZ gespeichert und dort für den Datenbankserver schnell zugänglich auf Festplatte vorgehalten.

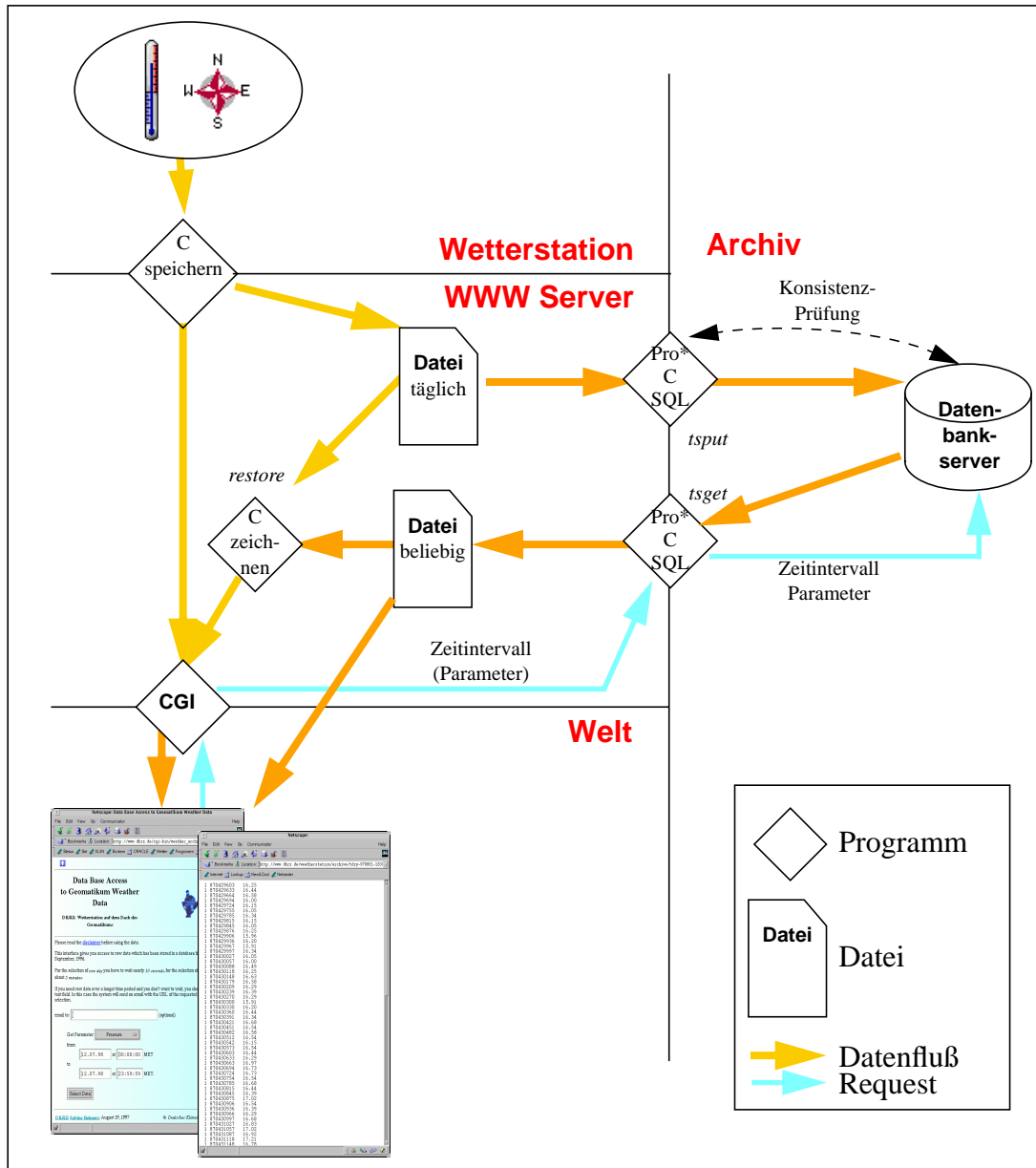


Abb. 1: Datenfluß der Wetterdaten, Archivierung und WWW-Gateway

Der WWW-Zugang zu diesen Daten ist prototypisch für die Zielgruppe „interessierte Laien“ implementiert ([http://www.dkrz.de/cgi-bin/weather\\_archive](http://www.dkrz.de/cgi-bin/weather_archive)). Die WWW-Abfrage ermöglicht eine Selektion nach einem Parameter (Temperatur, Windstärke, usw., oder alle Parametern gleichzeitig) in einem wählbaren Zeitintervall. Falls der Nutzer Daten über einen größeren Zeitraum ordert und die Bereitstellung dieser größeren Daten-

menge somit einen längeren Zeitraum beansprucht, hat er die Möglichkeit, sich über das Eintreffen der Daten auf dem WWW-Server durch eine E-Mail benachrichtigen zu lassen.

In Abb. 1 ist das Gesamtkonzept zur Speicherung und Bearbeitung der Daten dargestellt. Es sind zwei Datenflüsse realisiert. Neben der „quasi Echtzeit-Anzeige“ der Meßwerte werden Diagramme mit dem Tagesverlauf der Parameter aus den gespeicherten Dateien generiert und im WWW angeboten. Die in der Datenbank gespeicherten Meßwerte können entweder direkt als „Rohdaten“ ausgeliefert oder in einem weiteren Verarbeitungsschritt zu Diagrammen umgesetzt werden, die dann als GIF-Bilder zur Verfügung stehen.

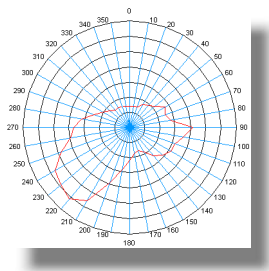


Abb. 2: Relative Häufigkeit der Windrichtung 1997

Die gute Nutzer-Akzeptanz dieser Schnittstelle zeigt sich beispielsweise dadurch, daß es gibt Nutzer gibt, die uns ihre aus den Daten generierten Auswertungen zur Verfügung stellen. Ein Beispiel ist die Darstellung der relativen Häufigkeit der Windrichtung 1997 (siehe Abb. 2).

## Klimamodelldaten

Im Gegensatz zu der oben beschriebenen Bereitstellung von Wetterdaten mittels Datenbank/WWW stellt das CERA WWW-Gateway (<http://www.dkrz.de/forschung/project/cerahome.html>) ein Interface dar, das Datenbankabfragen nach kompletten Datensätzen ermöglicht und vor allem von Wissenschaftlern aus dem Bereich der Klimaforschung genutzt wird. Für diese Nutzergruppe wurden Suchstrategien implementiert, die zur Selektion von Experimenten und Datensätzen führen. Bei der „Search“-Variante ist die Abfrage in Anlehnung an die Schlüsselwort-Hierarchie der GCMD aufgebaut. Die „Browse“-Variante ermöglicht ein Eingrenzen des Suchraums durch Auswahl von Schlüsselworten, geographischer Ausdehnung oder Projekten.

Die Metadaten eines Experimentes oder eines Datensatzes lassen sich mit dem Interface direkt im WWW abrufen. Die dazu gehörigen Klimadatensätze, die im GRIB-Format der WMO gespeichert sind, werden über das WWW-Interface angefordert und mit Unterstützung eines E-Mail Mechanismus im AFTP bereitgestellt.

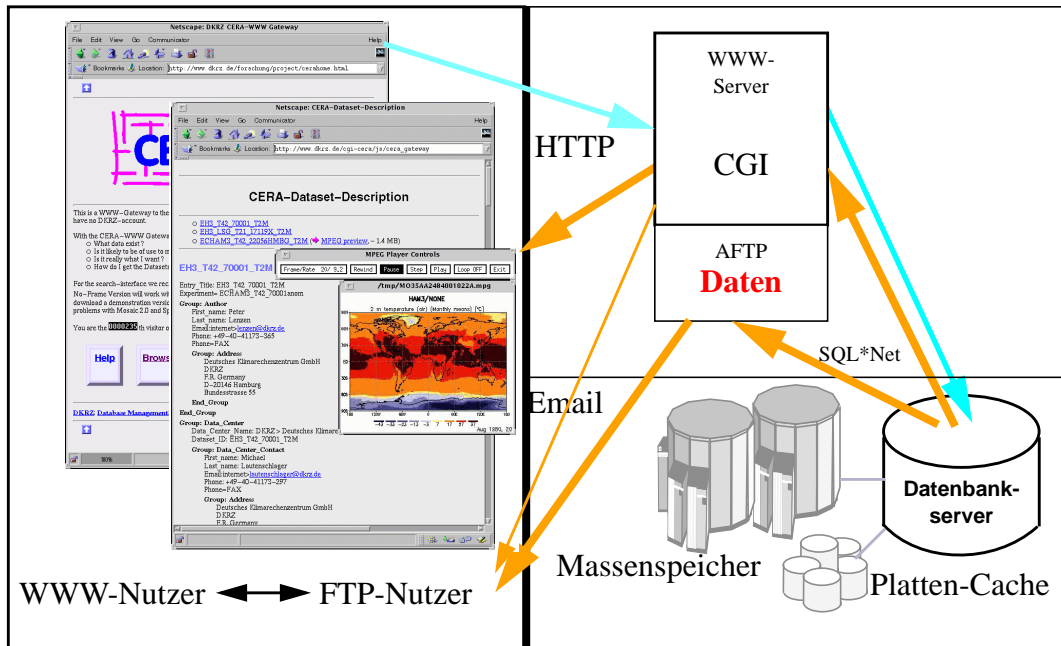


Abb. 3: Architektur des Gateways zur Klimadatenbank

Das hat seine Ursache in der Organisation der Daten innerhalb der Datenbank. Da die Datenmenge im TeraByte-Bereich liegt, können nicht alle Daten schnell verfügbar direkt auf dem Datenbankserver gespeichert werden. Ein Teil ist im Platten-Cache und ein weiterer im Tertiärspeicher (Cartridges) ausgelagert. Falls Daten angefordert werden, kann das je nach Auslastung der Systeme von Minuten bis zu Stunden dauern.

Um den Nutzer über den Stand seines Auftrags zu informieren, werden aus der Datenbank heraus E-Mails generiert (siehe Anhang B). Nachdem die Daten erfolgreich in den AFTP-Bereich transferiert wurden, wird ein Visualisierungsprogramm gestartet, welches einen Preview der Daten in Form eines MPEG-Films generiert (siehe „Automatische Generierung wissenschaftlicher Animationen mit GrADS“ auf Seite 31). Dieser Film wird „gecached“ und steht dann den nächsten anfragenden Nutzern sofort im WWW zur Verfügung.

Die DKRZ-Datenbank wird zur Zeit unter Oracle 8.0.3/Solaris 2.5 betrieben. Abb. 3 beschreibt die Komponenten der Schnittstelle. Die Verbindung zwischen Datenbank und WWW-Server wird über SQL\*Net (Oracles Netzwerkprotokoll) hergestellt. Pro\*C-Programme kommunizieren unter UNIX über diese Schnittstelle mit der Datenbank. Die Pro\*C-Programme werden per CGI vom WWW-Server aufgerufen.

### 3.1.3 Strukturierte Daten und Volltextinformationen mit Harvest

Zu den Vorteilen der Suche in strukturierten Metadaten zählen die – vor allem für Experten – mächtigen Möglichkeiten zur Eingrenzung eines Problems und der Abschluß von „Scheinergebnissen“. Im Gegensatz dazu bietet eine Suchmaschine auf Basis von Volltextindizierung Vorteile wie eine einfache nutzerseitige Bedienung, die schnelle Antworten ermöglicht. Es wird keine zeitverzögernde Kooperation und Nor-

mung benötigt und dennoch ein breites Spektrum an Informationen zugänglich gemacht. Für die Erschließung klimarelevanter HTML-Seiten im WWW war eine Volltextindizierung nötig.

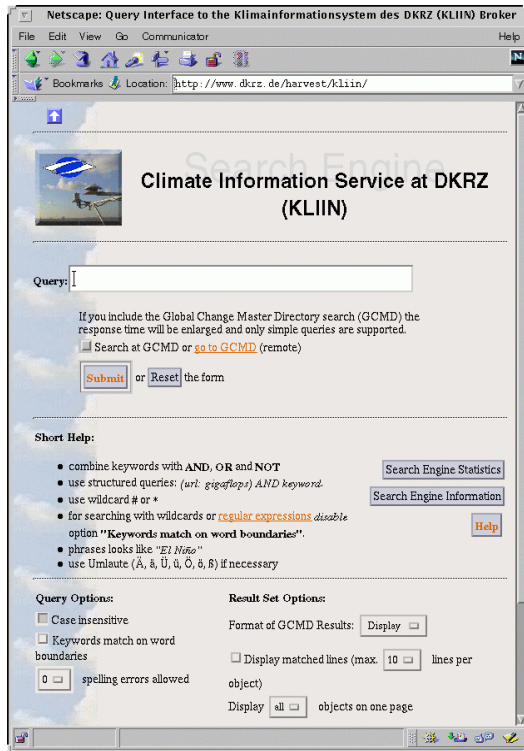


Abb. 4: WWW-Interface zum Harvest-Suchsystem

„Glimpse“ mit der aktuellen Version 4.1 wird noch von der Gruppe weiterentwickelt, die das Projekt etabliert hat. Die Software besteht aus einem Paket von Modulen, die unabhängig voneinander konfigurierbar sind und in verschiedenen Kombinationen zusammengefügt werden können. Die Intention von Harvest besteht darin, gezielt eine endliche Menge von Dokumenten unter einem bestimmten Gesichtspunkt bereitzustellen und nicht einen „Robot“ zu etablieren, der das gesamte Web indiziert.

## Struktur

Die Komponenten der Harvest-Software lassen sich hierarchisch kombinieren und erlauben in jeder Stufe den Einsatz von Filtern. Die Hierarchie der Harvest-Komponenten in dem von uns aufgebauten Suchsystem ist in Abb. 5 dargestellt.

Wir haben vier Gatherer konfiguriert, die lokal bzw. auf entfernten Servern Dokumente durchsuchen und diese als Objekte im „Harvest Summary Object Interchange Format“ (SOIF) speichern. Dabei wird der Inhalt eines Dokumentes Attributen zugeordnet, die beispielsweise bei HTML-Seiten aus den HTML-Tags und den Bezeichnern der Meta-Tags generiert werden. Es ist außerdem möglich, einem Objekt Attribute per Hand (oder per Programm) zuzufügen. Gatherer sind jedoch nicht nur in

Wir setzen für diese Aufgabe die Harvest-Software ein. Die Software zeichnet sich durch architektonischen Konzepte wie eine modulare Architektur bestehend aus Indexer, Gatherer und Broker und die Möglichkeit der Integration verschiedener Metadatenformate und SGML-Strukturen aus. Die Dateistruktur der Originaldokumente bleibt als Attributwert der URL zugänglich und auswertbar und es können eigene Attribute, die beispielsweise räumliche (geographische) oder zeitliche Informationen berücksichtigen, integriert werden.

Die Harvest-Software wurde in einem inzwischen beendeten Projekt an der „University of Colorado at Boulder“ entwickelt. Die zur Zeit aktuelle Version 1.5 wird von einer Nutzergruppe aus England betreut, nur die Indizierungs- und Suchsoftware

der Lage HTML-Dokumente zu extrahieren, sondern sie können auch Postscript-, PDF- und Textdateien, sowie Tar-Archive und andere Formate durchsuchen und auf Attribute abbilden.

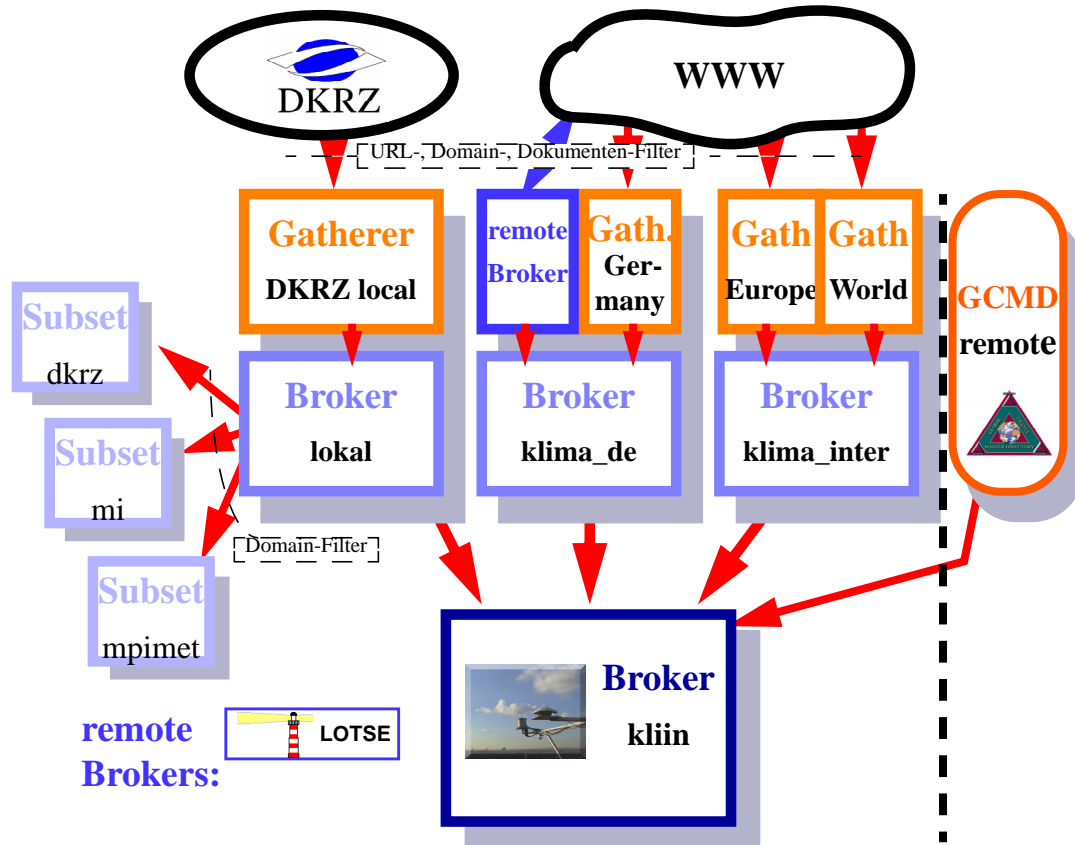


Abb. 5: Datenfluß und Filter zwischen den Harvest-Modulen des Klimainformationssystems

Drei Broker übernehmen die Objekte, überprüfen dabei die Lebensdauer und löschen gegebenenfalls mehrfach auftretende Objekte. In Kooperation mit dem GKSS-Forschungszentrum wurde der dort im Projekt LOTSE (Land/Ocean Thematic Search Engine) laufende Broker ebenfalls integriert. Ein vierter Broker vereinigt die Informationen der drei Datenquellen und macht sie über das WWW-Interface <http://www.dkrz.de/harvest/kliin/> verfügbar (siehe Abb. 4). In dieses Interface wurde eine Abfragemöglichkeit der GCMD integriert, die auf Wunsch des Nutzers aktiviert werden kann.

Jeder Broker besitzt eine Such- und Indizierungskomponente, die unter Verwendung von CGI-Scripten die Schnittstelle zum Anwender darstellt. Wir setzen dafür die im Harvest-Projekt entwickelte Software „Glimpse“ ein, es ist jedoch auch möglich, andere Indexer, wie z.B. „wais“ oder „swish“ zu verwenden. Die Syntax der Anfragen und das WWW-Interface werden durch diese Programmkomponente bestimmt.

Lokal beim DKRZ werden WWW-Server verschiedener Institute betrieben, die von einem Gatherer/Broker-Paar verfügbar gemacht werden. Die Schnittstellen zum

DKRZ, dem Meteorologischen Institut der Universität Hamburg und zum Max-Planck-Institut für Meteorologie bilden die Suchsysteme für die Institute und sind durch einfache Formulare realisiert. Damit wird auf der gesamten Datenbasis des Brokers „lokal“ gesucht. Die Nutzeranfrage wird durch einen mit „AND“ verknüpften Term, der die WWW-Server-Domain selektiert, erweitert.

### Komponenten: Gatherer

Prinzipiell läßt sich ein Gatherer um so spezifischer konfigurieren, je mehr Wissen über die Datei-Struktur des Servers und über die Dokumente einfließt, die eingesammelt und als Objekte gespeichert werden sollen. Es ist möglich, eine Auswahl bezüglich der Server, bezüglich der URL's oder der Dokumententypen zu treffen. In Listing 1 ist exemplarisch die Konfigurationsdatei des Gatherers „lokal“ abgebildet, um einen Eindruck über mögliche Spezifikationen zu vermitteln. Mit Hilfe von URL- und Domainfiltern kann man beispielsweise den gesamten Inhalt eines Servers, bestimmte Dokumententypen (z.B. Audio-Dokumente mit der Endung .au) oder URL's mit zu definierenden Pfadkomponenten (z.B. /statistik) vom Einsammeln oder der Objektbildung ausschließen. Listings von Filtern und Regeln sind exemplarisch in Anhang C aufgeführt.

Außerdem stehen Werkzeuge zur Verfügung, mit denen die erzeugten Objekte editiert werden können. Es ist jedoch nicht möglich, anhand einer Liste von Stichworten zu entscheiden, ob ein Dokument als Objekt aufgenommen wird oder nicht.

```
#
# lokal.cf - configuration file for the
# Deutsches Klimarechenzentrum GmbH Gatherer
#
# Created by reimers@dkrz.de on Tue Nov  4 13:24:43 MET 1997
#
#-----
#-----konfiguration-----
#-----
Gatherer-Name:   Deutsches Klimarechenzentrum GmbH
Gatherer-Port:   8510
Top-Directory:  /var/opt/harvest/gatherers/lokal
Lib-Directory:  /var/opt/harvest/gatherers/lokal/lib

#Refresh-Rate; 604800: 1 week, Time-To-Live: 2419200: 1 month (default)
#Refresh-Rate; 86400: 1 day, Time-To-Live: 604800: 1 week
Refresh-Rate:    86400
Time-To-Live:    604800
#-----
#Post-Summarizing fuer file:...-URLs
Post-Summarizing: lib/urllrules
#-----
#Local-Mapping fuer http:...-URLs
#Local-Mapping:  http://www.dkrz.de/ /www/
#Local-Mapping:  http://www.mi.uni-hamburg.de/ /afs/dkrz.de/www/mi/
#Local-Mapping:  http://www.mpimet.mpg.de/ /afs/dkrz.de/www/mpi/
#Debug-Options:  -D20,1

#-----
#-----RootNodes-----
#-----
<RootNodes>
# Enter URLs for RootNodes here
file://info.dkrz.de/www/klima/elnino/ URL=30,/pf/k/k202039/harvest/gatherers/lokal/lib/
urlfilter.file
file://info.dkrz.de/www/ URL=9000,/pf/k/k202039/harvest/gatherers/lokal/lib/urlfilter.file
file://info.dkrz.de/afs/dkrz.de/www/mi/ URL=750,/pf/k/k202039/harvest/gatherers/lokal/lib/
urlfilter.file
file://info.dkrz.de/afs/dkrz.de/www/mpi/ URL=750,/pf/k/k202039/harvest/gatherers/lokal/lib/
urlfilter.file
file://info.dkrz.de/pf/k/k202045/html/tnp/ URL=100,lib/urlfilter.file
file://info.dkrz.de/ipcc/ddc/ URL=200,lib/urlfilter.file
```

```

</RootNodes>

#-----
#-----LeafNodes-----
#-----
<LeafNodes>
# Enter URLs for LeafNodes here
</LeafNodes>
    
```

### Listing 1: Beispiel einer Gatherer-Konfiguration

Zum Erfassen möglichst umfangreicher Klimainformationen sind vier Gatherer aktiv. Der Gatherer „lokal“ durchsucht die physikalisch am DKRZ lokalisierten WWW-Server. Dazu gehören der WWW-Server des DKRZ, der Server des Meteorologischen Instituts der Universität Hamburg und der Server des Max-Planck-Instituts für Meteorologie. Da dieser Gatherer keine TCP-Verbindungen aufbauen muß, kann die Datenerfassung besonders effizient erfolgen. Mit dem „Post-Summarizing“-Mechanismus werden die Dokumente des lokalen Filesystems in URL's übersetzt.

Die drei anderen Gatherer „klima\_de“, „klima\_eu“ und „klima\_inter“ durchsuchen – ausgehend von den WWW-Seiten des Klimadatenkatalogs und anderen bei uns lokalisierten WWW-Seiten mit Klima- und Wetter-Links – entfernte Server. Eine Aufspaltung in mehrere Gatherer, wobei einer die Domain „de“, eine zweiter die Domains europäischer Länder (z.B. „uk“, „nl“) und der dritte alle anderen Domains einsammelt, wurde aus zwei Gründen vorgenommen: Das Durchsuchen der klimarelevanten WWW-Server in Deutschland benötigt erheblich weniger Zeit und Ressourcen und kann deshalb öfter durchgeführt werden als ein Sammellauf auf aus Netzsicht weiter entfernten Servern. Aber auch aus inhaltlicher Sicht ist eine Aufteilung sinnvoll, denn die Trefferquote deutschsprachiger Texte wird durch Anfrage an den entsprechenden Broker erhöht.

Abhängig von der Häufigkeit eines Gatherer-Laufes, die mit Hilfe des Unix-„Cron“-Mechanismus festgelegt wird, müssen die Attribute „Refresh-Rate“ und „Time-To-Live“ in der Konfiguration des Gatherers definiert werden. Bei jedem Lauf werden die Dokumente mit einer „Update-Time“ versehen. Der Broker löscht alle Objekte, die älter als „Update-Time + Time-To-Live“ sind. Das Attribut „Refresh-Rate“ sorgt dafür, das die „Update-Time“ regelmäßig aktualisiert wird.

Zur Zeit wird der lokal arbeitende Gatherer zweimal in der Woche gestartet, der Gatherer für die Domain „de“ läuft einmal im Monat, während der europäische und der international suchende Gatherer nur alle zwei Monate aktiviert werden. Das hat seine Ursache in dem Zeitaufwand eines Sammellaufes, der lokal durchgeführt nur knapp einen halben Tag, auf deutschen Klimainformationen bereitstellenden Servern ca. eine Woche und für die internationalen Suchläufe ca. zwei Wochen dauert.

Das Erzeugen einer Gatherer-Datenbank erfolgt erst, nachdem alle durch die Konfiguration erreichbaren Links durchlaufen worden sind. Wenn der Rechner, auf dem der Gatherer läuft, vorher heruntergefahren wird, sind alle bis dahin gesammelten Informationen wertlos. Das kann vor allem dann zu einem Problem werden, wenn ein Gatherer mehrere Tage mit einem Lauf beschäftigt ist.

Um dieses Problem zu entschärfen, wurden für die Gatherer mehrere Konfigurationen erstellt, die sich jeweils durch die URL unterscheiden, bei der der Sammellauf gestar-

tet wird. Mit jeder Konfiguration wird ein Sammellauf initiiert, die Daten vereinen sich jedoch in einer Datenbasis.

Diese Vorgehen hat auch den Vorteil, daß damit einerseits die Gesamtzeit eines Gathererlaufes reduziert wurde und daß die Datenbasis inkrementel erneuert und damit eher aktualisiert wird. Es wurde ferner ein Mechanismus erstellt, der auch dann automatisch erkennt, mit welcher Konfiguration der nächste Sammellauf gestartet wird, falls beispielsweise wegen eines Rechnerausfalls der Prozeß von Hand angestoßen werden muß.



Abb. 6: Ergebnis einer Suche

bereit. Abb. 6 zeigt das Ergebnis einer Suche nach dem Begriff „El Niño“ auf deutschen Servern.

Wir setzen vier Broker ein, die es ermöglichen, entweder in den beim DKRZ lokalisierten Servern, auf deutschen Servern, auf internationalen Servern oder in der gesamten Datenbasis zu suchen. Exemplarische Brokerkonfigurationen befinden sich in Anhang D.

Das Suchinterface der Broker wurde durch weitere Funktionalität ergänzt. Dazu gehört ein Sortieren der Ergebnisse anhand der Anzahl der Zeilen, die das Suchwort enthalten, gewichtet mit einem Wert, der für jedes Attribut (z.B. Body, Title) festgelegt wird. Außerdem ist es möglich, die Ergebnisliste aufgeteilt auf mehreren Seiten

## Komponenten: Broker

Broker erhalten Objekte entweder von Gatherern oder von anderen Brokern. Diese Übernahme kann unter Verwendung eines Booleschen Ausdrucks gefiltert werden. Allerdings schien uns die Software bei der Verwendung dieser Möglichkeit nicht voll funktionsfähig zu sein, da die so gewonnene Datenbasis eines Brokers mit der Zeit alle Objekte als veraltet löscht und sie nicht erneuert. Daher haben wir diese Art der Filterung wieder eingestellt.

Aufgaben eines Brokers sind die regelmäßige Abfrage der Datenquelle(n), die Überprüfung der Gültigkeit von Objekten, das Löschen von doppelt auftretenden Objekten und die Indizierung der Daten. Außerdem stellen Broker sowohl ein WWW-Interface zur Suche als auch eins zur Administration

ausgeben zu lassen. In das Interface, welches die Informationen aller Datenquellen zusammenführt, wurde eine Anbindung an die GCMD integriert. Allerdings ist die Netzverbindung zur GCMD nicht sehr optimal, so daß durch ein zusätzliches Abfragen dieser Datenbasis die Antwortzeit des Systems und damit die Wartezeit des Nutzers steigt.

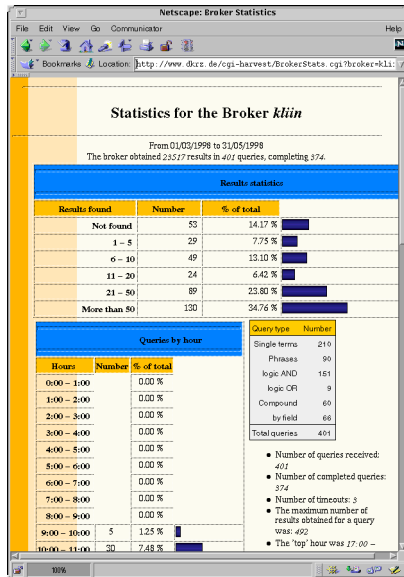


Abb. 6: Statistik eines Brokers

Jede Aktion eines Brokers wird protokolliert. Zusätzlich haben wir ein Programm installiert, das eine Statistik der Anfragen grafisch darstellt und über die Komplexität der Anfragen sowie die Antwortzeiten des Systems Auskunft gibt. Diese Statistik ist jedoch nur für DKRZ Mitarbeiter zugänglich.

## Probleme

Zur korrekten Behandlung von Umlauten mußte eine Schwierigkeit bei der Installation überwunden werden. Der benötigte Zeichensatz „iso\_8859\_1“ kann mit der Definition der Environment-Variablen „LC\_CTYPE“ eingebunden werden. Die Definition mußte jedoch nicht nur in den Scripten zum Starten der Harvest-Komponenten eingefügt werden, sondern abhängig vom Betriebssystem auch in einigen Makefiles vor dem Kompilieren von Harvest ergänzt werden. Dies war nicht dokumentiert.

Ferner tritt beim Ausschalten der Option „Display matched lines“ bei der Suche über das WWW-Interface eine fehlerhafte Interpretation der Anfrage auf, so daß das Suchergebnis verfälscht wird.

Ein Problem tritt bei der Speicherung der SOIF-Objekte auf, falls eine WWW-Seite fehlerhaftes HTML beinhaltet. Die Übernahme der Informationen in das Objekt bricht an einer solchen Stelle ab. Als fehlerhaftes HTML wird beispielsweise auch die Verwendung von Listenelementen erkannt, die nicht in die entsprechende Listenumgebung eingebettet wurden.

## 3.2 Visualisierung

### 3.2.1 Echtzeit Video/Audio im WWW

In diesem Projektbereich sollten qualitativ hochwertige Präsentationsfilme aus dem Klimabereich angeboten werden. Hierzu fanden zunächst vergleichende Untersuchungen von Audio/Video-Systemen statt, die über proprietäre Protokolle Daten übertragen und quasi gleichzeitig (im „streaming“-Verfahren) darstellen können. Da die im Projektantrag vorgesehene Software „Vosaic“ inzwischen in den kommerziellen Bereich übergewechselt war, war es sinnvoll, zunächst einen Vergleich von inzwischen verfügbaren, ähnlichen Produkten vorzunehmen. Aufgrund einer Kooperation mit den Entwicklern stand uns zwar eine kostenlose Server- und Konvertierungssoft-

ware einer alten Version von Vosaic zur Verfügung, neue Versionen mußten aber bezahlt werden. Andere Systeme, wie beispielsweise RealAudio/Video, konnten über einen befristeten Zeitraum in einer Testversion untersucht werden. Die Audio/Video-Clients aller Firmen sind frei verfügbar.

Getestet wurde auf einem PC (Pentium 200Pro) unter WIN NT 4.0 Server mit der WWW-Server Software Microsoft IIS2.0. Auf der Client-Seite wurden der WWW-Browser Netscape Navigator auf verschiedenen Plattformen sowie der Microsoft Internet Explorer verwendet.

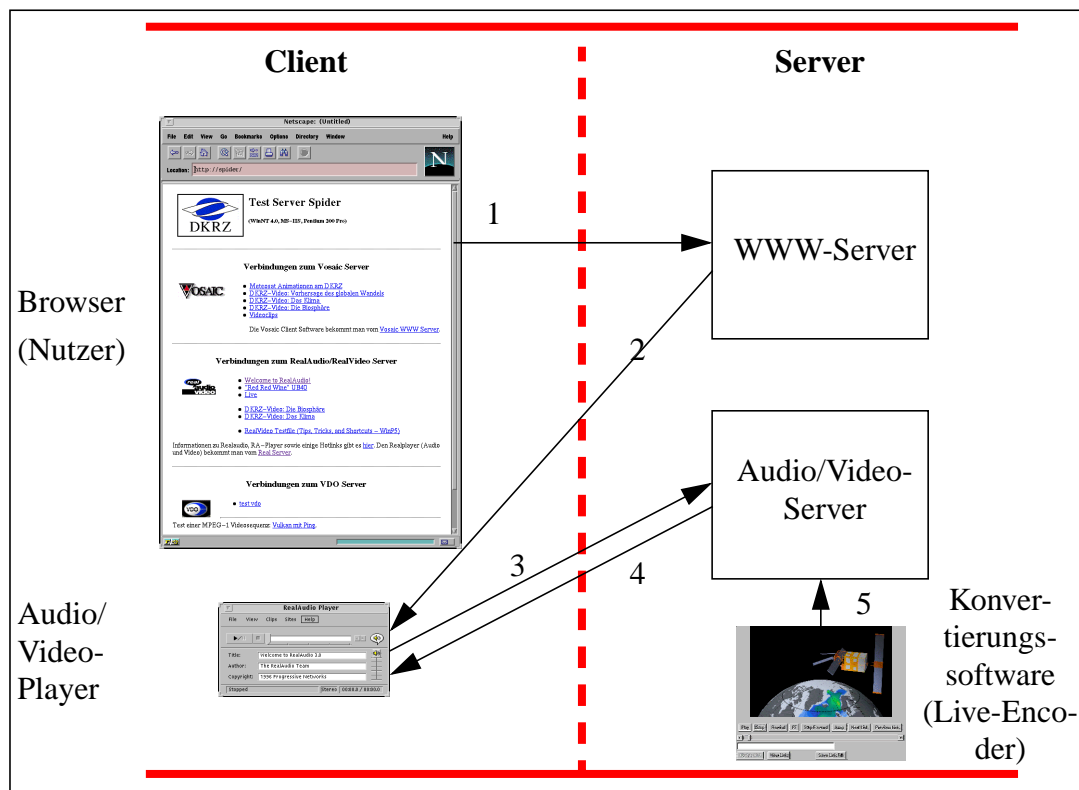


Abb. 7: Prinzipielles Zusammenspiel der Audio/Video-Software im WWW

Die Audio/Video-Software gliedert sich in die drei Bereiche Serversoftware, Clientsoftware und Werkzeuge zur Konvertierung und Bearbeitung der Video- und Audiosequenzen. Der Ablauf der Kommunikation zwischen den Komponenten ist in Abb. 7 dargestellt und verläuft folgendermaßen:

1. Der Benutzer wählt einen Audio/Video-Link in einem WWW-Dokument. Daraufhin wird der Audio/Video-Player lokal gestartet.
2. Der A/V-Player erhält Informationen in Form einer URL, die es ihm ermöglichen, sich mit dem A/V-Server in Verbindung zu setzen.
3. Der Player fordert die A/V-Sequenz vom Server an.
4. Der Server sendet die gewünschten Daten zum Player, der sie dekomprimiert und abspielt. Die Wiedergabe findet „parallel“ (im „streaming“-Verfahren) zur

Übertragung statt.

- Die A/V-Sequenzen können entweder vorher konvertiert oder aber live (in Echtzeit) bereitgestellt werden.

### Untersuchung von echtzeitfähigen Server/Client-Systemen

Tabelle 1 liefert einen Vergleich der berücksichtigten Systeme. Neben technischen Voraussetzungen wie unterstützte Plattformen oder Eingabeformate sind auch Leistungsmerkmale wie Multicastfähigkeit, Bandbreitenadaption oder Echtzeitübertragungsfähigkeit aufgeführt. Es sei an dieser Stelle darauf hingewiesen, daß die Untersuchung bereits Anfang 1997 stattfand und sich Leistungsmerkmale und Zustand der Software verändert haben können.

Tabelle 1: Vergleich von WWW-fähiger Audio/Video-Software im zweiten Quartal 1997

Produkt	Vosaic	RealVideo	Xing	VDOLife	VCR
Serverplattformen	Win95, WinNT, Mac, Unix	WinNT, FreeBDS, Linux, Solaris2.5	Irix, Solaris, Linux, WinNT	Win95, WinNT, FreeBDS, Linux, Solaris	SunOS HP-UX Solaris SGI IRIX FreeBSD LINUX
Clientplattformen	Win95, WinNT, Mac, Unix	Win3.1, Win95, WinNT, Mac, Unix (nur Audio)	Win95, WinNT, Mac, Unix	Win3.1, WinNT, Win95, Mac	UNIX (s.o.)
Toolplattformen	Win95, WinNT mit JDK 1.1	Win95, WinNT, Mac	Irix, Solaris, WinNT	WinNT, Win95, Mac	keine Tools
Eingabeformate	Video: AVI, AVS, MPEG, WAV, SunAU, Quicktime → MPEG Audio: +AIFF → GMS	AVI, Quicktime	AVI, MPEG	AVI, AVS	MEPG1, AU (Sun- Audio)
Übertragungs- formate	MPEG (Video) GMS (Audio)	eigene Kodierung	MPEG, LBR- Audio-Sreams	eigene Kodierung	MPEG (Video) AU(Audio)
Protokoll, CODEC	Stream, VDP, 2 Kanäle	RV-Server- stream	Stream, multi- castfähig	Stream, VTP	Stream, 2 Kanäle
Bandbreiten- adaption	4 Stufen	3 Stufen	ja	nein	nein
Echtzeit- übertragung	nein	ja	ja	ja	nein

Tabelle 1: Vergleich von WWW-fähiger Audio/Video-Software im zweiten Quartal 1997

Produkt	Vosaic	RealVideo	Xing	VDOLife	VCR
Einbindung der Daten in HTML (Bsp.)	<code>&lt;embed src="file.vos" ...&gt; start&lt;/embed&gt;</code> d.h. Verwendung als Inline-Video	<code>&lt;A HREF="text.ram"&gt; start&lt;/A&gt;</code> in text.ram: pnm://server/file.ra (oder .mov)	<code>&lt;A HREF="text.xdm"&gt; start&lt;/A&gt;</code> in text.xdm: STREAM = XDMA://server/file.ply (oder: .mpg)	<code>&lt;A HREF="text.vdo"&gt; start&lt;/A&gt;</code> in text.vdo: vdo://server/file.avi oder <code>&lt;embed src="file.avs" ...&gt; start&lt;/embed&gt;</code> (inline oder extern)	<code>&lt;A HREF="text.prog"&gt; start&lt;/A&gt;</code> in text.prog: Titel server: file.mpg server: file.au
Verfügbarkeit der Serversoftware	<a href="http://www.vosaic.de/">http://www.vosaic.de/</a> Bereitstellung der Software aufgrund einer Kooperation	<a href="http://www.realaudio.com/">http://www.realaudio.com/</a> 30-Tage Testversion	<a href="http://www.xingtech.com/">http://www.xingtech.com/</a> keine freie Testversion verfügbar	<a href="http://www.vdo.net/">http://www.vdo.net/</a> 30-Tage Testversion	<a href="http://cse.ogi.edu/~scen/Player/">http://cse.ogi.edu/~scen/Player/</a> frei verfügbar
getestete Software	MediaServer, MediaClient, Mediastudio	Server, Encoder und Tools auf WinNT, Clients auf Win95, WinNT	nur Clientsoftware (StreamWorksPlayer)	nur VDOLife-Client	modifizierter Server, Client
Bemerkungen	zusätzlich zur Multimedia-Datei muß eine Informationsdatei übertragen werden	demnächst sollen auch Video-Clients unter Unix verfügbar sein, Nutzer muß seine Bandbreite kennen	sowohl Server-, Client- als auch Toolplattform unter Solaris lauffähig	wird bei der WWW-Tagesschau eingesetzt	eingeschränkter Einsatzbereich (UNIX)
Bewertung	Es treten Probleme sowohl bei der Konvertierung, als auch bei der Übertragung auf	Der Audio-Player unter Solaris läuft instabil, gute Ergebnisse im PC-Umfeld	Vor allem vormittags (US-Leitung ist noch nicht überlastet) erstaunliche Übertragungsqualität	Die Server-Testversion war unter WinNT nicht lauffähig, gute Übertragungsqualität der Tagesschaubeispiele	sehr einfacher Server, keine Serverroot konfigurierbar (Quellcode vorhanden, änderbar), Synchronisationsprobleme

Vosaic hinterließ den Eindruck von Beta-Software. Die mit dem mitgelieferten Konvertierungswerkzeug erstellten MPEG-Dateien wurden zum Teil stark verändert, die Synchronisation von Ton und Bild funktionierte nur mangelhaft. Da die mitgelieferten Beispiele eine bessere Qualität aufweisen, stellte sich die Frage, welches Quellformat sich tatsächlich zur Konvertierung in das benötigte MPEG-Format eignet. Beispiele auf fremden Servern ließen sich nur teilweise und sehr bruchstückhaft übertragen.

Die Übertragung erfolgt im Unicast-Modus, damit hat der Nutzer die Möglichkeit, die Übertragung zu stoppen und „zurückzuspulen“. Dem Vosaic-Player fehlt jedoch eine Regelmöglichkeit der Lautstärke und eine Colormap-Anpassung, so daß die Ausgabe in Falschfarben-Darstellung erfolgen kann und sich dadurch zu nicht erkennbarem Geflimmer verändert.

*RealAudio/Video* erzielte recht gute Übertragungsergebnisse. Es wurde die „30-Tage-Testversion“ des A/V-Servers unter WIN NT 4.0 installiert. Auf der Serverseite konnten mit Hilfe der Konvertierungswerkzeuge Audio/Video-Dokumente für verschiedene Bandbreiten erzeugt werden, zwischen denen der Nutzer dann je nach persönlicher Anbindung (28.8 kBit, 64 kBit, usw.) wählen kann. Der Server dient lediglich dazu, die netzweite Übertragung zu regeln, d.h. RealAudio/Video-Dateien können lokal auch ohne Server abgespielt werden. Die Verwendung im Unix-Umfeld war nur sehr eingeschränkt möglich. Ein Video-Client ist hierfür erst in letzter Zeit und nur für Solaris erhältlich. Er läßt sich nur dann aufrufen, wenn noch genügend Farben vorhanden sind, sonst erzeugt er nur eine Fehlermeldung.

Die *Xing*-Software war vor allem deshalb interessant, weil sowohl Server als auch Client unter Unix lauffähig sind und daher im Wissenschaftsbereich leicht eingesetzt werden können. Zur Übertragung werden verschiedene MPEG-Ströme unterstützt. Da es keine kostenlose Testversion des Servers gab, konnte der Test jedoch nur eingeschränkt (nur die Client-Seite) vorgenommen werden. Der Client hat keine Lautstärkeregelung, d.h. es mußte zusätzlich ein Audiotool gestartet werden. Reine Audiodokumente wiesen eine deutlich bessere Qualität auf als Audio/Video-Dokumente. Bei geringerer Bandbreite wurde der Audiostrom bevorzugt übertragen, während die Videosequenz in eine Abfolge von Standbildern übergeht. Der Xing-Client (StreamWorksPlayer) und Vosaic behinderten sich gegenseitig bei der Freigabe des Audiodevices.

Der vom *VDO*-Softwarehersteller angebotene Testserver war auf der WinNT-Plattform nicht lauffähig. Daher wurde auch hier nur die Clientsoftware am Beispiel der Tagesschau untersucht, die einen VDO-Server zum Zeitpunkt des Tests einsetzte. Die dort bereitgestellten Dokumente ließen sich in einer akzeptablen Qualität übertragen. Inzwischen wird auch bei der Tagesschau die RealNetworks-Software verwendet.

*VCR* ist ein A/V-System aus der UNIX-Welt und bot eine geringe Funktionalität: Die Software beinhaltete nur einen Viewer und einen nicht konfigurierbaren Server. Werkzeuge zur Erstellung der Videos wurden nicht mitgeliefert oder vorgeschlagen. Da der Quellcode verfügbar ist, wurde der Server von uns soweit modifiziert, daß die Konfiguration einer „Serverroot“ möglich ist. Damit wurden die Lese- und Schreibrechte auf definierte Verzeichnisebenen eingeschränkt. Ein Problem trat bei der Synchronisation von Video und Audio auf, die als getrennte Ströme übertragen werden. Je geringer die Framerate des MPEG-Videos bei der Erstellung gewählt wird, desto eher ist der Film beendet, unabhängig von der Laufzeit des Audio-Clips.

Exemplarisch für den PC-Bereich wurde *NetShow* vom Microsoft untersucht. Das Produkt beruht auf der ActiveX-Technologie, ist nur unter Win95 einsatzfähig und wurde daher nicht in den tabellarischen Vergleich aufgenommen. Eine Beta-Version des Servers war frei verfügbar. Die Qualität der lokal abgerufenen Multimedia-Dokumente überzeugte bei NetShow. Hierzu wurde kein Server benötigt. Externe vom Ser-

ver bereitgestellte Dokumente wurden mit einem klaren, gut verständlichen Ton wiedergegeben, während der Videoanteil aus einer Abfolge von Standbildern bestand.

### **Eingesetzte Server-Software**

Im Rahmen des Projektes wurde der Vosaic-MediaServer in der Version 1.05 für Solaris 2.5 beschafft. Dieser Server kann sechs Nutzer gleichzeitig mit „Streams“ versorgen. Mitgeliefert wurde außerdem das Konvertierungswerkzeug Vosaic-MediaStudio in der Version 2.0b5, welches unter Windows NT 4.0 betrieben und über eine grafische Oberfläche bedient wird. Die Wurzeln des MediaStudios liegen (bei der Verarbeitung von MPEG1-Dateien) in der Unix-Welt bei den „Berkely MPEG Tools“, einem frei verfügbaren Archiv zur Erzeugung, Analyse und Darstellung von MPEG-Filmen. Zur automatischen Erzeugung von aktuellen „Streaming“-Satellitenfilmen ([http://www.dkrz.de/svideo/vosaic/sat/d3\\_64\\_5.html](http://www.dkrz.de/svideo/vosaic/sat/d3_64_5.html)) wurde dieses Archiv in leicht modifizierter Form verwendet.

Von der Firma RealNetworks wurde für nicht-kommerzielle Zwecke ein frei verfügbarer Video-Server mit eingeschränktem Leistungsumfang angeboten, der bei uns eingesetzt wird. Er liegt in der Version 4.0.0 vor und kann unter Solaris 2.5 betrieben werden. Der Server kann 60 „Streams“ gleichzeitig bereitstellen. Die mit dem „Light“-Server ausgelieferte Konvertierungssoftware bietet keine Möglichkeiten zum Schneiden oder Editieren der Videos. Das Hilfe-Menü läßt darauf schließen, daß solche Funktionen in der Voll-Version vorhanden sind. Zur Übertragung wird das firmeneigene „RealMedia“-Dateiformat eingesetzt, das Video- und Audiospur in einer Datei vereint. Daher treten bei der Übertragung mit der Real-Software keine Synchronisationsprobleme auf.

### **Arbeitsablauf zur Erstellung eines Angebotes**

Wie aus Abb. 8 hervorgeht, mußte zunächst ein Video in einem gängigen digitalen Format erzeugt werden. Die Qualität des digitalen Videos bestimmte die Qualität der daraus resultierenden „Streams“. Voraussetzung für diesen Arbeitsschritt war ein leistungsfähiger Rechner mit ausreichendem Arbeitsspeicher. Zusätzlich war eine Hardwareunterstützung (Grafikkarte) erforderlich.

Als Ausgangsformat setzten wir das aus der PC-Welt stammende AVI-Format (Audio Visual Interleaved) ein, da dieses Format von den Konvertierungswerkzeugen beider Firmen unterstützt wird und Bild und Ton in einer Datei speichert.

Da uns in der Unix-Welt leistungsfähige Rechner zur Verfügung standen, verwendeten wir außerdem MPEG1 als digitales Video-Ausgangsformat, das nur von der Vosaic-Software weiterverarbeitet werden kann. Bei dieser Variante wird die Tonspur in einer zweiten Datei gespeichert.

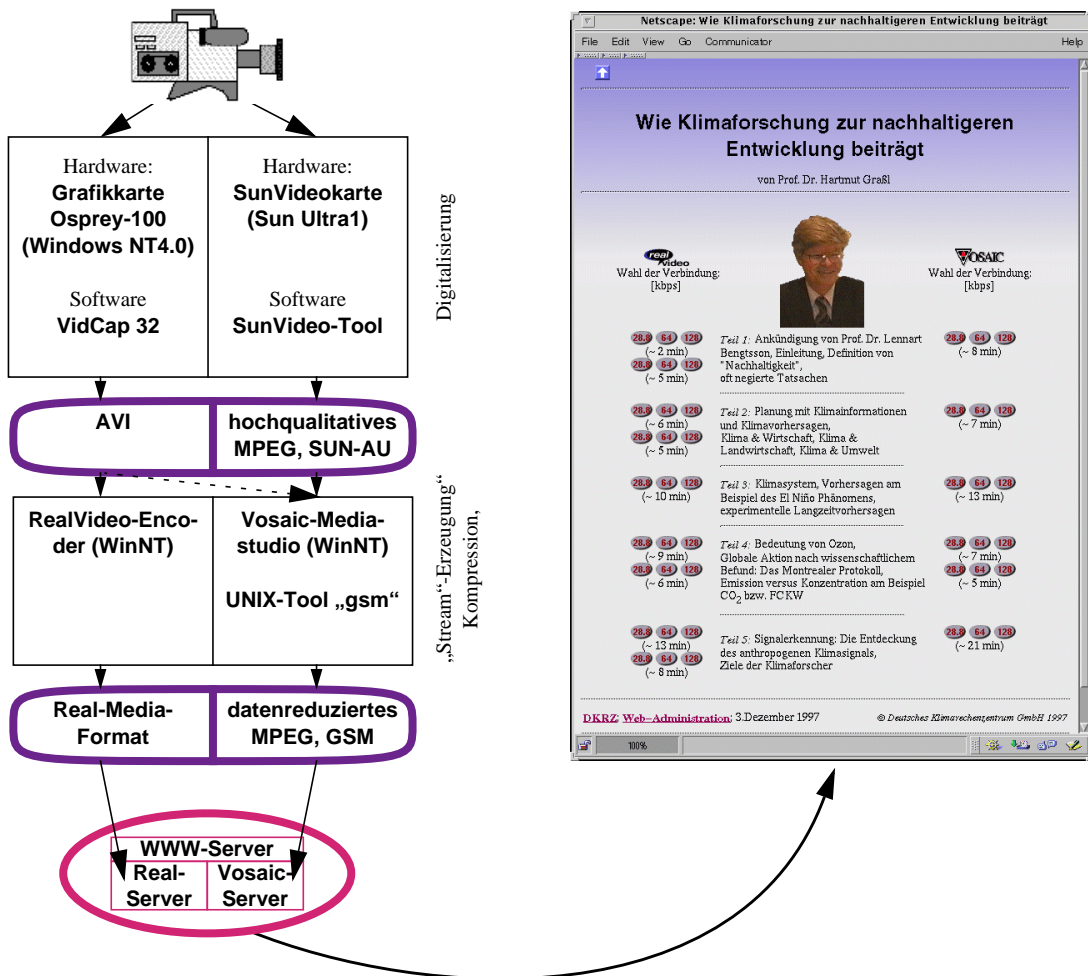


Abb. 8: Arbeitsschritte vom Video-Band zum digitalen „Streaming“-Video

Der Digitalisierung folgte die Konvertierung und die Erstellung von „Streams“ für definierte Bandbreiten. Zum Schluß wurden die Videosequenzen annotiert.

### Vosaic: Digitalisierung

Die Vosaic-Software überträgt einen separaten Video- und einen Audio-„Stream“. Daher kann schon bei der Digitalisierung eine Trennung der Medien erfolgen. Dabei muß jedoch beachtet werden, daß bei diesem Schritt ein Video mit konstanter Frame-rate erzeugt wird, damit die (Audio/Video) Datenströme später synchronisiert übertragen werden können. Wir verwendeten einen Sun Ultra1-Rechner mit der SunVideo-Karte und der dazugehörigen Software, die neben dem digitalen MPEG1-Video eine Audio-Datei im Sun-Audio-Format speichert.

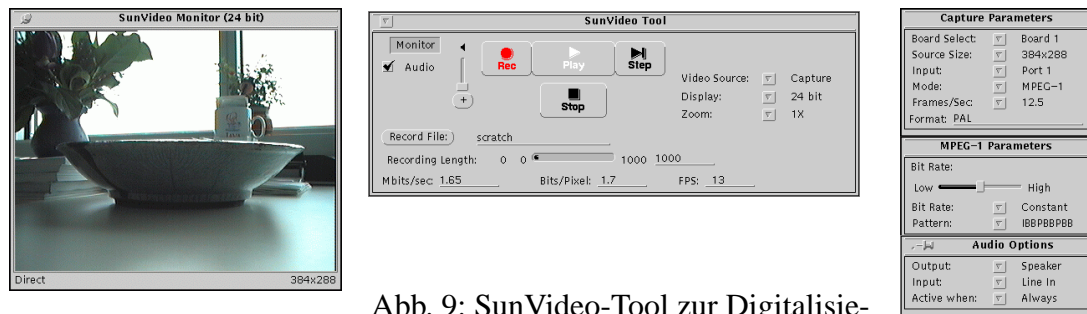


Abb. 9: SunVideo-Tool zur Digitalisierung analoger Videoquellen

Bei der Generierung von MPEG-Videos wird ein verlustbehaftetes Kompressionsverfahren verwendet. Da sich aufeinanderfolgende Bilder in Videosequenzen häufig nicht stark unterscheiden, werden nur wenige Bilder, die sogenannten „I“-Frames („Intra-Frames“), komplett gespeichert. Die „Predicted“ oder „P-Frames“ sind dazwischenliegende Bilder und werden nur als Änderung gegenüber dem letzten I- oder P-Frame abgespeichert. B-Frames („Between“) enthalten Informationen über die Änderungen bezogen auf Vergangenheit und Zukunft. Die Berechnung der Änderungen erfolgt auf sogenannten Macroblocks, das sind 16x16 Bildpunkte große Bildflächen. Die Abfolge der Frames wird durch das Kodierungspattern beschrieben, das als Parameter bei der Videoaufnahme gewählt wird. Ein weiterer Aufnahme-Parameter ist die Framerate.

Ein Nachteil des SunVideo-Tools bestand darin, das überwiegend nicht ganzzahlige Frameraten (5, 6.2, 8.3, 12.5, 25 Frames/Sekunde) gewählt werden konnten. Die Weiterverarbeitung mit den Vosaic-Werkzeugen erwartet jedoch ganzzahlige Raten, um eine optimal Synchronisation zu gewährleisten. Da eine Framerate von 25 Frames/Sekunde auch von einem Sun Ultra-1-Rechner nicht in Echtzeit generiert werden konnte und eine Framerate von 5 Frames/Sekunde kein flüssiges Video mehr erzeugt, haben wir die Filme in der Regel mit der Framerate 12.5 aufgenommen. Als Kodierungspattern wurde IBBPBBPBB verwendet.

### Vosaic: „Stream“-Generierung und Kompression

Mit dem frei verfügbaren Unix-Tool „toast“ konnte aus der Sun-Audio-Datei das von Vosaic unterstützte Audio-Format GSM (Global System for Mobile telecommunication) erzeugt werden, das eine konstante Bitrate von 13 kbit/s besitzt. Der Audio-Konverter wird unter der URL <http://kbs.cs.tu-berlin.de/~jutta/toast.html> bereitgestellt.

Das Vosaic-MediaStudio bot neben der Möglichkeit der Konvertierung und der Analyse auch ein Menü „Downsample“ für qualitativ hochwertigen MPEG-Dateien an. Hier konnten die Bildgröße und die Framerate verringert sowie eine gewünschte Datenrate vorgegeben werden. Die Dimensionen der Bildgröße mußten immer als Vielfaches von 16 und die Framerate als Teiler der ursprünglichen Rate gewählt werden. Der Parameter Datenrate diente nur als Vorschlag und wurde vor allem dann nicht eingehalten, wenn die Bildfolge schnellen Änderungen unterworfen ist.

Das wird in Abb. 10 deutlich. Im Diagramm ist der zeitliche Verlauf der Datenrate des El Niño Videos (<http://www.dkrz.de/dkrz/kommunikation/svideo/klimasimulation/>

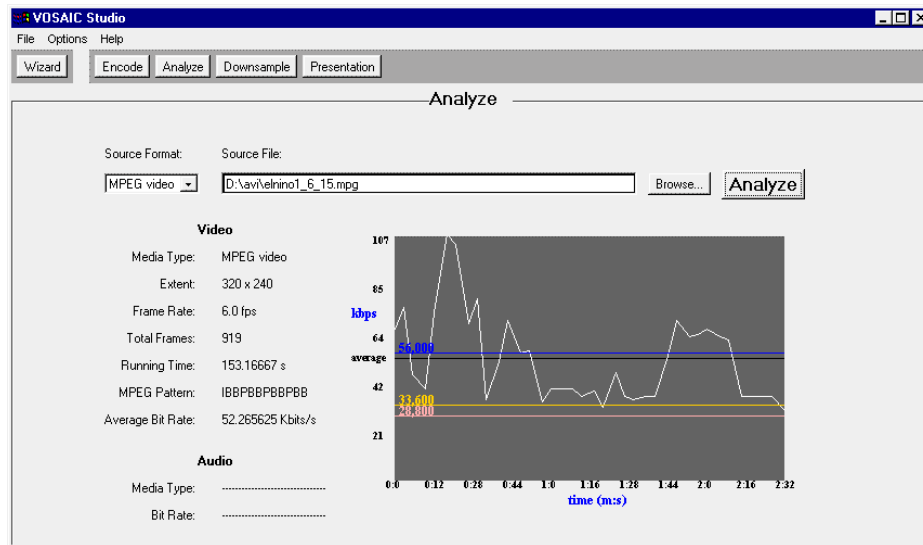


Abb. 10: Analyse-Werkzeug des Vosaic-MediaStudios

elnino1\_6\_15.html) dargestellt. Der in dieser Abbildung analysierte Film beginnt mit einem schnellen Wechsel von Bildern, was sich in den Peaks am Anfang des Diagramms widerspiegelt. Darauf folgt die Visualisierung einer Klimasimulation, bei der sich nur wenige Bereiche im Bild verändern und die Datenrate auf einen niedrigeren quasi konstanten Wert eingeegelt werden kann.

Zur Analyse und zur Übertragung von MPEG-Dateien benötigt Vosaic eine Beschreibung der MPEG-Codierung (siehe Anhang E). Jeder Frame wird mit seiner Funktion (I, B, P), seiner Größe und seiner Position innerhalb des Datenstroms beschrieben. Diese Information wird bei der Konvertierung oder beim „Downsample“-Arbeitsgang erzeugt und in einer Datei mit der Endung .par gespeichert. Die Beschreibung dient bei der real-time Übertragung zur Steuerung des Datenstromes.

Mit dem Ziel, eine längere Sequenz aktueller Satellitenbilder als Streaming-Video anzubieten, haben wir mit Hilfe der Berkely MPEG Tools *mpeg\_stat* und *mpeg\_play* einen Parser konstruiert, der unter Unix lauffähig ist und im Hintergrund automatisch eine Beschreibungsdatei erzeugen kann. Das Tool *mpeg\_encode* generiert aus Einzelbildern, die bei uns per Satellit empfangen werden, stündlich ein aktuelles MPEG-Video (siehe [http://www.dkrz.de/svideo/vosaic/sat/d3\\_64\\_5.html](http://www.dkrz.de/svideo/vosaic/sat/d3_64_5.html)), das zusammen mit dieser Beschreibungsdatei vom Vosaic-Server bereitgestellt wird.

Für die Einbindung der „Streams“ in HTML-Seiten wurden als Objekte eingebundene Textdateien verwendet, die alle zur Lokalisation benötigten Informationen enthalten (siehe Listing 2).

```

Class      vidaud
VideoType  mpeg1
VideoProtocol  vdpvideo
VideoLocation  elnino/elnino1_6_15.mpg
VideoHost      info.dkrz.de
VideoPort      1235
VideoStartTime 0
AudioType      gsm
AudioProtocol  vdpaudio
    
```

AudioLocation elnino/elnino2.au.gsm  
 AudioHost info.dkrz.de  
 AudioPort 1235  
 AudioStartTime 0

Listing 2: Beschreibung eines „Stream“-Objektes, abgelegt in „datei.vos“

## Vosaic: Client-Plattformen und Anforderungen

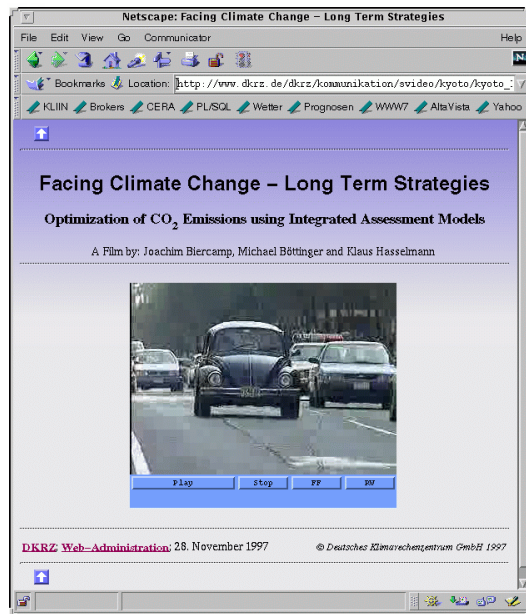


Abb. 11: Vosaic-Client (Plugin)

Zur Wiedergabe der Filme wird beim Nutzer ein Plugin für den Netscape Navigator benötigt. Der *Vosaic-MediaClient* in der Version 1.06 ist frei verfügbar und zur Zeit für folgende Betriebssysteme erhältlich:

- Macintosh
- Windows 95/NT
- SPARC Solaris 2.5 (nicht IPC)
- IRIX 5.3

Weitere Voraussetzungen sind der Netscape Navigator ab Version 3.0, wobei ab Version 4.x eine bessere Ausnutzung der Ressourcen erreicht wird. Außerdem sollte eine Grafikkarte vorhanden sein, die mehr als 8-Bit Farbtiefe erlaubt, da

Netscape allein schon viele Farben verbraucht und für ein Plugin keine separate Colormap alloziiert wird. Es werden eine Soundkarte und Lautsprecher oder Kopfhörer benötigt. Da der Vosaic-Client keine Lautstärkenregelung besitzt, muß diese über die zur Soundkarte gehörende Software gesteuert werden.

## RealAudio/Video: Digitalisierung

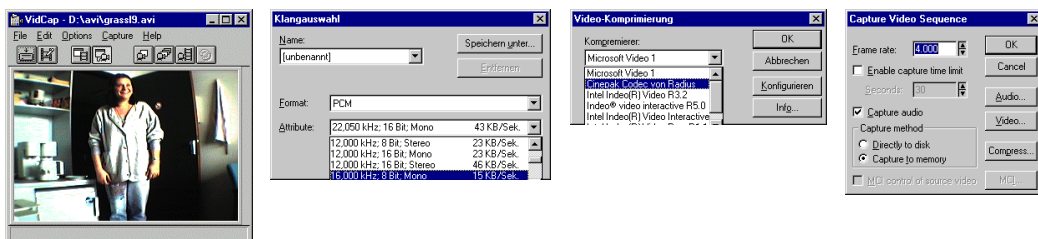


Abb. 12: VidCap32 zur Digitalisierung analoger Videoquellen

Als Ausgangsformat für die Datenreduktion mit dem Real-Encoder wurde das AVI- (Audio Video Interleaved) Format benötigt. Da dieses Format seinen Ursprung im

PC-Bereich hat, werden auch die Digitalisierungswerkzeuge für PC's angeboten. Wir verwendeten zur Erzeugung eine Osprey-100 Grafikkarte mit der Software „VidCap32“ (siehe Abb. 12) auf einem Pentium 200Pro mit 98 MB Arbeitsspeicher unter Windows NT 4.0.

Bei diesem Verfahren beschränkt die Größe des Arbeitsspeichers sowohl die Qualität als auch die mögliche Länge einer Sequenz. Falls man die gewonnenen Frames direkt auf die Festplatte schreibt, werden viele Frames verworfen, da die Hardware dies in Echtzeit nicht erlaubt. Daher sollte mit der Option „Capture to memory“ gearbeitet werden. Eine weitere Schwierigkeit bei der Digitalisierung bestand darin, daß der Filmschnitt im kleinen Sucher der Videokamera durchgeführt werden mußte, da das Bildfenster während der Aufnahme nicht erneuert wurde.

### RealAudio/Video: „Stream“-Generierung und Kompression

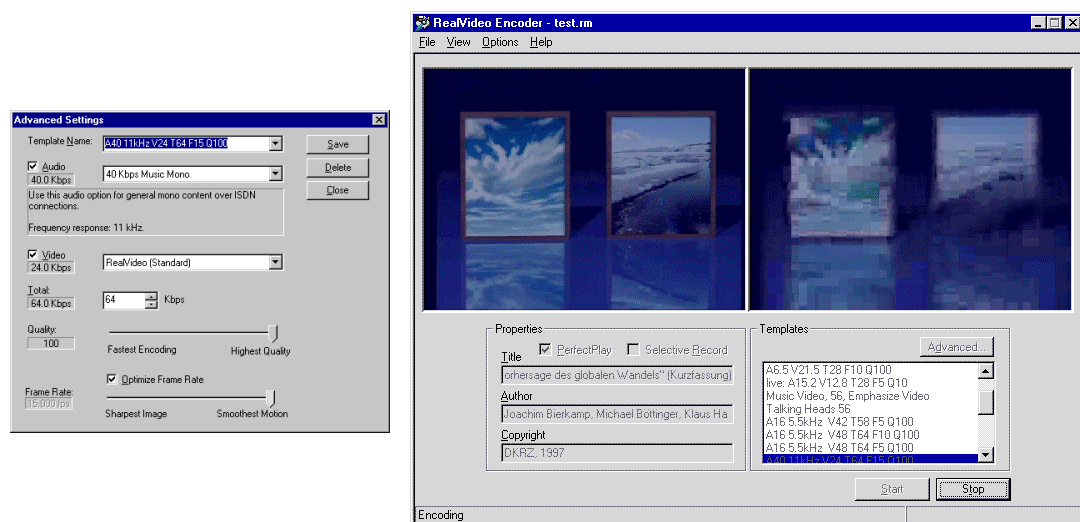


Abb. 13: Datenkomprimierung mit dem Real-VideoEncoder

Die Komprimierung der AVI-Daten zum RealNetworks-eigenen Real-Media-Format (.rm) war sehr rechenzeitaufwendig. Abb. 13 macht die Qualitätseinbuße bei der Datenreduktion deutlich. Das linke Videofenster zeigt das Original, im rechten Fenster wird das Ergebnis der Komprimierung dargestellt.

Für die Kompression werden Parameter-Kombinationen als Templates vorgeschlagen, die Anwendungsfälle wie „Talking Head“, „Music-“ und „High-Action-Video“ für verschiedene Zielbandbreiten unterscheiden. Es lassen sich Frameraten bis zu 15 Frames pro Sekunde generieren.

Die WWW-Einbindung der Real-Media-Daten erfolgte mit Hilfe einer als Link in eine HTML-Seite integrierten Metadatei, deren Endung durch den MIME-Typ „audio/x-pn-realaudio“ beschrieben wird. Sie enthält die eigentliche URL (z.B.: pnm:/info.dkrz.de/klim1/klim1\_a16\_5.5kHz\_t90\_f15.rm) des „Streams“ und vermittelt sie an den Real-Server.

## RealAudio/Video: Client-Plattformen und Anforderungen

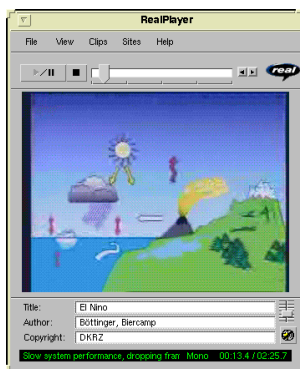


Abb. 14: RealVideo-Client als Helper-Applikation

Der *RealPlayer* liegt zur Zeit in der Version 5.0 vor und wird als Helper-Applikation in einem beliebigen WWW-Browser konfiguriert. Er ist für folgende Betriebssysteme erhältlich:

- Windows 95/98/NT
- Windows 3.1
- Mac OS 7.x/8.x
- Linux 2.0
- Solaris 2.5
- Irix 6.2/6.3
- OS/2

Außerdem ist ein leistungsfähiger Rechner zu empfehlen (beispielsweise hatte ein Pentium 90

Probleme bei der Darstellung von „Streams“ mit 64 kbps und mehr); es sollte schon ein 166 MHz PC oder höher eingesetzt werden. Falls die System-Leistung für die gewünschte Bandbreite nicht ausreicht, macht sich das (neben einem Hinweis in der Statuszeile des Viewers) durch unangenehmes „Knattern“ in der Tonspur und Stoppen der Bildfolge bemerkbar. Unter Solaris traten Probleme mit einigen Grafikkarten auf, die 24-Bit Farbtiefe ermöglichen (MGX, SX, nicht jedoch die Parallaxkarte).

Wer über einen leistungsfähigen Rechner, aber eine schmalbandigen Internetanbindung verfügt, kann den RealPlayer so konfigurieren, das die Video-Daten lokal gepuffert werden, bis das störungsfreie Abspielen eines „Streams“ mit größerer Bandbreite möglich ist. Im Extremfall wird allerdings die gesamte Datenmenge geladen, so daß man dann nicht mehr von einer „streaming“-Übertragung reden kann.

Der RealPlayer bietet die Möglichkeit, Statistiken wie beispielsweise übertragene Bandbreite oder Datenverlust anzuzeigen. Damit ist er auch in der Lage, die Last auf einer Leitung widerzuspiegeln.

### Angebot: Klima-Videos und Vorträge

Für eine optimale Wiedergabe muß ein entsprechend der Internetanbindung aufbereiteter „Stream“ ausgewählt werden. Üblicherweise werden vom Hersteller bzw. Anbieter der Filme Sequenzen für unterschiedliche Bandbreiten bereitgestellt. Wir haben nach den oben beschriebenen Verfahren Filme für die Bandbreiten 28.8 kbps, 64 kbps und 128 kbps erzeugt. Damit wird eine Telefonleitung mit Modemanschluß, eine ISDN-Verbindung und eine doppelte ISDN-Verbindung (und mehr) unterstützt. Die Videos können über die URL <http://www.dkrz.de/dkrz/kommunikation/svideo/svideo.html> erreicht werden.

Im Online-Angebot sind das DKRZ-Video „Klimasimulationen – Vorhersage des globalen Wandels“ und der zur Weltklimakonferenz in Kyoto im Dezember 1997 produzierte DKRZ-Film „Facing Climate Change – Long Term Strategies“. Der zuerst genannte ca. 19 Minuten lange Film wurde in vier Abschnitte unterteilt, um Ein-

sprungmöglichkeiten zu den verschiedenen Themen zu bieten. Es ist mit keinem der am DKRZ eingesetzten Videoservertypen möglich, anhand eines Inhaltsverzeichnis in eine längere Sequenz themenbezogen hineinzuspringen.

Der zweite ca. 15 Minuten lange Film steht als ein zusammenhängender „Stream“ bereit. Bei einem Film dieser Länge wird in der Vosaic-Variante ein Synchronisationsproblem deutlich, welches durch die Kombination der Digitalisierungssoftware (SunVideo) und der Kompressionssoftware von Vosaic entsteht. Ton und Bild laufen am Ende des Filmes deutlich auseinander.

Neben den DKRZ-Videos wurde der ca. eineinhalb Stunden dauernde Vortrag, den Prof. Dr. Hartmut Graßl 1996 zum „Tag der Forschung zum Thema Klima und Wetter“ gehalten hat, in mehreren Abschnitten digitalisiert und im „Streaming“-Verfahren präsentiert. Bei diesem Vortrag ist die Tonspur deutlich wichtiger als das Bild. Die auf Folien dargestellten Kernaussagen des Vortrags wurden bei der Vosaic-Variante als Text oder Bild in die HTML-Seite integriert, in die auch der „Stream“ als Plugin-Objekt eingebunden ist.

Eine vollständig automatisch ablaufende Video-Generierung aus Einzelbildern liefert stündlich einen aktuellen Satellitenfilm, der mit Vosaic betrachtet werden kann. Hier wird besonders deutlich, daß die Bildqualität eines MPEG-Film mit einer Bandbreite, die ein quasi gleichzeitiges Laden und Ansehen ermöglicht, gering ist. Als Vergleich stehen die „Download“-MPEG-Sequenzen der URL <http://www.dkrz.de/sat/dkrz/> zur Verfügung, die nur aus wenigen Frames bestehen, aber eine deutlich bessere Qualität aufweisen.

Die Abrufzahlen der DKRZ-WWW-Statistik zum Start des Angebots im Dezember plazierte die „Streaming“-Videos auf Rang 10 (siehe <http://www.dkrz.de/statistik/analog9712.html#Directory> und Listing 3) der „Hitliste“.

Directory Report

(Go To: Top: Daily Report: Hourly Summary: Domain Report: Request Report: Referrer Report: Browser Summary: Browser Report: Status Code Report: Error Report)

Printing all directories with at least 0.05% of the requests for pages, sorted by number of page requests. Printing directories to depth 3.

```
pages: %reqs: %bytes: #reqs: Mbytes: directory
-----
27087: 4.02%: 0.80%: 34590: 92: [root directory]
18219: 2.68%: 9.29%: 23114: 1068: /sat/
14710: 2.30%: 1.03%: 19809: 118: /ngdoc/ng4.0.1/ref/
14706: 2.30%: 1.03%: 19776: 119: /ngdoc/ng4.0/ref/
12368: 1.44%: 1.07%: 12436: 123: /mirror/tnp/
6868: 6.67%: 16.66%: 57457: 1915: /sat/meteo_france/
6553: 0.76%: 0.67%: 6566: 78: /~k202045/tnp/
4558: 0.81%: 0.34%: 7003: 39: /dkrz/
4030: 0.47%: 0.15%: 4030: 18: /info/
3300: 0.43%: 0.19%: 3723: 21: /dkrz/kommunikation/svideo/
```

Listing 3: Ausschnitt aus der DKRZ-WWW-Serverstatistik im Dezember 1997

## 3.2.2 Visualisierung von Klimadaten

### 3D-Visualisierung: Vis5D versus VRML

Vis5D ist ein frei verfügbares 3D-Visualisierungsprogramm, das in der Lage ist,

meteorologische Modelldaten in fünf Dimensionen darzustellen. Damit sind die drei Raumkoordinaten, die Zeitkomponente und die Dimension der physikalischen Größen erfaßt. Der Benutzer erhält keine fertige Darstellung der Daten, sondern die Möglichkeit, räumliche und zeitliche Veränderungen unterschiedlicher Parameter interaktiv zu visualisieren, zu animieren und zu verstehen (siehe Abb. 15).

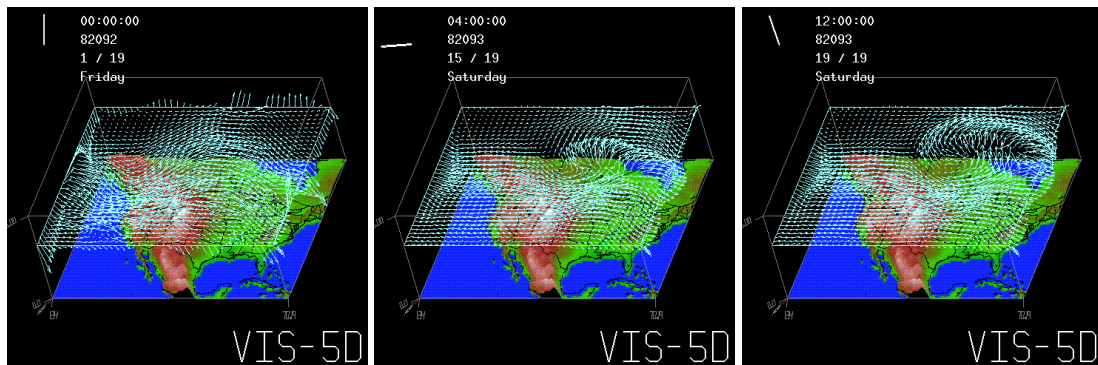


Abb. 15: Zeitliche Veränderung *einer* physikalischen Größe (vertikale Winde) auf *einer* Grenzfläche

## Vis5D: Quelle und Konfiguration

Die jeweils aktuelle Software wird in Absprache mit den Entwicklern des „Space Science and Engineering Center“ (SSEC) der „University of Wisconsin-Madison“ vom DKRZ gespiegelt und bereitgestellt (<ftp://ftp.dkrz.de/pub/visu/ssec/>). Es gibt Versionen für alle gängigen UNIX-Derivate. Beispiele und Anleitungen sind unter der URL <http://www.dkrz.de/dkrz/kommunikation/kliin/Vis5D/> zusammengestellt. Die Software besteht aus einem Viewer, Werkzeugen zur Konvertierung von Gitterdaten und Topologie- bzw. Küstenlinien-Datensätzen.

Der Viewer ist für den Einsatz im WWW konfigurierbar, wenn als MIME-Type „application/Vis5D v5d“ eingetragen wird und der Vis5D-Aufruf mit folgender Zeile (Datei .mailcap) festgelegt wird: „application/Vis5D; <Verzeichnis, in dem sich Vis5D befindet>Vis5D %s -path <Verzeichnis mit der mitgelieferten Topographie>“.

## Vis5D: Einsatz und Grenzen

Das Programm läßt sich besonders gut auf Grafik-Workstations einsetzen, die über eine 3D-Hardware-Beschleunigung verfügen. Speziell auf Maschinen, bei denen die Grafik-Sprache OpenGL direkt über die Grafik-Hardware unterstützt wird (z.B. Silicon Graphics), bietet das Programm ein sehr hohes Maß an Interaktivität selbst bei komplexen Darstellungen. Dies beinhaltet die Animation zeitabhängiger Abläufe des modellierten Wettergeschehens ohne die Notwendigkeit der expliziten Zwischenspeicherung der Bilder.

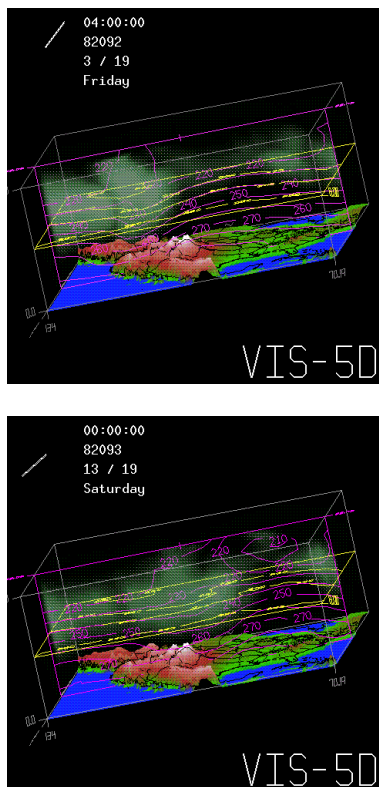


Abb. 16: Volumen-Rendering, zwei Zeitschritte

Doch auch mit weniger leistungsfähigen Workstations läßt sich dieses Programm sinnvoll einsetzen, um die zeitlichen Änderungen visualisierter Modellgrößen zu verstehen. Hier muß die Darstellung allerdings für jede Änderung der 3D-Perspektive komplett neu berechnet werden, was je nach Komplexität und CPU-Leistung einige Sekunden dauern kann. Besonders das Volumen-Rendering (siehe Abb. 16) eines Parameters ist sehr rechenintensiv.

Die zugrundeliegenden Datensätze sind sehr groß (5 bis 50 MB und mehr). Daher muß der lokale Rechner über genügend temporären Plattenplatz und Hauptspeicher verfügen.

Ein Grundverständnis für das Modell, aus dem die Daten stammen, ist Voraussetzung bei der Visualisierung prägnanter Phänomene.

Vis5D erlaubt das Abspeichern von Einzelbildern und mit Hilfe der Script-Sprache Tcl auch das automatische Abspeichern von Bildsequenzen. Diese lassen sich mit geeigneten Werkzeugen zu (MPEG) Filmen verarbeiten. Die Filme können im WWW entweder auf herkömmliche Weise als MPEG-Datei zum Laden bereitgestellt oder mit Hilfe von

„Streaming Video“ angeboten werden.

Auf diese Weise vorprozessierte Bilder oder Filme sind einerseits als „Daten-Preview“ für Meteorologen nützlich. Andererseits sind sie als Informationsmaterial für den interessierten Laien meistens ausreichend, da die Visualisierung mit Vis5D auch Verständnis für die Daten und Erfahrung mit dem Werkzeug erfordert.

### Unterschied von Vis5D und VRML

Mit VRML hat man die Möglichkeit, eine Welt mit 3D-Geometriedaten aufzubauen und sich darin zu bewegen. Es sind Verknüpfungen mit Hyperlinks möglich und es können 3D-Animationen angestoßen werden.

Im Gegensatz dazu arbeitet Vis5D auf einem 5-dimensionalen Datensatz und bietet Werkzeuge zur dynamischen Visualisierung. Der Benutzer kann nicht nur entscheiden, was er aus welchem Blickwinkel betrachten möchte, sondern er kann auch bestimmen, wie er ein Phänomen darstellen will. Dazu kann er neben dem Blickwinkel auch die physikalischen Parameter (wie Temperatur, relative Feuchte, usw.), deren Farben und Transparenz und die Darstellungsform (Schicht, Grenzfläche) auswählen. Allerdings ist Vis5D nur für die Domäne der Klimamodelldaten einsetzbar, während sich mit VRML beliebige Geometriedaten darstellen lassen.

## Automatische Generierung wissenschaftlicher Animationen mit GrADS

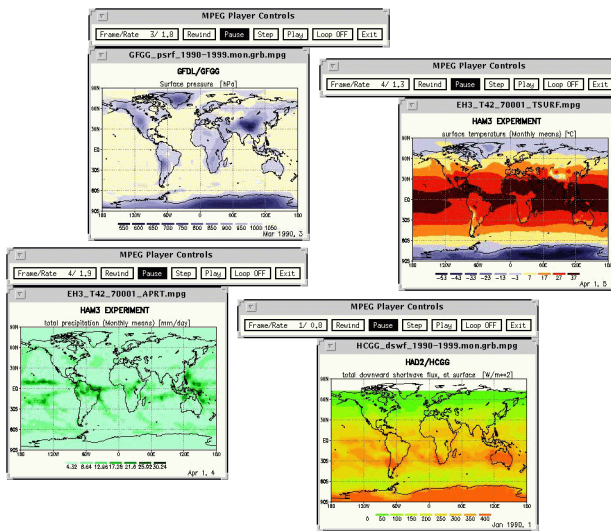


Abb. 17: Previews der Klimamodelldaten

Um Klimamodelldaten auch Nutzern nahe zu bringen, die nicht Meteorologie studiert haben, wurde ein Mechanismus zur automatischen Generierung von Animationen aus Zeitserien entwickelt. Klimamodelldaten sind Tupel von Variablen in einem Gitter, dessen Koordinaten Längen- und Breitengrade, die Höhe und die Zeit sind. Sie werden beim DKRZ und bei kooperierenden Institutionen im Datenaustauschformat GRIB (GRIdded Binary) archiviert. Dieses Format wurde von der WMO Commission for Basic-Systems (CBS) für großvolu-

mige Gitterdaten entwickelt und ist selbstbeschreibend.

Zur Visualisierung haben wir die GrADS-Software eingesetzt. Das Grid Analysis and Display System ist ein Public-Domain-Softwarepaket, welches speziell auf meteorologische Daten zugeschnitten ist, die auf Gittern vorliegen. Da dies für die meisten der am DKRZ erzeugten Modelldaten zutrifft, können diese also auch mit GrADS analysiert und dargestellt werden. Obwohl das Datenmodell des Paketes zeitabhängige dreidimensionale Daten unterstützt, ist die Graphik auf rein zweidimensionale Darstellungen beschränkt.

GrADS ist außerordentlich interaktiv; langwieriges Übersetzen, Laden und Ausführen von Programmen ist nicht notwendig. Über eine „Command-Language“ können Kommandos direkt ausgeführt werden. In Form von Skripten können vorgefertigte interaktive Applikationen entwickelt werden, die auf die jeweiligen Anforderungen eines Benutzers zugeschnitten sind. Außerdem ist es möglich, das Programm im Hintergrund arbeiten zu lassen. Animationen über die Zeit oder eine der räumlichen Dimensionen gehören ebenfalls zum Funktionsumfang von GrADS. Zum Sichern einer Grafik von GrADS kann eine binäre Ausgabedatei erzeugt werden. Diese Datei ist ein Metafile und kann mit einem speziellen Programm (*gxgif*) in eine GIF-Bild verwandelt werden.

GrADS benötigt neben dem GRIB-Datensatz eine Descriptordatei. Dabei handelt es sich um eine ASCII-Beschreibung der Daten, die zunächst automatisch mit dem Programm *grb2gas* aus den Daten generiert werden kann (siehe Listing 4).

```
GrADS Descriptor file
* generated by grb2gas RELEASE 2.6
TITLE HAM3 EXPERIMENT
*
OPTIONS YREV
*
```

```

DSET ^EH3_T42_70002_GPH30
DTYPE grib
INDEX ^EH3_T42_70002_GPH30.gmp
*
UNDEF -1.0e+34
*
XDEF 128 LINEAR 0.0000 2.8125
*
YDEF 64 LEVELS -87.8638
-85.0965 -82.3129 -79.5256 -76.7369 -73.9475 -71.1578 -68.3678 -65.5776
-62.7874 -59.9970 -57.2066 -54.4162 -51.6257 -48.8352 -46.0447 -43.2542
-40.4636 -37.6731 -34.8825 -32.0919 -29.3014 -26.5108 -23.7202 -20.9296
-18.1390 -15.3484 -12.5578 -9.7671 -6.9765 -4.1859 -1.3953 1.3953
4.1859 6.9765 9.7671 12.5578 15.3484 18.1390 20.9296 23.7202
26.5108 29.3014 32.0919 34.8825 37.6731 40.4636 43.2542 46.0447
48.8352 51.6257 54.4162 57.2066 59.9970 62.7874 65.5776 68.3678
71.1578 73.9475 76.7369 79.5256 82.3129 85.0965 87.8638
*
ZDEF 1 LEVELS 30
*
TDEF 144 LINEAR 12Z30Jan0001 1mo
VARS 1
GPH 1 156,100,0 :geopotential height:gpm:::
ENDVARS
    
```

#### Listing 4: Descriptordatei des GRIB-Datensatzes EH3\_T42\_70002\_GPH30

Klimamodelldaten eines Experimentes werden häufig in Datensätzen gespeichert, die den monatliche Mittelwert einer Variablen über einen Zeitraum beschreiben. Für diese Datensätze haben wir einen Mechanismus entwickelt, der den zeitlichen Ablauf der Variable in einem MPEG-Film darstellt. Der Mechanismus wurde mit Hilfe eines Perl-Scripts realisiert.

Wir haben dazu eine Zuordnung von Farbpalette und Parameter definiert. Eine Darstellung der Definitionen kann Anhang F entnommen werden.

Der von dem Perl-Script gesteuerte Ablauf zur automatischen Generierung der MPEG-Filme ist in Abb. 18 schematisch dargestellt. Das Perl-Script erzeugt die Descriptordatei und sammelt beschreibende Informationen zum Datensatz. Es erzeugt daraus ein GrADS-Script und ruft es auf. Falls Datensätze aus der DKRZ-Datenbank verarbeitet werden, stammt ein Teil der beschreibenden Informationen zum Datensatz ebenfalls daher. Dazu wurde ein Pro\*C-Programm (*parameter\_info*) entwickelt.

Die von GrADS erzeugten Grafiken werden als Metadateien gespeichert, zu GIF-Bildern konvertiert und in das JPEG-Format umgewandelt. Dieses Format ist ein mögliches Ausgangsformat zur Erzeugung von MPEG-Filmen. Das Perl-Script, welches den Ablauf steuert, erzeugt eine Parameter-Datei und stößt die MPEG-Generierung an. Die Erzeugung von MPEG-Filmen wurde bereits in Kap. 3.2.1 "Echtzeit Video/Audio im WWW" auf Seite 16 beschrieben.

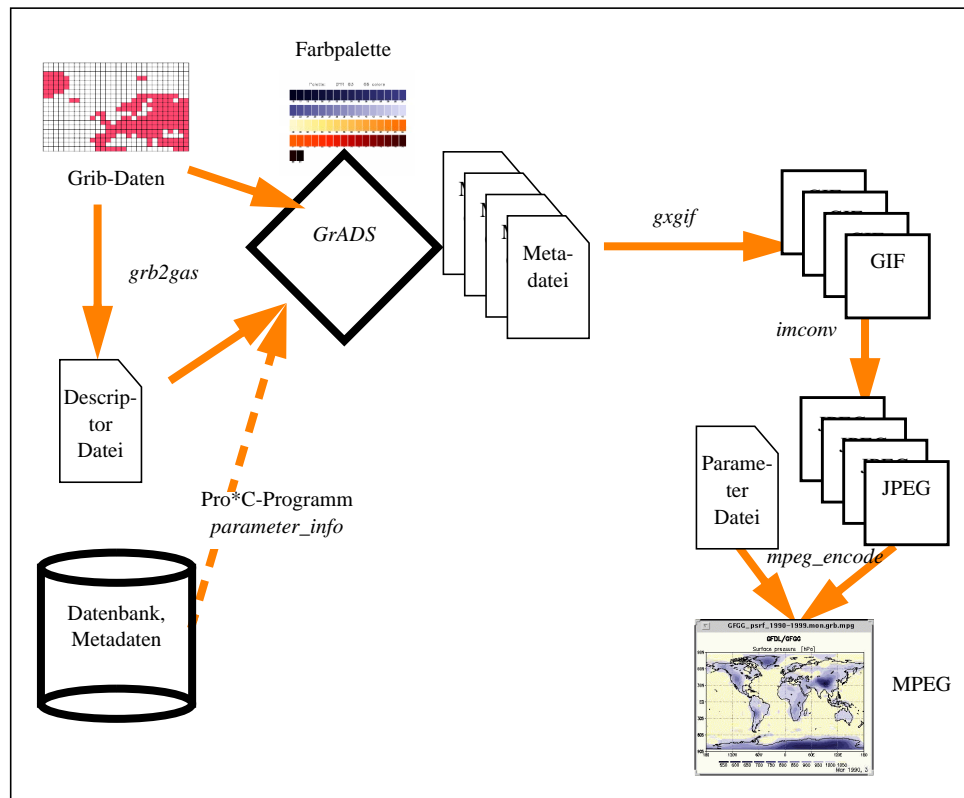


Abb. 18: Erzeugung eines MPEG-Videos aus Klimamodell Datensatz

Die Erzeugung der MPEG-Previews wird über das WWW-Datenbank-Interface angestoßen, falls ein Datensatz geordert wird. Ferner wurden für die Klimamodell-Datensätze des Data Distribution Centre (DDC) des Intergovernmental Panel on Climate Change (IPCC), die beim DKRZ lokalisiert sind, Filme erzeugt ([http://ipcc-ddc.cru.uea.ac.uk/dkrz/dkrz\\_index.html](http://ipcc-ddc.cru.uea.ac.uk/dkrz/dkrz_index.html)).

### 3.3 Präsentationen und Veröffentlichungen

Das Projekt war mit folgenden Vorträgen auf Arbeitskreisen und Workshops vertreten:

- Sabine Reimers, „Multimediale Klimainformation im Internet: Vorstellung des Projektes“, Arbeitskreis Informationssysteme, Januar 1997 in Berlin
- Sabine Reimers, Michael Lautenschlager, „Access to Distributed Climate and Environmental Data“, EO/GEO Workshop, Februar 1997, Washington, DC, USA
- Sabine Reimers, „Datenbankintegration ins WWW am Beispiel einer Klima-Datenbank“, Arbeitskreis Informationssysteme, Februar 1998 in Berlin

In den Zeitschriften des DKRZ („Gigaflops“) und des DFN-Vereins („DFN Mitteilungen“) sowie im WWW wurden Teilaspekte des Projektes vorgestellt:

- Sabine Reimers, „Streaming Audio/Video im WWW“, Gigaflops 4/1997, (<http://>

[www.dkrz.de/gigaflops/gigaflops97-4/GIGA-12.html](http://www.dkrz.de/gigaflops/gigaflops97-4/GIGA-12.html))

- Sabine Reimers, Gerrit Henken, „Multimediale Klimainformation im Internet (KLIIN), Daten und Informationen rund ums Klima“, DFN-Mitteilungen 48, November 1998, (<http://www.rtb-nord.uni-hannover.de/dfn/mitteilungen/html/heft48/index.html>), in Vorbereitung

Im Projektverlauf wurden zwei Zwischenberichte erstellt und dem DFN-Verein übergeben:

- Sabine Reimers, Gerrit Henken, „Multimediale Klimainformation im Internet: Zwischenbericht zum Meilenstein M1“, Deutsches Klimarechenzentrum GmbH, Juli 1997
- Sabine Reimers, Gerrit Henken, „Multimediale Klimainformation im Internet: Zwischenbericht zum Meilenstein M2“, Deutsches Klimarechenzentrum GmbH, Dezember 1997

## 4 Zusammenfassung



Im Projekt wurde ein offener Zugang zu Klimainformationen sowie zu Informationen des allgemeineren multidisziplinären Problemkreises des „Globalen Wandels“ („Global Change“) geschaffen. Ein zweiter Schwerpunkt lag in der Visualisierung von Klimadaten und -informationen: Präsentationsfilme aus dem Global-Change-Bereich wurden im WWW bereitgestellt, es wurden Verfahren zur wissenschaftlichen Animation von Klimadaten in den Datenbereitstellungs-Prozeß integriert und es wurden Möglichkeiten der dreidimensionalen

Visualisierung (Vis5D, VRML) untersucht. Die Einstiegsseite des Projektes findet man unter <http://www.dkrz.de/dkrz/kommunikation/kliin/>.

Eine Bestandsaufnahme von WWW-Datenquellen und -Projekten aus den Bereichen „Earth Science“, „Earth Observation“ zeigte auf, daß die Standardisierungsbemühungen von Daten bzw. Metadaten zwar weit vorangeschritten sind, aber der Datenzugang über WWW in der Praxis häufig aus unterschiedlichen Gründen nicht im gewünschten Maße möglich ist. Dies ist zum Teil technisch begründet, da das Archivieren von und Suchen in extrem großen Datenmengen (bis in den TeraByte Bereich) völlig neue Anforderungen an die Systeme stellt. Aber auch der Wert der Ware „Klimadaten“ gewinnt immer stärker an Bedeutung und somit wird – besonders in Europa – der Zugang zu den Klimadaten häufig eingeschränkt. Das ist beispielsweise in den USA insofern anders, als daß dort der freie Zugriff auf Meßwerte als ein Grundrecht angesehen wird. Erst die Darstellung, Visualisierung und Interpretation der Daten wird als Ware bzw. Handelswert betrachtet.

Zur Erfassung der im WWW verfügbaren Klima- und Global-Change-Informationen wurde die Harvest-Software ausgewählt und installiert (<http://www.dkrz.de/harvest/>). Es hat sich gezeigt, daß die Konfiguration eines themenbezogenen Suchsystems, welches unstrukturierte Volltextinformationen von unterschiedliche Quellen erschließen soll, ständige Kontrollarbeiten erfordert. Es müssen beispielsweise Verzeichnisse ausgeschlossen werden, die Informationen zu anderen Themen (z.B. Serverstatistiken) liefern. Dies ist bei einem eigenen Server mit einer bekannten Struktur leicht durchführbar, auf fremden Servern jedoch nur eingeschränkt möglich.

Solche Anpassungsarbeiten lassen sich reduzieren, wenn die Aufbereitung der Daten jeweils vor Ort erfolgt. Dabei werden ebenfalls erheblich bessere Ergebnisse erzielt, da der Datenerzeuger die Struktur der bereitgestellten Daten besser kennt und dies bei der Konfiguration berücksichtigen kann. Hierzu bietet sich die Harvest-Software an, sie kann verteilt installiert, konfiguriert und betrieben werden. Eine Kooperation findet mit dem LOTSE-Projekt des GKSS-Forschungszentrums statt. Dort sind Harvest-Komponenten vor Ort installiert, die mit den Komponenten beim DKRZ Daten austauschen.

Die Aufnahme von Dokumenten im Data-Interchange-Format zeigte, das sich strukturierte Informationen unproblematisch in das Angebot integrieren lassen. Bei der Vorgabe, ein einfach zu verwendendes Interface anzubieten, geht jedoch die Möglichkeit verloren, eine Selektion anhand der speziellen Datenstruktur durchzuführen. Daher ist es sinnvoll, verschiedene Zugänge zu den Daten bereitzustellen und es dem Anwender zu überlassen, die zu seinem Wissensstand passende Oberfläche zu wählen.

Aus diesem Grund wurde die Datenbank-WWW Kopplung mit zwei Einstiegsmöglichkeiten ausgestattet, die verschiedene Suchstrategien unterstützen („Search“ [http://www.dkrz.de/cgi-cera/js/cera\\_gateway](http://www.dkrz.de/cgi-cera/js/cera_gateway), „Browse“ [http://www.dkrz.de/cgi-cera/js/browse\\_button](http://www.dkrz.de/cgi-cera/js/browse_button)). Zur Visualisierung der Klimamodelldaten wurde in diese Schnittstelle ein Verfahren zur Erzeugung von Filmen, die als Daten-Preview dienen können, integriert.

Das Klimainformationssystem wurde ergänzt durch Echtzeit Video/Audio im WWW (<http://www.dkrz.de/dkrz/kommunikation/svideo/svideo.html>). Die DKRZ-Filme „Klimasimulation – Vorhersage des globalen Wandels“ (<http://www.dkrz.de/dkrz/kommunikation/svideo/klimasimulation.html>) und der zur Weltklimakonferenz in Kyoto im Dezember 1997 produzierte Film „Facing Climate Change – Long Term Strategies“ (<http://www.dkrz.de/dkrz/kommunikation/svideo/kyoto.html>) sowie ein Vortrag von Prof. Dr. Graßl ([http://www.dkrz.de/dkrz/kommunikation/svideo/tdf\\_grassl.html](http://www.dkrz.de/dkrz/kommunikation/svideo/tdf_grassl.html)) wurden für das „Streaming“-Verfahren fertiggestellt. Es wurden Varianten für die Software der Firmen Vosaic und RealNetworks erstellt.

Die zur Zeit (am DKRZ) verfügbare Audio/Video-Hard- und Software läßt noch viele Wünsche offen. Angefangen bei den Problemen der Digitalisierung, den Schwierigkeiten bei der Synchronisation, den Limitierungen durch die Bandbreite und der Leistungsfähigkeit der Client-Rechner lassen sich noch viele Schwachstellen nennen. Langfristig erfordert ein professionelles Angebot an digitalen, „streaming“-fähigen Videos eine sehr leistungsfähige Digitalisierungs-Hardware und eine gute Videoschnitt- und Editierungssoftware. Für ein übersichtliches Angebot im WWW wäre es außerdem wünschenswert, wenn die Serversoftware Möglichkeiten zuließe, auf Teilbereiche eines „Streams“ zu linken, d.h. Einsprünge zu definieren.

Wenn man sich jedoch klar macht, daß ein Computer (noch) kein Fernseher ist, eröffnet das „Streaming“-Video für das World Wide Web eine neue Dimension im Hinblick auf eine multimediale Informationsbereitstellung.

Die Generierung der Videos ist mit einem hohen Arbeitsaufwand verbunden, unter anderem auf Grund der am DKRZ zur Verfügung stehenden Tools. Es können praktisch keine Arbeitsgänge automatisiert werden, da die Werkzeuge auf unterschiedlichen Systemen laufen, die kein einheitliches Zugreifen von einem Arbeitsplatz aus zulassen. Außerdem können die Werkzeuge nicht im Hintergrund betrieben werden.

Trotz der Verkürzung der Projektlaufzeit konnte das Projekt erfolgreich abgeschlossen werden. Das Klimainformationssystem sowie das Angebot an Klima-Videos wurden etabliert und gut angenommen.

## A Akronyme

ADL	Alexandria Digital Library
AFTP	Anonymous FTP (File Transfer Protocol)
ASCII	American Standard Code for Information Interchange
AVI	Audio Visual Interleaved
AWI	Alfred-Wegener-Institut für Polar- und Meeresforschung
BMBF	Bundesministerium für Wissenschaft, Bildung, Forschung u. Technologie
BLOB	Binary Large Object
CBS	Commission for Basic Systems
CERA	Climate and Environmental Data Retrieval and Archive
CGI	Common Gateway Interface
DAAC	Distributed Active Archive Center
DDC	Data Distribution Centre
DIF	Directory Interchange Format
DFN	Deutsches Forschungsnetz
DKRZ	Deutsches Klimarechenzentrum
DLR	Deutsche Forschungsanstalt fuer Luft- und Raumfahrt
EOSDIS	Earth Observing System Data and Information System
FGDC	Federal Geographic Data Committee
GCM	Global Climate Model
GCMD	Global Change Master Directory
GIF	Graphics Interchange Format
GILS	Government Information Locator Service
GKSS	Gesellschaft zur Erforschung von Kernbrennstoffen in Schiffbau und Schiffahrt
GrADS	Grid Analysis and Display System
GRIB	Gridded Binary
HTML	Hypertext Markup Language
IMS	Information Management System
IPCC	Intergovernmental Panel on Climate Change
ISIS	Interactive Satellite Image Server
JPEG	Joint Photographic Experts Group
KLIIN	(Multimediale) Klimainformationen im Internet
LOTSE	Land/Ocean Thematic Search Engine
MARC	Machine-Readable Cataloging
MEL	Master Environmental Library
MPEG	Moving Picture Experts Group
NASA	National Aeronautics and Space Administration
NODC	U.S. National Oceanographic Data Center
Perl	Practical Extraction and Report Language
RTP	Real Time Protocol
SFEP	South Florida Ecosystem Program, Datenbank mit Metadaten
SOIF	Summary Object Interchange Format
SQL	Structured Query Language
SSEC	Space Science and Engineering Center

---

URL	Uniform Resource Locator
VDP	Video Datagram Protocol
VRML	Virtual Reality Markup Language
WAIS	Wide Area Information Server
WMO	World Meteorological Organization
WWW	World Wide Web