

## 26 Bewertung des erreichten Standes und Ausblick

Kap. 26 dient zur Bewertung des erreichten Standes und zum Ausblick auf die darauf aufbauende künftige Ausweitung der Datenerschließung. Kap. 26.1 bewertet den im Vergleich zur Ausgangssituation erzielten Nutzen für das Potsdam-Institut für Klimafolgenforschung; in Kap. 26.2 wird die gewählte iterative Vorgehensweise als entscheidender Faktor für das Erreichen der vorgegebenen Ziele eingestuft. Kap. 26.3 gibt einen Ausblick auf die im Rahmen eines dreijährigen TOPIK-Projektes vorgesehene Ausweitung auf gegitterte und polygonverortete Zeitreihen.

### 26.1 Erzielter Nutzen für das Institut

Der im Vergleich zur Ausgangssituation erzielte Nutzen für das Potsdam-Institut für Klimafolgenforschung kann anhand der Behebung der Ausgangsdefizite (Kap. 26.1.1) sowie einer verbesserten Ausgangsbasis für künftige Kooperationen (Kap. 26.1.2) und für künftige Erweiterungen (Kap. 26.1.3) beschrieben werden.

#### 26.1.1 Behebung der Ausgangsdefizite

Die durchgeführte Arbeit kann bezogen auf die vorgegebenen Ziele durchgängig als Erfolg gewertet werden. Im Vergleich mit den vorab festgestellten Ausgangsdefiziten bei der Erschließung der ausgewählten Datenräume durch individuelle Anwender (vgl. Kap. 8.3) konnten in allen Aspekten deutliche Verbesserungen erzielt werden:

- |                                       |  |
|---------------------------------------|--|
| Integrierte Datenbasis                | <ul style="list-style-type: none"> <li>▶ Während zu Beginn dieser Arbeit die Auswertung der im Potsdam-Institut für Klimafolgenforschung vorgehaltenen Zeitreihenmetadaten und punkterorteten Zeitreihen durch deren Heterogenität erschwert war, steht den Anwendern über die Schnittstelle heute der Zugriff auf eine integrierte Datenbasis offen. Zeitreihenmetadaten zur Beschreibung von Stationen, die sich zuvor in getrennten Datenräumen befanden, können nun gemeinsam ausgewertet und selektiert werden. Ferner ist nun ein direkter und flexibler Online-Durchgriff von Zeitreihenmetadaten auf sämtliche durch diese beschriebenen und im Institut vorgehaltenen Zeitreihen möglich, die so in komfortabler Weise bereitgestellt werden können.</li> </ul> |
| Verbesserte Orientierung der Anwender | <ul style="list-style-type: none"> <li>▶ In direktem Zusammenhang hiermit sind die deutlich verbesserten Möglichkeiten zur Orientierung über die verfügbaren Datenbestände zu nennen. Über die Schnittstelle besteht sowohl Zugriff auf die allgemeine Metadatenbank PIK CERA-2, die über unterschiedlichste Datenressourcen informiert, wie auf alle relevanten Datenbanken des Institutes, die zur Dokumentation von Zeitreihenmetadaten verwendet werden.</li> </ul>  |
| Autonomie der Anwender                | <ul style="list-style-type: none"> <li>▶ Individuelle Anwender können nun über die entstandene Schnittstelle autonom, schnell, komfortabel und flexibel auf diese Datenbestände zugreifen. Der bereitgestellte Funktionsumfang unterstützt einen effizienten und anforderungsgerechten Zugang zu den Daten ohne Mithilfe dritter Personen und kann sowohl orts- wie zeitunabhängig durchgeführt werden. Die Bedienung der Schnittstelle kann zudem leicht erlernt und angewendet werden; neu eingebundene Datenräume können ohne zusätzlichen Lernaufwand des Anwenders einbezogen werden.</li> </ul>  |
| Akzeptanz der Anwender                | <ul style="list-style-type: none"> <li>▶ Die entstandene Lösung wird von den Wissenschaftlern des Institutes angenommen. Sämtliche im Rahmen dieser Arbeit adressierten Datenräume wurden von den jeweiligen Datenbereitstellern für einen Zugriff über die Schnittstelle zur Verfügung gestellt; diese wird von Wissenschaftlern unterschiedlicher Disziplinen für die individuelle Datenversor-</li> </ul>   |

gung genutzt. Die als unabdingbar für einen Erfolg der durchgeführten Entwicklung eingestufte Akzeptanz durch die jeweils heterogenen Gruppen der Datennutzer und Datenbereitsteller (vgl. Kap. 9.2) konnte damit für beide Gruppen erreicht werden.

- Erweiterbare Datenbasis
- ▶ Der Datenbestand des Institutes ist nicht statisch, sondern wird durch die Zusammenführung neuer Daten aus externen Quellen beständig erweitert. Dem Institut steht hierfür heute eine Infrastruktur zur Verfügung, in die neue Metadatenätze zur Beschreibung beliebiger Datenressourcen (PIK CERA-2) wie zur Beschreibung von Stationen (Zeitreihenmetadatenbanken und Zeitreihendatenbanken) ebenso aufgenommen werden können wie punktverortete Zeitreihen. Solche neu hinzugefügten Datenressourcen können den Anwendern nun schnell und unaufwendig gemeinsam mit den bereits vorhandenen Daten zugänglich gemacht werden.

### 26.1.2 Verbesserte Ausgangsbasis für Kooperationen

Zugleich ist auf diese Weise eine in mehrfacher Hinsicht verbesserte Ausgangsbasis für künftige Kooperationen des Institutes entstanden. Hierunter fallen:

- Verbesserte Möglichkeiten zur unaufwendigen institutsexternen Datenbereitstellung
- ▶ Der am PIK zusammengeführte Pool von Zeitreihenmetadaten und punktverorteten Zeitreihen hoher Qualität - oder Teilausschnitte aus diesen - kann im Rahmen von Forschungsk Kooperationen unaufwendig und potentiell weltweit mit dem realisierten Funktionsumfang verfügbar gemacht werden. Aufgrund der Konfigurationsmöglichkeiten der Schnittstelle können dabei ohne Reprogrammierung auch externe Zugriffsmöglichkeiten bereitgestellt werden, die sich von der PIK-internen Nutzung unterscheiden, bspw. durch Einschränkung auf bestimmte Datenräume und Attribute. Während sich die Installation einzelner entfernter Clients auf Applikationsbasis, bspw. in Jena, Reykjavik/Island und Norwich/UK auf unterschiedlichen Plattformen in kurzer Zeit problemlos durchführen ließ, ist allerdings für eine weiter gefächerte Bereitstellung im Interesse einer problemlosen Wartung und Aktualisierung der Clients entweder ihre Bereitstellung als Applet oder als Applikation unter Nutzung der mittlerweile verfügbaren Technologie Java Web Start<sup>356</sup> vorzuziehen.
- Verbesserte Möglichkeiten zur unaufwendigen Einbindung externer
- ▶ In umgekehrter Weise können in die Schnittstelle in kurzer Zeit und ohne zusätzlichen Programmieraufwand bei Bedarf potentiell beliebig verteilte Datenbanken von Kooperationspartnern des Institutes eingebunden<sup>357</sup> und so dessen Mitarbeitern zugänglich gemacht werden. Als

<sup>356</sup> Java Web Start ist ein von Sun entwickeltes Konzept, das es erlaubt, Java-Applikationen über das World Wide Web bereitzustellen. Applikationen können durch Aufruf einer HTML-Seite angefordert werden, werden jedoch außerhalb des Web-Browsers interpretiert. Unterstützt werden dabei u.a. Vorgaben über die zu verwendende Virtuelle Maschine, deren automatische Bereitstellung, lokales Vorhalten und transparente Aktualisierung der Applikation im Bedarfsfall oder die Integration der Applikation in den Desktop der Anwendermaschine. Applikationen können dabei ähnlich wie Applets sowohl mit limitierten Zugriffsrechten oder mit erweiterten Rechten betrieben werden. Ferner können auch Applikationen mit limitierten Rechten den Anwender auffordern, die Erlaubnis zum Öffnen einer lokalen Datei oder zum Speichern einer Datei auf dessen Rechner zu erteilen [Sun JWS 2002ab].

<sup>357</sup> Dies ist ohne Reprogrammierung der Schnittstelle möglich, sofern die einzubindenden Ressourcen die zugrundegelegten Anforderungen - Bereitstellung über ein via Java Database Connectivity (JDBC) ansprechbares relationales Datenbankmanagementsystem sowie Bereitstellung jedes Datenraumes in Form einer Datenbanktafel (vgl. Kap. 13.5 und 13.6) - erfüllen. Eine gegebenenfalls für externe Quellen erforderliche Lockerung dieser

- Datenbanken
- Vorteil ist hierbei zu werten, dass eine derartige Bereitstellung verteilter Ressourcen dem Anwender keinen zusätzlichen Aufwand abverlangt, sondern im Gegenteil von den bereits vertrauten Konzepten bspw. der Anfrageerstellung über Filtermodule profitieren kann. Im Rahmen der Entwicklung wurde die prinzipielle Durchführbarkeit einer gleichzeitigen Einbindung lokal vorgehaltener und entfernter Datenräume exemplarisch anhand unterschiedlicher Ausschnitte einer CERA-1-Metadatenbank des Deutschen Klimarechenzentrums (DKRZ) in Hamburg ausgewertet, die jeweils problemlos durch entsprechende Rekonfiguration der Schnittstelle eingebunden werden konnten. Die Einbindung der CERA-2 Metadatenbank des DKRZ / MPIfM (Deutsches Klimarechenzentrum / Max Planck-Institut für Meteorologie) als zweite allgemeine Metadatenbank neben PIK CERA-2 ist für die nahe Zukunft vorgesehen; ein weiteres potentiell Einsatzgebiet stellt die gegenwärtig diskutierte Nutzung der Schnittstelle für einen wechselseitigen Zugriff auf Datenressourcen im PIK und dem britischen Tyndall Centre dar (vgl. Kap. 24.3).
- Verbesserte Ausgangslage für die Beteiligung an Grids
- ▶ Mit dem Konzept der Grids wird ein auf dem Internet aufsetzender, transparenter Zugriff auf verteilte und kostspielige Daten- und IT-Ressourcen angestrebt (vgl. Kap. 5.3). Da dem Institut nun ein integrierter Bestand punktverorteter Zeitreihen zur Verfügung steht, kann dieser künftig - etwa gemeinsam mit am Institut entwickelten Simulationsmodellen - als Datenressource im Rahmen von Kooperationen zum Aufbau eines Grid bereitgestellt werden. Eine solche perspektivische Nutzung der Institutsdatenbestände kann nun mit weitaus weniger Aufwand erfolgen, als wenn hierfür zunächst - wie bei Beginn der Arbeit gegeben - verschiedene heterogene Datenbanken des Institutes zusammengeführt werden müssten. Das PIK besitzt damit eine gute Ausgangslage, um künftig durch Nutzung von Grids an multi-institutionalen „virtuellen“ Organisationen (vgl. Kap. 5.3.4) zur Lösung komplexer Forschungsaufgaben zu partizipieren.

### 26.1.3 Offenheit für weiterführende Konzepte

Durch das Entstehen einer integrierten Datenbasis sowie einer modularen und erweiterbaren Software zum Zugriff auf diese ist zudem nun eine gute Basis für eine beständige Erweiterung der bislang realisierten Funktionalität sowie eine künftige Integration weiterführender Konzepte gegeben. Zu nennen sind hier insbesondere:

- Vereinfachte Kommunikation mit den Anwendern
- ▶ Durch die über die Schnittstelle demonstrierten und von den Anwendern real genutzten Möglichkeiten ist nun eine verbesserte Kommunikationsbasis mit den Wissenschaftlern über Erweiterungen der Funktionalität und die Ausweitung auf weitere Datenräume gegeben. Dabei reduziert sich für die Beteiligten die Komplexität der zu behandelnden Themen, da nun auf vorhandene und den Anwendern bereits vertraute Komponenten und Funktionalität Bezug genommen werden kann. Zudem treten nun einzelne Wissenschaftler verstärkt mit eigenen wertvollen Vorschlägen an die Scientific Data Management Group heran; so wurde bspw. die Visualisierung von Zeitreihen über spezielle Diagramme, die vom Autor als eine der Nachnutzungen entwickelt wurde

(vgl. Kap. 25.2), auf diese Weise angeregt.

- Basis für schnelle Prototypisierung
- ▶ Auf der Basis der entstandenen modular entworfenen Software können schnell Prototypen für neue Anwendungen erstellt werden. Eine besondere Rolle spielt hierbei immer wieder der interaktive digitale Atlas IDA (vgl. die in Kap. 25 beschriebenen Nachnutzungen) mit seiner Doppelfunktion zur Raumauswahl und zur raumbezogenen Visualisierung. Letztere wurde zudem prototypisch auf die Kartengenerierung aus Zeitreihen auf regelmäßigen Gittern sowie aus polygonverorteten Zeitreihen ausgeweitet, so dass für entsprechende Prototypen jeweils ansprechende interaktive Formen der Datendarstellung bereitgestellt werden konnten (vgl. Kap. 26.3). Solche Möglichkeiten zur schnellen Prototypisierung unterstützen die realistische Bewertung von Lösungsalternativen für künftige Entwicklungen, da sie eine frühzeitige Einbeziehung der Anwender und ihres Fachwissens erlauben.
- Möglichkeiten zur Erweiterung der Schnittstellenfunktionalität
- ▶ Aus Rückmeldungen der Anwender wird deutlich, dass die Schnittstelle - unabhängig vom jeweils realisierten Funktionsumfang - beständig weiter verbessert werden kann. So wurde in jüngerer Zeit vermehrt Bedarf nach zwei konkreten Erweiterungen geäußert. Dabei soll einerseits die Selektion von Zeitreihenmetadaten durch die Vorgabe einer beliebigen Liste von Stationsnamen oder Stationsidentifikatoren ermöglicht werden; zum anderen soll die Zeitreihenvisualisierung auch den Vergleich von Werten jeweils eines nutzerdefinierbaren Monats über einen längeren Zeitraum hinweg ermöglichen<sup>358</sup>. Solche und ähnliche Erweiterungen, die für die nahe Zukunft vorgesehen sind, können durch gezielte und vergleichsweise unaufwendige Eingriffe, die sich auf die hier von betroffenen Komponenten beschränken - in diesem Fall das Filtermodul `SingleAttributeFilter` (vgl. Kap. 20.2) bzw. die beiden Auswertungsmodule `TimeSeriesSelector` (vgl. Kap. 22.3.3) und `TimeSeriesVisualizer` (vgl. Kap. 22.3.4) - vorgenommen werden.
- Möglichkeiten zur Erweiterung um Konzepte der virtuellen Integration
- ▶ Zur Integration heterogener Datenbestände stehen virtuelle und materielle Konzepte zur Verfügung (vgl. Kap. 2.4.1). Während die von der Scientific Data Management Group durchgeführte Homogenisierung durch den Aufbau eines Data Warehouse für Zeitreihen sowie von integrierten Zeitreihenmetadatenbanken und Zeitreihendatenbanken (vgl. Kap. 18.1.2), also durch eine *materielle* Integration neben den Originalbeständen, realisiert wurde, ist ein solches Vorgehen bei einer perspektivischen Ausweitung auf institutsexterne Datenräume nicht mehr ohne weiteres durchführbar. Um hier dennoch einen integrierten, homogenen Zugang zu heterogene Quellen zu erreichen, kann im Bedarfsfall auf Konzepte der *virtuellen* Integration wie Mediatoren (vgl. Kap. 2.4.2) oder föderierte Datenbanksysteme (vgl. Kap. 2.4.4) zurückgegriffen werden. Der modulare Aufbau der Schnittstelle erlaubt es, entsprechende Funktionalitäten zu integrieren, ohne hierfür wesentliche Teile der Software verändern zu müssen; damit könnten den Anwendern künftig auch heterogene Quellen, die nicht vorab physisch zusammengeführt werden können, wie *ein* homogener Datenraum zugänglich gemacht werden.

<sup>358</sup> Auf diese Weise wird es beispielsweise möglich, Niederschlags- und Temperaturwerte im Monat April für die Jahre 1950 bis 2000 miteinander zu vergleichen. Eine derartige Funktionalität wird bereits von der Komponente für die Visualisierung von Zeitreihen mit dem Applet `InteractiveScenarioVisualizer` (vgl. Kap. 25.3) bereitgestellt.

Durch den Einsatz von Mediatoren könnten hier neben Daten aus Datenbanken auch semi- oder unstrukturierte Daten (vgl. Kap. 1.2) einbezogen werden.

Möglichkeiten zur Erweiterung des visuellen Zugangs zu den Daten

- ▶ Konzepte zur Visualisierung von Daten (vgl. Kap. 4) wurden mehrfach für die gewählte Lösung eingesetzt<sup>359</sup>. Dabei wurden bisher ausschließlich „klassische“, der Mehrheit der Anwender vertraute und damit leicht interpretierbare Formen der Darstellung gewählt. Da die Schnittstelle offen für nachfolgende Erweiterungen sowohl von Selektions- und Auswertungsfunktionalität ist, ist hier der künftige Einsatz weiterführender Konzepte in mehrfacher Weise möglich. So könnten etwa in einem Filtermodul wie dem HierarchyBrowser (vgl. Kap. 20.7) zur graphisch-interaktiven Navigation durch die Baumstruktur eines Thesaurus anstelle der eingesetzten, traditionellen Darstellung auch Konzepte der Informationsvisualisierung wie ConeTrees (vgl. Kap. 4.8.1) oder Hyperbolic Browser (vgl. Kap. 4.8.3) zum Einsatz kommen. In gleicher Weise könnten bei Bedarf bestehende Auswertungsmodule auf erweiterte Konzepte der Visualisierung umgestellt oder zusätzliche Auswertungsmodule, die diese anbieten, eingebunden werden. Beispiele für solche möglichen Erweiterungen sind die Einbeziehung von Techniken zur Visualisierung multivariater Daten (vgl. Kap. 4.6), der Einsatz dynamischer Darstellungen (vgl. Kap. 4.3.3) oder die Anwendung von Konzepten der Informationsvisualisierung (vgl. Kap. 4.8). Da die Zahl der Filter- wie der Auswertungsmodule erweiterbar ist, können neue Varianten dabei zusätzlich zu den den Anwendern bereits vertrauten Formen des Zugangs angeboten werden. Auf diese Weise kann die Akzeptanz einzelner Darstellungen im laufenden Betrieb erprobt werden, ohne den Anwendern einen Wechsel ihrer gewohnten Arbeitsweise abzuverlangen; ebenso können alternative Varianten - etwa mehrere visuelle Zugänge zum Thesaurus von PIK CERA-2 - auch auf Dauer bereitgestellt und von einzelnen Anwendern je nach Vorliebe ausgewählt werden.

Möglichkeiten zur Erweiterung auf Methoden des Data Mining

- ▶ Unter dem Begriff Data Mining (vgl. Kap. 3) werden Konzepte zum Auffinden zuvor unbekannter Muster in Daten zusammengefasst. Während derartige Ansätze aufgrund ihrer speziellen Ausrichtung für die hier angestrebte Lösung nicht relevant waren und daher entsprechend nicht berücksichtigt wurden, besteht nun durch die entstandene integrierte Datenbasis an punktverorteten Zeitreihen eine verbesserte Ausgangslage auch für die Einbeziehung solcher Ansätze. Vor dem Hintergrund des hohen Aufwandanteils, den Datenaufbereitung und -bereitstellung für Data Mining-Projekte ausmachen (vgl. Kap. 3.5.1), ist insbesondere als Vorteil zu werten, dass für diesen Datenbestand des Institutes keine aufwendigen Zusammenführungs- und Homogenisierungsprozesse vor Beginn von Auswertungen - sei es durch Data Mining oder andere Werkzeuge wie bspw. spezielle Visualisierungs- oder Statistiksoftware - mehr erforderlich sind. Ausschnitte solcher Funktionalität könnten bei Bedarf zudem durch die Realisierung entsprechende Filter- und Aus-

<sup>359</sup> Dies umfasst sowohl die Auswahl von Raumbezügen anhand ihrer graphischen Darstellung (vgl. Kap. 20.3) sowie die Präsentation von Ergebnismengen aus allgemeinen Metadaten (vgl. Kap. 21.1.5), Zeitreihenmetadaten (vgl. Kap. 21.2.4) und Zeitreihen (vgl. Kap. 22.3.4).

wertungsmodule künftig auch direkt in die Schnittstelle integriert werden.

- Möglichkeiten zur Personalisierung der Schnittstellen-Clients
- ▶ Die konkrete Ausgestaltung der Schnittstellen-Clients wird zur erst Laufzeit durch Auswertung der Konfigurationsinformationen und die darauf basierende Generierung der entsprechenden Oberfläche festgelegt (vgl. Kap. 14.3). Während dies gegenwärtig - mit Ausnahme institutsexterner Clients - für alle Anwender in gleicher Weise geschieht, erlaubt diese Konzeption auch eine Bereitstellung unterschiedlicher Clients für spezifische Anwenderanforderungen. Solche „maßgeschneiderten“ Angebote für einzelne Nutzergruppen oder konkrete Anwender könnten etwa in der Vorauswahl der jeweils bevorzugten Datenbank und einer Vorbelegung von Filtermodulen entsprechend häufig eingesetzter Kriterien bestehen. Zur Ableitung der hierfür erforderlichen Informationen können einerseits automatische Prozesse - von der einfachen Speicherung und Wiederherstellung individueller Nutzerprofile bis hin zur Auswertung der bei der Nutzung der Schnittstelle anfallenden Sekundärdaten<sup>360</sup> - eingesetzt werden; ebenso können künftig Möglichkeiten zu einer manuellen Konfiguration des Client durch den Anwender integriert werden.

#### 26.1.4 Fazit

Zusammenfassend kann festgehalten werden, dass bezogen auf sämtliche adressierten Datenräume die anvisierten erforderlichen Verbesserungen erreicht werden konnten. Das Potsdam-Institut für Klimafolgenforschung verfügt heute über eine von den Anwendern akzeptierte und zur Unterstützung der täglichen Arbeitsprozesse genutzte Infrastruktur zur komfortablen Bereitstellung der dort vorgehaltenen allgemeinen Metadaten, Zeitreihenmetadaten und punktverorteten Zeitreihen. Zudem wurde eine günstige Ausgangsbasis für weitere diesbezügliche Entwicklungen am Institut geschaffen.

### 26.2 Bewertung der gewählten Vorgehensweise

Das zugrundegelegte *iterative Vorgehen* bei der Realisierung der Schnittstelle konnte wesentlich dazu beitragen, die aufgrund der gegebenen Komplexität von Kontext, Anwendern und Anforderungen bestehenden Verständnisschwierigkeiten schrittweise abzubauen. So stellte sich während des Entwicklungsprozesses immer wieder heraus, dass einzelne Fragestellungen weniger gut anhand theoretischer Erörterungen, sondern weitaus besser anhand von Prototypen, die einen realen Eindruck von der Umsetzung angedachter Funktionalität vermitteln, geklärt werden konnten. Ferner konnten auf diese Weise während der Erstellung der Software kontinuierlich Anwender einbezogen und auf diese Weise wertvolle Hinweise auf geeignete Formen der Umsetzung einzelner Aspekte von Funktionalität und Gestaltung gewonnen werden. Da die Konzeption der Schnittstelle ihre iterative Fortentwicklung unterstützte, war es möglich, basierend auf Rückmeldungen der Anwender mit vergleichsweise geringem Aufwand gezielte Verbesserungen an einzelnen Komponenten vorzunehmen und so bspw. die generelle Nutzerführung oder die Ausgestaltung einzelner graphischer Oberflächen sukzessive zu verbessern und den Wünschen und

<sup>360</sup> Daten über die Nutzung der Schnittstelle werden von dieser in einer in einer speziellen Log-Datenbank dokumentiert. Hierbei werden Informationen über die verwendeten Filtermodule, die Verwendung des Hilfesystems etc. in anonymisierbarer Form aufgezeichnet. Dieses automatische „Mitschreiben“ wurde bereits zu Beginn von Betriebsphase I eingeführt, um sowohl zu Rückschlüssen über die Art der Nutzung der Schnittstelle zu gelangen (vgl. Kap. 24.2) wie gegebenenfalls auftretende Fehler schnell rekonstruieren und beheben zu können. Diese Daten könnten - in gewisser Analogie zur Optimierung von Websites durch Web Usage Mining (vgl. Kap. 3.4.4) - in anonymisierter Form ausgewertet und für eine weitere Verbesserung der allgemeinen Nutzerführung sowie für die Personalisierung der Schnittstelle genutzt werden.

Bedürfnissen der Anwender anzunähern.

Die gewählte Vorgehensweise einer iterativen Realisierung der Schnittstelle kann damit als *entscheidender Faktor* für das Erreichen der vorgegebenen Ziele angesehen werden. Ein vergleichbares Vorgehensmodell wird daher auch der im Rahmen eines neuen, 2004 beginnenden TOPIK-Projektes verfolgten Ausweitung auf Datenräume mit gegitterten sowie polygonverorteten Zeitreihen zugrundegelegt werden.

### **26.3 Ausblick: Ausweitung auf gegitterte und polygonverortete Zeitreihen**

Die am PIK betriebene disziplinübergreifende Erdsystemanalyse basiert in hohem Maße auf computergestützter Modellierung und erfordert daher einen möglichst effizienten und anforderungsgetriebenen Zugriff auf die hierfür relevanten, komplexen Datenräume. Für wesentliche Ausschnitte dieser Datenräume - allgemeine Metadaten, punktverortete Zeitreihen und die zu deren Erschließung erforderlichen Zeitreihenmetadaten - konnten im Rahmen dieser Arbeit in enger Zusammenarbeit mit der Scientific Data Management Group des PIK Lösungen entwickelt werden, die sich als geeignet und erfolgreich erwiesen, so dass nun ein integrierter und funktionaler Zugriff auf einen hochqualitativen Datenpool von über die gesamte Erde verteilten Stationen besteht. Für zwei weitere, ebenfalls für die Modellierung bedeutsame Gruppen von Datenräumen - Zeitreihen auf regelmäßigen Gitterstrukturen sowie Zeitreihen auf Polygonstrukturen - stellt sich die Situation bis heute anders dar. Für diese Datenräume besteht am Institut noch keine vergleichbare Infrastruktur, die ein effizientes zentralisiertes Datenmanagement mit geeigneten internetbasierten, adaptiven, funktionalen und intuitiven Zugriffs- und Visualisierungsschnittstellen verbindet. Entsprechend verlangt hier die Erschließung relevanter Datenräume ebenso wie die Bereitstellung von im Forschungsprozess generierten Ergebnisdaten dem individuellen Wissenschaftler immer noch wiederkehrende, umständliche und zeitaufwendige Prozesse ab.

Daher ist es naheliegend, die für punktverortete Zeitreihen erreichte Funktionalität nun auch auf diese Datenräume auszudehnen, die ihrerseits jeweils eigene Charakteristika aufweisen und neue Herausforderungen aufwerfen. Erste vorbereitende Schritte wurden mit auf im Rahmen dieser Arbeit entwickelten Softwarekomponenten basierenden Prototypen unternommen; eine umfassende Lösung wird im Rahmen eines 2004 beginnenden, dreijährigen TOPIK-Projektes entwickelt und umgesetzt werden.

#### **26.3.1 Zeitreihen auf regelmäßigen Gittern**

Unter *Zeitreihen auf regelmäßigen Gittern* (im Folgenden auch kurz als *gegitterte Zeitreihen* bezeichnet) werden dabei solche Zeitreihen verstanden, deren geographischer Bezug jeweils durch eine eindeutige Position auf einem regelmäßigen Gitter beschrieben ist. Hierunter fallen basierend auf gemessenen (punktverorteten) Zeitreihen berechnete Daten ebenso wie Ergebnisse von Simulationen, also Ausgabedaten von Modellen. Ein Datenraum mit gegitterten Zeitreihen kann dabei als regelmäßiger multidimensionaler Datenwürfel aufgefasst werden; typische Dimensionen sind Raumbezug (geographische Breite und Länge der Gitterzellen sowie gegebenenfalls die Höhe über dem Meeresspiegel), Zeitbezug, einzelne Variablen und gegebenenfalls unterschiedliche zeitliche Auflösungen.

Da hier in Abhängigkeit von der Zahl der betrachteten Parameter sowie der räumlichen und zeitlichen Auflösung große Datenvolumen anfallen können, werden diese Zeitreihen aus Effizienzgründen gegenwärtig zumeist nicht in Datenbanken, sondern in Dateien vorgehalten. Ein prominentes Beispiel für eine wichtige Sammlung gegitterter Zeitreihen, die ebenfalls über Dateien zur Verfügung steht und auch vom PIK bisher in dieser Form genutzt wird, ist die von der Climatic

Research Unit (CRU)<sup>361</sup> der University of East Anglia, Norwich, U.K. bereitgestellte Gridded Monthly Meteorology<sup>362</sup>. Hierfür wird die Erde in ein Gitter von 0.5° Länge und 0.5° Breite aufteilt; auf dem so entstehenden Gitter werden für die Landmassen mit Ausnahme der Polargebiete unterschiedliche Zeitreihen mit je einem Wert pro Parameter, Gitterzelle und Monat für einen Zeitraum von gegenwärtig 98 Jahren bereitgestellt.

Durch die bisher aus praktischen Gründen eingesetzte Organisation der Daten über Dateien ist von den einzelnen Anwendern jeweils einzelfallbezogene Arbeit erneut und wiederholt zu leisten. So erfordert etwa die Auswahl von Untermengen oder die Aggregation von Werten in der Regel Programmierfähigkeiten, um entsprechende Datenextraktionen auf den Dateien vorzunehmen; für eine Berechnung statistischer Parameter und nachfolgende Visualisierungen sind entsprechende Werkzeuge einzubeziehen, so dass jeweils eine Vielzahl von Einzelschritten erforderlich ist. Durch eine datenbankgestützte Bereitstellung solcher Daten in Kombination mit geeigneten und weitgehend generischen Auswahl-, Verarbeitungs- und Visualisierungsfunktionen können hier viele dieser Prozesse zentralisiert angeboten und komfortabel zur Verfügung gestellt werden. Die zentrale Herausforderung in diesem Kontext bilden dabei Multidimensionalität und Größe der Datenräume, die sowohl besondere Anforderungen an eine performante Verarbeitung und Bereitstellung wie eine netzwerkfähige interaktive Visualisierung stellen.

### 26.3.2 Sozioökonomische Zeitreihen auf Polygonstrukturen

Neben punktverorteten Zeitreihen und Zeitreihen auf regelmäßigen Gittern bilden *Zeitreihen auf Polygonstrukturen* die dritte relevante Gruppe von Zeitreihen im Kontext der Erdsystemanalyse. Hierbei handelt es sich um solche Zeitreihen, deren geographischer Bezug jeweils durch ein oder mehrere Polygone beschrieben ist. Zu den typischen Vertretern dieser Gruppe zählen wirtschaftliche und soziologische Zeitreihen, die für einzelne Kontinente, Staaten oder andere vordefinierte geographische Einheiten erhoben werden; entsprechende Daten werden im PIK primär von Wissenschaftlern der Abteilung Global Change and Social Systems benötigt.

Eine geeignete Erschließung dieser Zeitreihen steht insbesondere vor der Herausforderung der Integration diverser Ressourcen. Anders als die beschriebenen gegitterten Zeitreihen liegen polygonverortete sozioökonomische Zeitreihen in extrem heterogener Form vor. Sie stammen aus einer Vielzahl unterschiedlicher Quellen und werden von den einzelnen Datengebern meist in Form von CD-ROMs oder über proprietäre Web-Zugriffsschnittstellen bereitgestellt. Die dabei angebotenen Nutzerschnittstellen sind durch uneinheitliche und unterschiedlich ausgeprägte Funktionalität gekennzeichnet, die zudem in der Regel den Anforderungen der Wissenschaftler des Institutes nicht genügt. So sind jeweils unterschiedlich funktionale oder auch überhaupt keine Möglichkeiten zur raum- oder zeitbezogenen Visualisierung gegeben; die Auswahlmöglichkeiten zum Export von Daten sind ebenfalls zumeist deutlich eingeschränkt. Weitere Defizite bestehen in fehlenden geeigneten Möglichkeiten zu nutzerdefinierten zeitlichen sowie räumlichen Aggregationen; eine integrierte Nutzung von Daten aus unterschiedlichen Quellen ist zudem nur mit jeweils erheblichem individuellen Aufwand möglich und potentiell fehleranfällig. Eine Zusammenführung sozioökonomischer Zeitreihen von übergreifendem Interesse sowie eine Bereitstellung geeigneter nutzerdefinierbarer Selektions-, Aggregations- und Visualisierungsfunktionalität kann daher die Anwender wesentlich bei der wissenschaftlichen Nutzung dieser Ressourcen unterstützen.

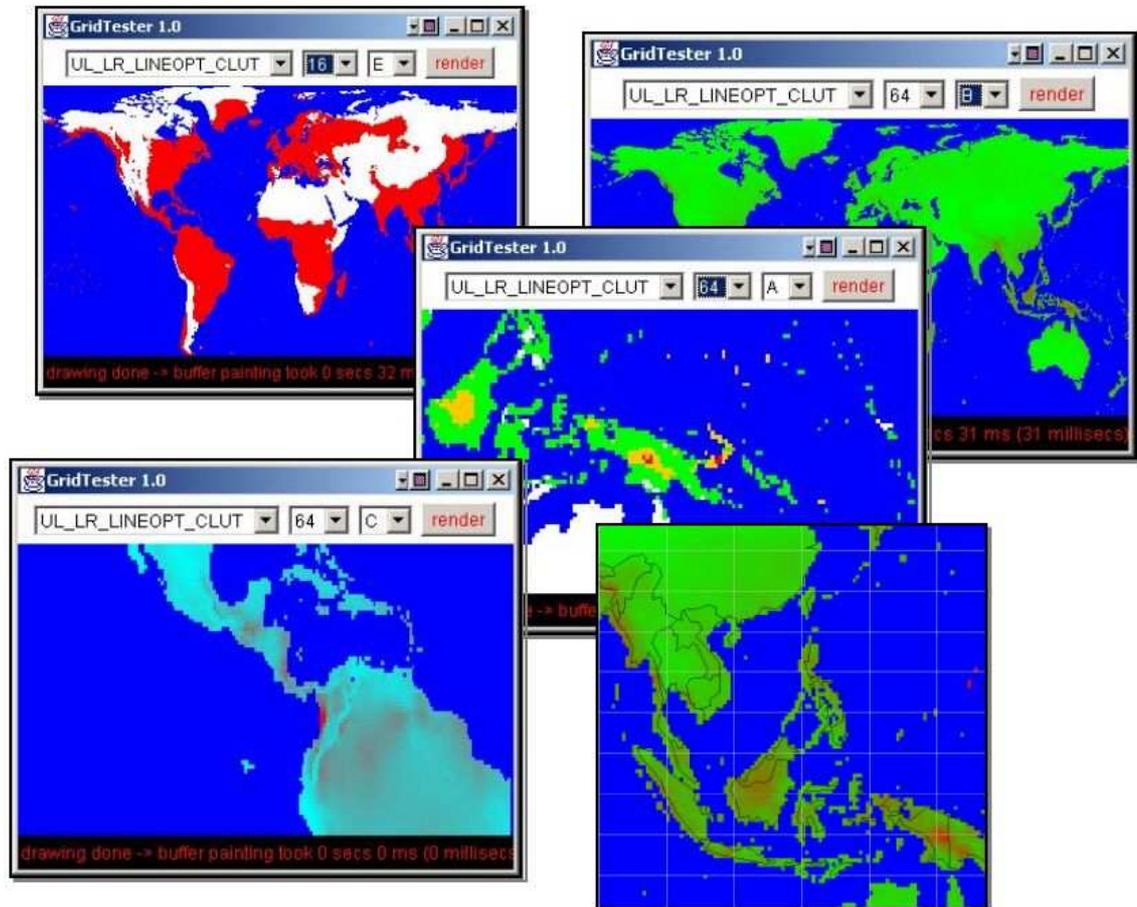
Im Rahmen einer Annäherung an diese komplexen Datenräume wurden basierend auf den bisher entstandenen Komponenten - und hier insbesondere durch Erweiterungen von IDA - zwei Prototypen entwickelt, die eine interaktive Generierung zweidimensionaler Visualisierungen aus gitter- bzw. polygonverorteten Zeitreihen adressieren.

<sup>361</sup> <http://www.cru.uea.ac.uk/>

<sup>362</sup> [http://www.cru.uea.ac.uk/~markn/cru05/cru05\\_intro.html](http://www.cru.uea.ac.uk/~markn/cru05/cru05_intro.html)

### 26.3.3 Prototyp 1 – Gegitterte Zeitreihen

Entsprechend der Multidimensionalität und Komplexität gegitterter Zeitreihen ist für einen intuitiven Zugang der Einsatz geeigneter Formen interaktiver Visualisierung von zentraler Bedeutung. Neben weiteren relevanten Formen wie Diagrammgenerierung oder Volumenvisualisierung<sup>363</sup> stellt die raumbezogene Darstellung anhand zweidimensionaler Karten ein wichtiges Hilfsmittel zur Auswertung dieser Daten dar. Um zu einer Abschätzung der Eignung von IDA für eine web-basierte, interaktive clientseitige Generierung solcher Karten aus gegitterten Zeitreihendaten zu gelangen, wurde ein entsprechender Prototyp entwickelt (vgl. Abb. 26.1).



**Abb. 26.1** - Prototypische interaktive Visualisierung gegitterter Zeitreihen.

Grundlegende Vorteile einer interaktiven Visualisierung, bei der die Ausgangsdaten auf dem Anwenderrechner verarbeitet werden, liegen in den Möglichkeiten zu einer schnellen nutzerdefinierten Veränderung der graphischen Darstellung beispielsweise durch interaktives Zoomen, nutzerdefiniertes Einfärben oder die Hervorhebung von Extrema (Hotspotting). Zusätzlich können so interaktiv Datenwerte durch Zeigen auf Teile der graphischen Darstellung ausgewählt und abgerufen werden; die Kombination einer entsprechenden Visualisierung mit IDA erlaubt zusätzlich den interaktiven Abruf von Koordinatenwerten, Ländernamen etc. und ermöglicht es, die Umrisse nutzerwählbarer administrativer Einheiten oder Flusseinzugsgebiete einzublenden.

Eine wesentliche Herausforderung bei einer solchen anwenderseitigen Visualisierung besteht in

<sup>363</sup> Als Volumenvisualisierung wird die Visualisierung skalarer Daten auf einem regelmäßigen dreidimensionalen Gitter bezeichnet. Sie dient dazu, innere Zusammenhänge in diesen Daten zu veranschaulichen. Als Methoden stehen Dekompositionen der Datenmenge bspw. in Punkte oder Schichten und deren Darstellung, die Extraktion und Darstellung sog. Isoflächen, die die Position eines spezifischen Datenwertes im Würfel repräsentieren sowie die Darstellung halbtransparenter Datenwürfel zur Verfügung (vgl. hierzu ausführlich [Schumann, Müller 2000, 251ff.]).

diesem Kontext in dem zu verarbeitenden Datenvolumen, das jeweils auf den Rechner des Anwenders zu transportieren und dort umzusetzen ist. Als Referenzdatensatz für die prototypische Entwicklung wurden aus der Gridded Monthly Meteorology der Climatic Research Unit Niederschlagswerte für einen Monat ausgewählt. Da diese Daten für ein Gitter von  $0.5^\circ$  Länge und  $0.5^\circ$  Breite vorliegen, bestand der Referenzdatensatz aus insgesamt  $720 \cdot 360 = 259.200$  Datenwerten. Setzt man pro Datenwert nur jeweils ein Byte an, sind zur Kodierung dieser Werte bereits rund 253 Kilobyte erforderlich. Um dieses Volumen zu reduzieren, wurden verschiedene Varianten einer komprimierteren Datenrepräsentation ausgewertet. Durch die Einteilung der Datenwerte in 64 unterschiedliche Klassen, die Reduzierung von Redundanzen sowie Binärcodierung konnte das Datenvolumen des Referenzdatensatzes auf knapp 44.000 Byte (weniger als 43 Kilobyte) reduziert werden.

Liegen diese Daten auf dem Rechner des Anwenders vor, erlaubt der Prototyp typische interaktive Rendering-Operationen (Generierung der Ausgangskarte, Zooming, Navigieren innerhalb von Kartenhierarchien, Abruf von Werten, Einfärben der Werte nach unterschiedlichen Klassen) auf einem Intel-PC mit Pentium 4-Prozessor (2,2 GHz) und 512 MB RAM jeweils ohne nennenswerte Verzögerung. Dieser Prototyp - der als Labormuster eingestuft werden kann - diente im wesentlichen dazu, zu Rückschlüssen über die Durchführbarkeit einer auf IDA basierenden clientseitigen Kartengenerierung mit entsprechendem Feedbackumfang zu gelangen, die als hinreichend performant umsetzbar eingestuft werden kann.

Allerdings ist zu berücksichtigen, dass der prototypisch realisierte Funktionsumfang nur einen der Bestandteile einer zukünftigen umfassenden Anwendung zur Erschließung gegitterter Zeitreihen darstellt. Bereits für die Erweiterung dieser Art von Visualisierung auf die Darstellung mehrerer zeitlich aufeinander folgender Karten in Form eines Filmes steigt das einzubeziehende Datenvolumen entsprechend an, so dass ein Transport dieser Daten zum Client hier schnell zu aufwendig werden und damit eine serverseitige Bilderzeugung sinnvoller sein kann. Sorgfältige Abwägungen sind auch für andere Formen der Visualisierung gegitterter Zeitreihen erforderlich. Während für Zeitreihendiagramme von ähnlichen Voraussetzungen wie bei punktverorteten Zeitreihen ausgegangen werden kann - so dass eine clientseitige Umsetzung mithin als unproblematisch einzustufen ist -, erfordert eine Volumenvisualisierung die Einbeziehung eines hohen Datenvolumens ebenso wie die Verwendung spezieller dreidimensionaler Rendering-Algorithmen, so dass hier verschiedene Varianten einer Client-Server-Aufteilung ebenso zu evaluieren sind wie hierfür verfügbare Visualisierungswerkzeuge.

Schließlich ist zu beachten, dass eine interaktive Bildgenerierung einen vorhergehenden effizienten Zugriff auf die Daten voraussetzt und damit zusätzlich der jeweils erforderliche Zeitaufwand für Datenextraktion, eventuelle räumliche oder zeitliche Aggregationen, Datentransformation sowie Netzwerktransport einzubeziehen ist. Insbesondere ist eine datenbankgestützte Bereitstellung, die es erlaubt, wesentliche Auswahl- und Vorverarbeitungsoperationen auf ein Datenbankmanagementsystem zu verlagern, als attraktiv einzustufen; allerdings ist hierfür die Verwendung geeigneter Modellierungskonzepte, die einen flexiblen und performanten Zugang zu den Daten erlauben, erforderlich.

#### **26.3.4 Prototyp 2 – Zeitreihen auf Polygonstrukturen**

Auch für die intuitive Auswertung polygonverorteter Zeitreihen stellt eine interaktive Generierung entsprechender Karten ein wichtiges Werkzeug dar. Um zu einer verbesserten Ausgangs- und Diskussionsbasis für eine zukünftige integrierte Erschließung sozioökonomischer polygonverorteter Zeitreihen zu gelangen, wurde hierfür ebenfalls ein entsprechender Prototyp entwickelt (vgl. Abb. 26.2). Anders als im Falle der gegitterten Zeitreihen wurde hier eine datenbankbasierte Anwendung konzipiert, die sowohl die interaktive Generierung zweidimensionaler Karten wie von Zeitreihendiagrammen (Säulendiagramme) unterstützt. Als Datenbasis wurde von der Scientific

Data Management Group in Zusammenarbeit mit der Abteilung Global Change and Social Systems eine prototypische Datenbank zur Integration sozioökonomischer Zeitreihen aufgebaut, in der Zeitreihen aus vier unterschiedlichen Quellen<sup>364</sup> zusammengeführt wurden. Ein einheitlicher Raumbezug der Datensätze wurde durch Adaption von hierarchischen räumlichen Klassifikatoren (vgl. Kap. 18.3.1) hergestellt, so dass eine Generierung entsprechender Karten durch automatische Zuordnung von Datenwerten zu den von IDA verarbeiteten Polygonen erfolgen kann, die über entsprechende Algorithmen eingefärbt werden.

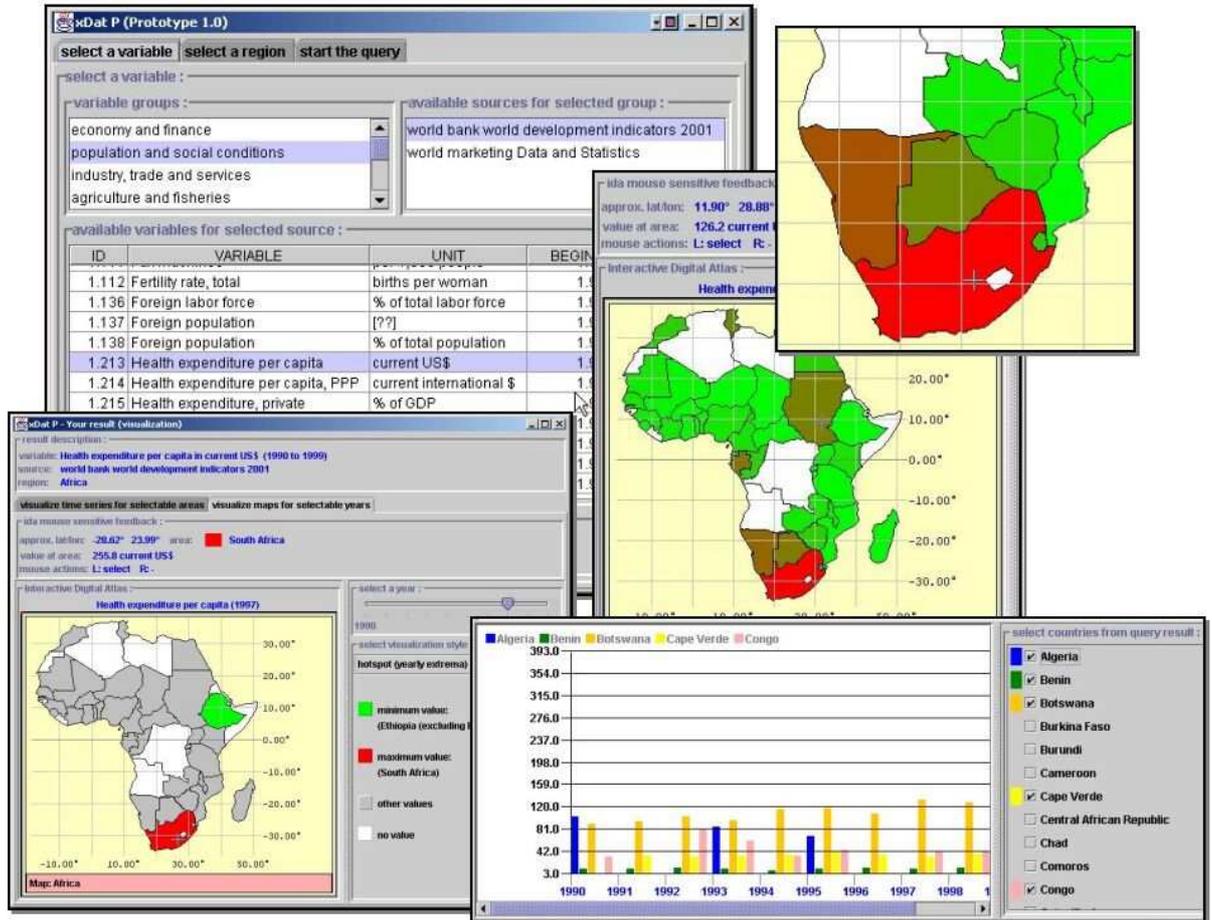


Abb. 26.2 - Der Prototyp zur Erschließung von Zeitreihen auf Polygonstrukturen.

Der Prototyp wurde primär zur Demonstration grundlegender Vorteile einer geeigneten Nutzer-schnittstelle konzipiert und im Interesse einer schnellen Realisierung auf wesentliche, stellvertre-tende Funktionalitäten limitiert. Der Anwender kann zur Auswahl von Zeitreihen aus der Daten-bank jeweils eine Variable, ein Zeitfenster sowie eine Kombination von Staaten definieren. Für die Auswahl des Raumbezuges kann dabei zwischen allen Staaten der Erde, allen Staaten eines Kontinentes oder einem von mehreren Clustern von Staaten<sup>365</sup> gewählt werden, die auf von der Abteilung Global Change and Social Systems entwickelten Einteilungen basieren. Für die inter-aktive raumbezogene Visualisierung wurde IDA entsprechend erweitert, so dass eine dynami-sche Kartengenerierung, ein schnelles Hintereinanderablaufen von Karten für einzelne Jahre („Film“), unterschiedliche Formen der Darstellung wie Hotspotting, Zooming sowie interaktiver Werteabruf unterstützt werden. Zusätzlich wurde eine interaktive Generierung von Säulendia-grammen für einen Vergleich ausgewählter Zeitreihen über die Zeit integriert. Die entstandene Anwendung erlaubt die Umsetzung der clientseitigen Interaktionen jeweils ohne spürbare Verzö-

<sup>364</sup> World bank development indicators 2001, World Marketing Data and Statistics, Carbon Dioxide Information Center ndp30 07/30/2001 sowie Zeitreihen der International Energy Agency.

<sup>365</sup> Die Zusammensetzung der auswählbaren Cluster kann extern über eine Datei konfiguriert werden.

gerung. Der Prototyp wurde Wissenschaftlern der Abteilung Global Change and Social Systems demonstriert und positiv bewertet.

Während eine Erweiterung der Software auf zusätzlich erforderliche Funktionalität als prinzipiell problemlos einzustufen ist, ergeben sich hingegen hinsichtlich der Bereitstellung einer geeigneten *integrierten* Datengrundlage erhebliche Herausforderungen:

- **Heterogenität**

Zeitreihen aus verschiedenen Quellen weisen eine hochgradige Inhomogenität etwa durch verschiedenartige räumliche Zuordnungen, Definitionen, verwendete Maßeinheiten etc. auf und müssen zunächst einer aufwendigen inhaltlichen Homogenisierung unterzogen werden.

- **Thematische Navigation**

Ferner ergab sich bereits bei den einbezogenen vier Datenquellen eine kaum übersehbare Vielfalt unterschiedlicher Variablen. Um Anwendern hier eine effiziente Navigation zu ermöglichen, ist der Einsatz eines geeigneten Thesaurus sowie eine entsprechende Zuordnung der einzelnen Zeitreihen unabdingbar.

- **Räumliche Aggregation**

Um eine räumliche Aggregation durch die dynamische Berechnung neuer Werte für eine nutzerdefinierbare Kombination bspw. von Staaten zu ermöglichen, müssen jeweils geeignete Abbildungsvorschriften für einzelne Variablen entwickelt werden, in die gegebenenfalls zusätzliche Informationen (Fläche der Staaten, Anzahl der Bevölkerung etc.) einzubeziehen sind. Ferner sind geeignete Strategien zur Behandlung von Fehlwerten erforderlich, um die Generierung wissenschaftlich verwertbarer Ergebnisdaten sicherzustellen.

- **Zeitlich variante Raumbezüge**

Nicht zuletzt ist zu berücksichtigen, dass die dokumentierten Raumbezüge von Zeitreihen in diesem Kontext einer zeitlichen Veränderlichkeit unterliegen können – so ist jeweils festzulegen, wie mit einzelnen Zeitreihen von Staaten, die in mehrere Einzelstaaten zerfallen sind (wie die Sowjetunion oder Jugoslawien) sowie von solchen, die wie BRD und DDR zu einem Staat vereinigt wurden, zu verfahren ist.

Die anhand der Prototypen skizzierten Herausforderungen verdeutlichen, dass auch die Entwicklung funktionaler Schnittstellen für einen Zugriff auf gitter- sowie polygonbezogene Zeitreihen nicht nur den Einsatz geeigneter Technologien, sondern zudem unabdingbar die Einbeziehung der Kompetenz von Fachwissenschaftlern aus den entsprechenden Themengebieten erfordert und entsprechend ebenfalls nur in einem interdisziplinären Ansatz zu erreichen ist.

### 26.3.5 Die nächsten drei Jahre – das neue TOPIK-Projekt PixDat

Die im Rahmen dieser Arbeit durchgeführte Entwicklung fand als internes Projekt unter der Schirmherrschaft der Abteilung *Data & Computation* statt. Basierend auf den positiven Erfahrungen mit der Erschließung der heterogenen punktverorteten Zeitreihen des Institutes werden die nächsten Schritte, beginnend im Januar 2004, nun im Rahmen eines dreijährigen multidisziplinären TOPIK-Projektes - bezeichnet als PixDat<sup>366</sup> - durchgeführt. In diesem Projekt werden zwei einander ergänzende Ansätze - bezeichnet als PIDA und xDat - zur Datenversorgung der für die Erdsystemanalyse zentralen multidisziplinäre Modellierung unter einem Dach vereint. So steht der Name PIDA für einen weiteren Ausbau der meteorologischen Datenbasis des PIK sowie deren unabdingbare Qualitätssicherung, während xDat [Wrobel et al. 2003] das Ziel verfolgt, den für punktverortete Zeitreihen erreichten Stand in analoger Weise auf gitter- und polygonverortete Zeitreihen auszuweiten.

Die Schlüsselfaktoren, die schon für die erfolgreiche Realisierung der beschriebenen Schnittstelle

---

<sup>366</sup> PixDat steht für *PIK eXtensible Database & Access Tools*.

ausschlaggebend waren, werden auch für die Erschließung der Gitter- und Polygondaten eingesetzt werden. Entsprechend wird wiederum die Kombination und Adaption moderner Technologien wie Datenbanken (hier ebenfalls unter Einsatz von Data Warehouse- und gegebenenfalls von relationalen OLAP-Konzepten (ROLAP, vgl. Kap. 2.3.4)) sowie Internet bzw. World Wide Web, die Entwicklung intuitiver graphischer Nutzeroberflächen, der Einsatz interaktiver Visualisierung sowie insbesondere ein iterativer Entwicklungsprozess zur Definition und Realisierung komfortabler und für die Anwender attraktiver Funktionalität verfolgt. Vor dem Hintergrund von Komplexität, Größenordnungen und Heterogenität der adressierten Datenräume wird auch für diese Entwicklung ein iteratives Systemdesign mit wohldefiniertem Kern sowie adaptierbarer und erweiterbarer Funktionalität als geeigneter Ansatz eingestuft. Das gewählte Konzept basiert auf der Entwicklung einer generischen Datenbankstruktur zur Aufnahme der Daten sowie der Entwicklung adaptiver Software, die - ähnlich wie schon für punktverortete Zeitreihen - die erforderliche Funktionalität ebenfalls weitgehend generisch bzw. konfigurierbar bereitstellt. Für die Durchführung wurden folgende Arbeitspakete (Working Packages) definiert:

#### ▪ **Data Management**

Im Arbeitspaket *Data Management* wird ein generisches Datenmodell entwickelt, das die effiziente datenbankgestützte Verwaltung von Daten in zeitlichen Hierarchien auf räumlichen Polygon- und Gitterstrukturen erlaubt. Hierfür werden basierend auf einem relationalen Datenbankmanagementsystem<sup>367</sup> Data Warehouse- und gegebenenfalls relationale OLAP-Konzepte verwendet. Zur schnellen Bereitstellung von räumlichen und zeitlichen Aggregationen werden diese vorberechnet und ebenfalls in der Datenbank vorgehalten. Für eine Raumverortung polygonbezogener Zeitreihen werden dabei die schon für Zeitreihenmetadaten eingesetzten hierarchischen räumlichen Klassifikatoren (vgl. Kap. 18.3.1) verwendet. Um die für Anwender attraktive Konvertierung von Zeitreihen auf Gitter- in Polygonstrukturen und *vice versa* anbieten zu können, werden ferner verschiedene Varianten einer entsprechenden Datenklassifizierung evaluiert, um eine geeignete dynamische Umsetzung dieser Konvertierungen zu ermöglichen<sup>368</sup>.

#### ▪ **Harmonisierung sozioökonomischer Zeitreihen**

Dieses Arbeitspaket adressiert die Bereitstellung einer geeignet weiterverarbeitbaren *integrierten* Datenbasis sozioökonomischer Zeitreihen. Die vorgesehenen Schritte beinhalten zunächst eine syntaktische Homogenisierung durch Transformation von Daten aus unterschiedlichen Quellen in das Datenmodell des xDat-Datenmanagements. Die Durchführung dieses Schrittes ist die Voraussetzung für einen einheitlichen Zugriff auf Daten aus unterschiedlichen Quellen; inhaltliche Heterogenitäten bleiben hier zunächst noch bestehen. Um auch diese sukzessive zu beseitigen, sollen nachfolgend Qualitätskontroll-Prozeduren für ausgewählte Datensätze in enger Zusammenarbeit mit anwendenden Projekten entwickelt werden; dies umfasst sowohl Plausibilitäts- und Konsistenzprüfungen wie die Konsolidierung und Komplettierung von Datensätzen. Schließlich sind für ausgewählte Variablen geeignete Funktionen zu definieren, die ihre flexible räumliche Aggregation erlauben, sowie Zyklen zur Aktualisierung und Erweiterung der Datengrundlage zu bestimmen.

#### ▪ **Data Processing**

Aufgabe des Arbeitspaketes *Data Processing* ist die Realisierung der nutzerdefinierbaren Verarbeitung von zuvor aus der Datenbank extrahierten Daten für nachfolgende Visualisierungen und Downloads. Hierfür wird ein xDat-Server entwickelt, der entsprechende Anforderungen von einer

---

<sup>367</sup> Vorgesehen ist der Einsatz von ORACLE 9i und/oder ORACLE 10g.

<sup>368</sup> Eine Konvertierung zwischen Gitter- und Polygondaten bedeutet eine räumliche Aggregation von Gitterzellen zu Polygonen bzw. eine Disaggregation von Polygonen in Gitterzellen. Dies wirft mehrere nichttriviale Problemstellungen auf, die eine geeignete Transformation von Datenwerten ebenso beinhalten wie Verfahren zur korrekten Zuordnung auch solcher Gitterzellen, die jeweils nur teilweise innerhalb eines Polygons liegen. Da die resultierenden Datenwerte die Grundlage wissenschaftlicher Arbeit bilden werden, sind hier sehr sorgfältige Untersuchungen und ein intensiver Austausch mit Fachwissenschaftlern der Anwendungsgebiete erforderlich.

Client-Anwendung entgegennimmt, die Daten aus der Datenbank extrahiert, weiter verarbeitet und anschließend zum Client überträgt. Der zur Verfügung gestellte Funktionsumfang wird die Auswahl einer thematischen, zeitlichen und räumlichen Untermenge aus multidimensionalen Datensätzen ebenso umfassen wie die Anwendung einer nutzerdefinierbaren Sequenz von analytischen Operatoren auf alle Werte der ausgewählten Daten; ebenso zeitliche und räumliche Aggregationen sowie Konvertierungen zwischen Gitter- und Polygondaten.

#### ▪ **Interaktive Visualisierung**

Ziel dieses Arbeitspaketes ist die Unterstützung des Anwenders bei der Evaluierung ausgewählter Daten sowie bei der Selektion von Untermengen für nachfolgende individuelle Modellierungsanforderungen durch geeignete Formen interaktiver Visualisierung. Vor dem Hintergrund der insbesondere für gegitterte Zeitreihen potentiell hohen Datenvolumina werden - sowohl basierend auf den bisher entwickelten Prototypen sowie verfügbaren Visualisierungswerkzeugen - geeignete Mischformen aus client- und serverseitiger Bildgenerierung definiert und adaptiert. Die vorgesehene Funktionalität umfasst die Generierung zweidimensionaler Karten für gegitterte sowie polygonverortete Zeitreihen zum Vergleich der räumlichen Verteilung von Datenwerten für einen nutzerdefinierbaren Zeitpunkt; ferner Sequenzen mehrerer Karten zur dynamischen Darstellung von Veränderungen über die Zeit (Film). Darüber hinaus ist - ebenfalls für gegitterte sowie polygonverortete Zeitreihen - die Erzeugung unterschiedlicher Arten von Zeitreihendiagrammen vorgesehen, um jeweils den Vergleich von Werteverläufen ausgewählter Zeitreihen über die Zeit zu ermöglichen; ferner werden für gegitterte Zeitreihen zusätzlich Varianten der Volumenvisualisierung integriert, um die Analyse dreidimensionaler Teilräume zu unterstützen.

#### ▪ **Download**

Die persistente Bereitstellung von prozessierten Daten ist die Aufgabe des Arbeitspaketes *Download*. Um eine flexible Weiterverwendbarkeit entsprechend der vielfältigen Bedürfnisse der Anwender sicherzustellen, werden verbreitete Formate wie ASCII, XML oder netCDF ebenso unterstützt werden wie verschiedene Bildformate zur Speicherung von Visualisierungsergebnissen oder Spezialformate, bspw. zur Weiterverarbeitung der Daten in Geoinformationssystemen.

#### ▪ **Nutzerschnittstelle**

Ziel dieses Arbeitspaketes ist die Entwicklung einer geeigneten netzwerkfähigen und plattformunabhängigen graphischen Oberfläche, die es dem Anwender erlaubt, die bereitgestellte Funktionalität in effizienter Weise zu nutzen. Dabei ist zwischen unterschiedlichen Anwenderinteressen, der Komplexität und Vielfalt der Funktionalität sowie intuitiver Bedienbarkeit zu vermitteln. Entsprechend wird die Realisierung - in Analogie zur bisher durchgeführten Schnittstellenentwicklung - wiederum in einem iterativen Prozess, basierend auf Modularität und Adaptivität, durchgeführt werden.

Das Projekt ist auf drei Jahre angelegt und kann auf den Erfahrungen und der Kompetenz aufbauen, die im Potsdam-Institut für Klimafolgenforschung bei der Erschließung allgemeiner Metadaten, Zeitreihenmetadaten und punktverorteter Zeitreihen gewonnen werden konnten. So ist es zu erwarten, dass innerhalb dieses Zeitraumes ein komfortabler und funktionaler Zugang für weitere, komplexe Datenressourcen der Erdsystemanalyse erfolgreich realisiert und so die autonome und individuelle Datenversorgung ihrer Wissenschaftler weiter verbessert werden kann.