

## 5 Internet, World Wide Web und Grids

„The Internet has revolutionized the computer and communications world like nothing before.“<sup>170</sup>

„The next IT revolution will be the Grid.“<sup>171</sup>

Gegenstand von Kap. 5 sind das Internet sowie die auf diesem aufsetzenden Konzepte World Wide Web und Grid. Kap. 5.1 dient der Einführung und stellt kurz Ursprünge und wesentliche Konzepte von Internet und World Wide Web vor; in Kap. 5.2 werden einige Aspekte der Nutzung von Internet und World Wide Web behandelt. Kap. 5.3 geht überblicksartig auf das Konzept der Grids ein, mit dem eine Nutzung des Internet zur komfortablen Kopplung verteilter IT-Ressourcen angestrebt wird. Ein abschließendes Fazit wird in Kap. 5.4 gezogen.

### 5.1 Einführung

Die Potentiale der globalen Netzwerkinfrastruktur Internet sowie des auf diesem aufsetzenden World Wide Web sind in den letzten Jahren breit diskutiert worden. Nicht zuletzt lassen sie auch eine neue Qualität wissenschaftlicher Arbeit greifbar nah erscheinen. So eröffnet etwa das Bundesministerium für Bildung und Forschung (BMBF) sein Strategisches Positionspapier zur Zukunft der wissenschaftlichen Information in Deutschland wie folgt:

„*Vision*

*Eine Wissenschaftlerin kommt morgens ins Labor. Sie holt sich aus der internen Datenbank des Rauminformationssystems den aktuellen Versuchsaufbau als Flussdiagramm auf die Anzeigetafel an der Laborwand. Während sie ihren Arbeitsplatz vorbereitet, lässt sie das System extern im Internet und in kommerziellen Datenbanken recherchieren, ob es zu einem Teilbereich ihres Versuches bereits bekannte Forschungsergebnisse gibt. Ein Reaktionszyklus, über den die Forscherin auf dem Weg zur Arbeit besonders intensiv nachgedacht hat, zeichnet sie als Strukturformel direkt auf die Tafel. Das System übernimmt das Formelbild als Frage und liefert wenig später mehrere Treffer, die es aus hochwertigen, qualitätsgeprüften Chemiedatenbanken hervorgeholt hat. Dann liest das Rauminformationssystem die Ergebnisse nach Rangfolge der Übereinstimmung zwischen Suchanfrage und Treffer laut vor. Auf Zuruf entscheidet die Wissenschaftlerin, welche Publikation verworfen und welche zur späteren Auswertung als abstract oder im Volltext in ihre persönliche, digitale Wissensbibliothek übernommen werden soll.*

*Science Fiction? Schon bald nicht mehr [...] [BMBF 2002, 1]*

Das BMBF adressiert in diesem Strategiepapier die Nutzung der neuen Technologien für einen verbesserten Zugriff auf wissenschaftliche Information, die es als eine Schlüsselressource für den Standort Deutschland wertet. Seit einigen Jahren wird in internationalen Bestrebungen unter dem Namen *Grid* bereits die Umsetzung einer anderen Vision in Angriff genommen – die Nutzung der Infrastruktur Internet zum komfortablen Zugriff auf vielfältige, dynamisch miteinander koppelbare, weltweit verteilte Ressourcen zur Wissensgewinnung aus digitalen Daten (vgl. Kap. 5.3). Nachfolgend werden zunächst kurz die Ursprünge und grundlegenden Konzepte von Internet (vgl. Kap. 5.1.1) und World Wide Web (vgl. Kap. 5.1.2) dargestellt.

<sup>170</sup> Leiner, B.M.; Cerf, V.G.; Clark, D.D.; Kahn, R.E.; Kleinrock, L.; Lynch, D.C.; Postel, J.; Roberts, L.G.; Wolff, S.: *A Brief History of the Internet* [Leiner et al. 2000].

<sup>171</sup> Statement des britischen Grid-Projektes GridPP [GridPP 2003].

### 5.1.1 Ursprünge und Konzept des Internet

Das *Internet* (vgl. bspw. [Leiner et al. 2000]) stellt eine heute nahezu weltumspannende Infrastruktur zur Vernetzung von Computern dar. Der Namensteil *inter* (lat. zwischen) verweist bereits darauf, dass es sich dabei nicht um ein einziges, homogenes Netzwerk handelt, sondern um einen Zusammenschluss aus einer Vielzahl einzelner Computernetze. Wichtige Prinzipien bilden der Zugriff auf verteilte Daten und Programme, eine paketbasierte Kommunikation und eine hohe Ausfallsicherheit. Diese Ziele lagen bereits dem 1966 von der amerikanischen Defense Advanced Research Projects Agency (DARPA) initiierten ARPANET zugrunde, das als Vorläufer des heutigen Internet gilt. Der Wunsch nach einer verteilten Rechnerinfrastruktur, die auch bei Ausfall von Teilsystemen noch funktionsfähig bleibt, war dabei nicht zuletzt militärisch motiviert – die Zerstörung einzelner Computer des Netzwerkes im Falle eines Angriffs sollte nicht zum Erliegen des Gesamtsystems führen. Im September 1969 wurde der erste Host-Rechner - am Network Measurement Center der UCLA - an das ARPANET angeschlossen. Einen Monat später kam der zweite - am Stanford Research Institute (SRI) - hinzu und die erste Nachricht von Host zu Host wurde vom Network Measurement Center zum SRI versandt. Bis Ende 1969 war die Zahl der Host-Rechner des ARPANET um weitere zwei (an der UC Santa Barbara und der University of Utah) auf insgesamt vier angewachsen. Die Zahl der angeschlossenen Rechner nahm in der Folge schnell zu. Von 1971 bis 1972 wurde zudem auf den beteiligten Rechnern das Network Control Protocol (NCP) implementiert, so dass eine Basis für die Entwicklung von Applikationen für das Netz bereitstand; die erste E-Mail-Software wurde bis zum Sommer 1972 realisiert. Im Oktober 1972 fand die erste öffentliche Demonstration des ARPANET bei der International Computer Communication Conference (ICCC) statt. Ein weiterer wichtiger Termin für die Entwicklung hin zum heutigen Internet war der 1. Januar 1983. An diesem Tag wurde das Network Control Protocol NCP durch die Einführung des für die Kommunikation zwischen heterogenen Rechnern ausgelegten Transmission Control Protocol / Internet Protocol (TCP/IP) abgelöst, das auch die Abspaltung des militärisch genutzten Teils des ARPANET als MILNET ermöglichte. Bis Mitte der 80er Jahre waren verschiedene weitere Rechnernetze entstanden, die bis dahin weitgehend isoliert voneinander betrieben wurden. Dies änderte sich, als ab 1985 das Wissenschaftsnetz NSFNET begann, ebenfalls TCP/IP als Protokoll zu übernehmen. In der Folge schlossen sich immer mehr Rechnernetze zusammen - [Leiner et al. 2000] nennen die Zahl von über 50.000 Netzen, davon allein 29.000 in den USA -, die nun in ihrer Gesamtheit das heutige Internet formen.

### 5.1.2 Ursprünge und Konzept des World Wide Web

Von zentraler Bedeutung für die umfassende Nutzung des Internet war die Entwicklung des *World Wide Web* ab 1990, das auf Arbeiten von Tim Berners-Lee am CERN (Organisation Européenne pour la Recherche Nucléaire)<sup>172</sup> bei Genf an der Grenze zwischen Schweiz und Frankreich zurückgeht (vgl. [Berners-Lee 1999]). Das World Wide Web erweitert das Internet um ein auf dieses aufsetzendes global verteiltes Hypertext-System, das auch die Integration von Bildern und zunehmend von multimedialen Elementen wie Ton und Film erlaubt; zudem wird eine komfortable, graphisch-interaktive Nutzung von Internet-Diensten wie E-Mail ermöglicht.

Wesentliche Bestandteile des World Wide Web bilden Dokumente, die über eine hierfür entwickelte Auszeichnungssprache formatiert werden (Hypertext Markup Language, HTML) und Verweise auf andere Dokumente enthalten können (sog. Hyperlinks). Zur Bereitstellung der Dokumente über das Netz dienen spezielle Softwareprogramme (Web-Server); der komfortable Abruf von Dokumenten und ihre graphische Darstellung auf dem Rechner

---

<sup>172</sup> [www.cern.ch/](http://www.cern.ch/)

des Anwenders erfolgt durch Client-Programme, die als Web-Browser bezeichnet werden. Das World Wide Web basiert auf dem Hypertext Transfer Protocol (HTTP) zur Kommunikation zwischen Web-Browsern und Web-Servern sowie der Verwendung von Universal Resource Identifiers (URIs)<sup>173</sup> zur eindeutigen Adressierung der verteilten Ressourcen.

Berners-Lee und der ebenfalls am CERN beschäftigte Robert Callieu begannen den Betrieb der ersten Vorläufer von zwei Web-Clients (ein kombinierter Browser und Editor namens WorldWideWeb, lauffähig auf NeXT-Rechnern) auf ihren jeweiligen Computern zu Weihnachten 1990; der ebenfalls von ihnen entwickelte erster Web-Server lief auf einem Rechner mit der Internetadresse info.cern.ch. In der Folge stieß ihre Konzeption zunehmend auf Interesse; so wurden erste graphische Web-Browser entwickelt (Erwise im April 1992 an der Universität Helsinki sowie ViolaWWW von Pei Wie an der Universität Berkeley). Im Februar 1993 stellte die NCSA den wesentlich von Marc Andreessen mitentwickelten Browser Mosaic zur Verfügung, der über das Web heruntergeladen und unaufwendig installiert werden konnte und daher große Verbreitung fand. Andreessen gründete in der Folge die Firma Mosaic Communications Corp., die im April 1994 in Netscape umbenannt wurde. Im Dezember 1994 fanden innerhalb von 72 Stunden weitere wesentliche Schritte in der Geschichte des World Wide Web statt. Am 14. Dezember 1994 kam zum ersten Mal das World Wide Web Consortium (W3C)<sup>174</sup> zusammen, das sich in der Folge als Gremium für Standardisierungen im Web etablierte. Am 15. Dez. 1994 gab Netscape bekannt, seinen Browser Mozilla unter dem Namen Navigator 1.0 kostenlos zur Verfügung zu stellen. Der Navigator 1.0 konnte über das Internet heruntergeladen und auf Windows, UNIX mit X-Windows sowie Macintosh betrieben werden, so dass dieser Browser bereits innerhalb weniger Monate von der Mehrzahl der Web-Anwender benutzt wurde. Am dritten Tag schließlich, dem 16. Dezember 1994, gab das CERN die Entscheidung bekannt, sich künftig primär auf die Entwicklung eines neuen Teilchenbeschleunigers (des Large Hadron Collider LHC, vgl. Kap. 5.3.4) zu konzentrieren und von der weiteren Entwicklung des World Wide Web zurückzuziehen.

## 5.2 Internet und World Wide Web als Infrastruktur

Das Internet und das darauf aufsetzende World Wide Web spielen heute eine zentrale Rolle als Infrastruktur zum Austausch von Daten zwischen potentiell beliebig verteilten Rechnern und Anwendern. Nachfolgend werden einige diesbezügliche Entwicklungstendenzen anhand von Größenordnungen (Kap. 5.2.1) und Formen der Nutzung (Kap. 5.2.2) dargestellt. Kap. 5.2.3 verweist auf die Bedeutung von Internet und World Wide Web für die Anbindung und Bereitstellung bestehender Ressourcen.

### 5.2.1 Größenordnungen

In den 90er Jahren fand eine stürmische Entwicklung des World Wide Web statt, die bis heute anhält. Waren im Januar 1993 rund 50 Web-Server aktiv [Berners-Lee 1999, 106], hatte sich ihre Zahl bereits bis zum Oktober des selben Jahres auf etwa 500 verzehnfacht [December, Randall 1994, 46]. Die Anzahl der heute aktiven Web-Server kann nur noch abgeschätzt werden. So berichten [Lawrence, Giles 1999] über ihre Abschätzungen des sog. *Publicly Indexable Web*. Hierunter wird von den Autoren derjenige Ausschnitt des World Wide Web verstanden, der potentiell von Suchmaschinen indiziert werden kann; ausgenommen sind hierbei bspw. HTML-Seiten, die nur durch vorherige Autorisierung oder durch einen Zugriff über Formulare abrufbar sind. Die Autoren testeten im Februar 1999

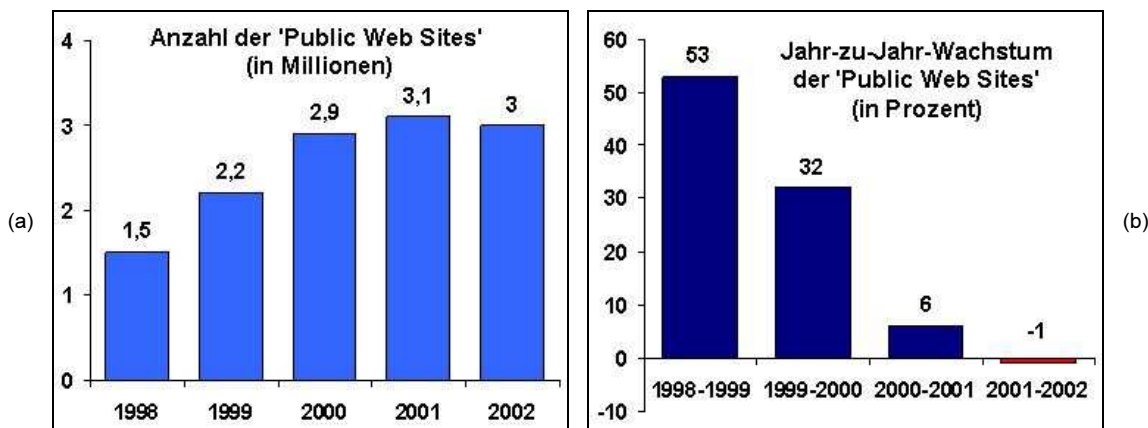
---

<sup>173</sup> URIs sind die allgemeinere Form der Uniform Resource Locator (URL).

<sup>174</sup> [www.w3c.org](http://www.w3c.org)

3,6 Millionen zufällig ausgewählte IP-Adressen am diesbezüglichen Standard-Port<sup>175</sup>; dabei fanden sie im Schnitt einen Web-Server bei einem von 269 Zugriffen. Durch Hochrechnung gelangten sie zu einer Anzahl von insgesamt 16 Millionen Web-Servern, von denen 2,8 Millionen als relevant für das Publicly Indexable Web klassifiziert wurden – wobei 83 Prozent der Server zur Bereitstellung kommerzieller Inhalte verwendet wurden und 6 Prozent auf Forschung und Lehre entfielen. Die Zahl der insgesamt zu diesem Zeitpunkt im Publicly Indexable Web verfügbaren Seiten (Web Pages) wurde von den Autoren mit rund 800 Millionen veranschlagt (im Vergleich zu mindestens 320 Millionen für Dez. 1997); das so bereitgestellte Datenvolumen auf 15 Terabyte.

Basierend auf Daten, die vom 1997 initiierten Web Characterization Project des Online Computer Library Center (OCLC)<sup>176</sup> erhoben wurden, vergleichen [O'Neill et al. 2003] Trends in der Entwicklung des sog. *Public Web* von 1998 bis 2002. Das Public Web setzt sich dabei aus denjenigen Websites zusammen, die vollständig frei zugänglich sind oder zumindest einen signifikanten Anteil von für alle Webbenutzer frei zugänglichem Inhalt bereitstellen (bezeichnet als *Public Web Sites*)<sup>177</sup>. Für Juni 2002 wird eine Anzahl von rund 3 Millionen (3.080.000) solcher Public Web Sites angesetzt, die den Autoren zufolge 35 Prozent des gesamten World Wide Web ausmachen. Die von einer Public Web Site durchschnittlich bereitgestellte Zahl von Webseiten wird mit 441 angegeben, die Gesamtanzahl der Seiten im Public Web mit annähernd 1,4 Milliarden.



**Abb. 5.1** - Entwicklung des Public Web von 1998 bis 2002 laut *Web Characterization Project*<sup>178</sup>.  
(a) Anzahl der dort enthaltenen Websites; (b) Wachstumsentwicklung von Jahr zu Jahr.

Die Diagramme in Abb. 5.1 veranschaulichen die Entwicklung der Anzahl der Public Web Sites im Beobachtungszeitraum. Es wird deutlich, dass sich die Zahl dieser Websites von 1998 bis 2002 verdoppelt hat; allerdings ist zugleich eine stetige Verlangsamung und für den Zeitraum von 2001 bis 2002 sogar ein leichter Rückgang zu verzeichnen. Die Autoren betonen, dass die beobachteten Werte nicht einem tatsächlichen Anwachsen der Zahl der verfügbaren Seiten oder des Datenvolumens im World Wide Web widersprechen; sie sehen jedoch, zumindest für den gegenwärtigen Zeitpunkt, ein Plateau in der Anzahl derjenigen Individuen, Organisationen und Unternehmen erreicht, die eine Web-Präsenz über eine von ihnen betriebene Website umsetzen.

<sup>175</sup> Portnummern erlauben die Adressierung einzelner Dienste an einer IP-Adresse; für viele Dienste wie E-Mail und Web-Server haben sich Standardnummern etabliert.

<sup>176</sup> <http://wcp.oclc.org>

<sup>177</sup> Das Web Characterization Project (vgl. <http://wcp.oclc.org/stats/size.htm>) unterscheidet zwischen *Public Web Sites*, *Private Web Sites* sowie *Provisional Web Sites*. Zu letzteren werden Websites gerechnet, die sich in einem unfertigen Zustand befinden oder bedeutungslose oder triviale Inhalte bereitstellen.

<sup>178</sup> Diagramme erstellt in Anlehnung an [O'Neill et al. 2003, Fig.1; Fig.2].

Im Report *How Much Information? 2003* [UC SIMS 2003ab] wird zwischen dem Datenvolumen statisch bereitgestellter HTML-Seiten (bezeichnet als sog. *Surface Web*) und dynamisch aus Datenbanken erzeugter HTML-Seiten (sog. *Deep Web*) unterschieden (vgl. Tab. 5.1). Dabei wird ein deutliches Anwachsen festgestellt; so wird etwa das Datenvolumen des Surface Web, das in einem vorangegangenen Report (vgl. [UC SIMS 2000]) noch auf 20 bis 50 Terabyte abgeschätzt wurde, für 2002 bereits auf 167 Terabyte veranschlagt. Die Autoren bezeichnen das Internet entsprechend als „*the fastest growing new medium of all time [...]*“ [UC SIMS 2003b].

Medium	Datenvolumen 2002 in Terabyte
Surface Web	167
Deep Web	91.850
E-Mails (Originals)	440.606
Instant messaging	274
<b>Total</b>	<b>532.897</b>

Tab. 5.1 - Datenvolumen im Internet 2002 laut Report *How Much Information? 2003*<sup>179</sup>.

### 5.2.2 Nutzung

2002 hatten weltweit rund 600 Millionen Menschen Zugriff auf das Internet, die sich allerdings ungleichmäßig auf die unterschiedlichen Regionen der Erde verteilten – so stellten Europa und Nordamerika gemeinsam bereits rund zwei Drittel aller Internetnutzer (vgl. Abb. 5.2a). Im globalen Durchschnitt verbringen Nutzer dabei im Monat 11 Stunden und 24 Minuten online, in den USA hingegen durchschnittlich 25 Stunden und 25 Min. zuhause sowie 74 Stunden und 26 Minuten am Arbeitsplatz [UC SIMS 2003b].

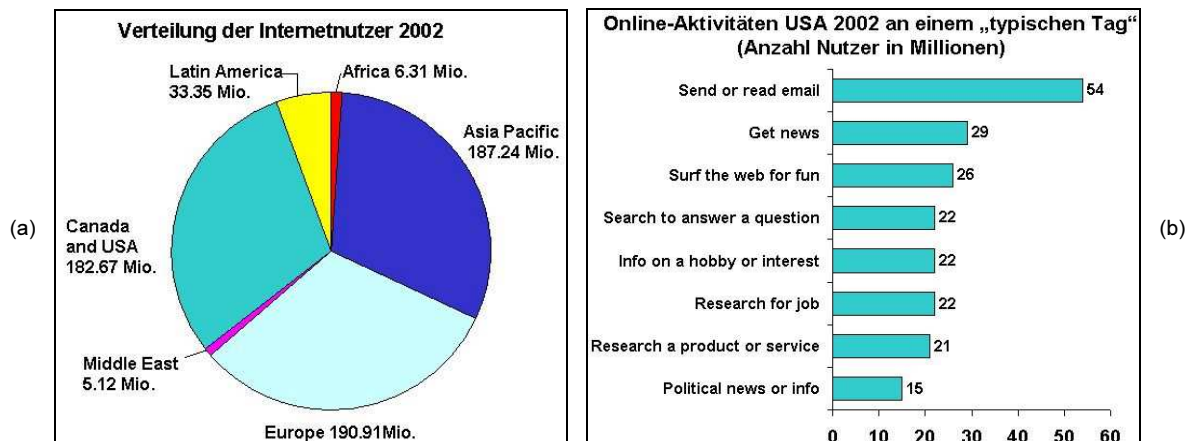


Abb. 5.2 - Nutzung des Internet: (a) Weltweite Verteilung der Internetnutzer 2002<sup>180</sup>; (b) die acht häufigsten Online-Aktivitäten der US-Amerikaner 2002<sup>181</sup>.

Im Diagramm in Abb. 5.2b werden anhand von Daten des Pew Internet & American Life Project [Madden 2003] die acht häufigsten Formen von Online-Aktivitäten dargestellt, die von US-Amerikaner „an einem typischen Tag“ unternommen werden. Auffällig ist, dass das Lesen und Versenden von E-Mail die bei weitem am häufigsten ausgeführte Tätigkeit darstellt. Hiermit korrespondiert das in [UC SIMS 2003b] für das Jahr 2002 abgeschätzte Datenaufkommen durch den Originalinhalt von E-Mails von bis zu 440.606 Terabyte (vgl. Tab. 5.1). In [Madden 2003] wird ferner auf deutliche demographische Unterschiede in den einzelnen Online-Aktivitäten der US-Amerikaner verwiesen – danach sucht bspw. ein großer Anteil der weiblichen Nutzer nach Informationen zu Gesundheitsthemen oder Religion,

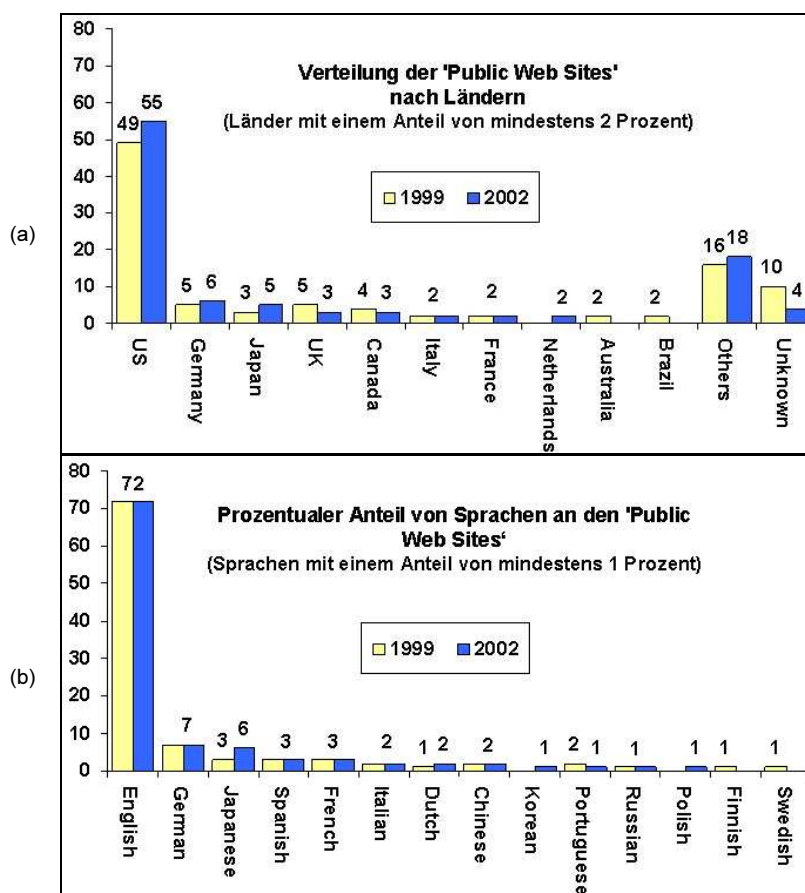
<sup>179</sup> Tabelle übernommen aus [UC SIMS 2003b, Tab. 8.1].

<sup>180</sup> Diagramm erstellt anhand von Daten aus [UC SIMS 2003a, Tab. 1.14]; vgl. auch [UC SIMS 2003b].

<sup>181</sup> Diagramm erstellt anhand von Daten aus [Madden 2003, viii].

während ein großer Prozentsatz der männlichen Nutzer eher verschiedene Arten von Nachrichten bevorzugt; ebenso finden sich beispielsweise jeweils spezifische Unterschiede zwischen Nutzern unterschiedlicher Einkommens- und Bildungsstufen, innerhalb ethnischer Gruppen, zwischen jungen und älteren Nutzern oder Land- und Stadtbevölkerung.

Bereits aus Abb. 5.2a wurde ersichtlich, dass für die Internetnutzung keine globale Gleichverteilung gegeben ist. [O'Neill et al. 2003] verdeutlichen, dass eine international ungleiche Verteilung ebenfalls für die Länder, in denen Public Web Sites (vgl. Kap. 5.2.1) bereitgestellt werden, sowie für die hierfür verwendeten Sprachen gilt. Danach (vgl. Abb. 5.3a) stammte die deutliche Mehrheit der 1999 betrachteten Public Web Sites mit 49 Prozent aus den USA, gefolgt von Deutschland mit 5 Prozent. Gegenüber 1999 haben sich für 2002 nur geringe Veränderungen ergeben; so stieg der Anteil der USA von 49 auf 55 Prozent, der Anteil japanischer Websites von 3 auf 5 Prozent und der deutsche Anteil von 5 auf 6 Prozent; Brasilien und Australien, 1999 noch jeweils mit 2 Prozent vertreten, fielen 2002 unter diesen Grenzwert.



**Abb. 5.3** - Aufteilung von Websites im Public Web für die Jahre 1999 und 2002<sup>182</sup>.  
(a) nach Ländern; (b) nach Sprachen.

Der Anteil einzelner Sprachen im Public Web (Abb. 5.3b) zeigt für 2002 eine deutliche Dominanz von Englisch mit 72 Prozent der Websites; Deutsch nimmt auch hier mit 7 Prozent den zweiten Rang ein. Auch hier sind im Vergleich zu 1999 wenig Veränderungen festzustellen. Der Anteil von Englisch und Deutsch blieb, ebenso wie der der meisten anderen Sprachen, unverändert. Ein deutliches Anwachsen ist für Japanisch festzustellen (um 50 Prozent von einem Anteil von 3 Prozent für 1999 auf einen Anteil von 6 Prozent für 2002; vgl. auch das Anwachsen der japanischen Websites im selben Zeitraum in Abb. 5.3.a);

<sup>182</sup> Diagramme erstellt in Anlehnung an [O'Neill et al. 2003, Fig.3 bis Fig.6].

Schwedisch und Finnisch fielen 2002 unter die 1 Prozent-Grenze; Polnisch und Russisch hingegen konnten sie nun überschreiten.

### 5.2.3 Anbindung bestehender Ressourcen

Auch wenn es ursprünglich als verteiltes Hypertext-System konzipiert wurde, wird das auf das Internet aufsetzende World Wide Web in zunehmendem Maße für die Anbindung vorhandener Ressourcen wie Datenbanken oder Unternehmensanwendungen genutzt. Als ausschlaggebend hierfür kann das Ineinandergreifen mehrerer Faktoren angesehen werden: Internetanbindung und Web-Browser können weitreichend vorausgesetzt werden; eine Anbindung bestehender Anwendungen an das World Wide Web erlaubt es, diese potentiell weltweit zur Verfügung zu stellen. Während insbesondere für ältere Anwendungen zudem bestimmte Hard- und Softwarevoraussetzungen erforderlich sind, kann eine webfähige Schnittstelle auf diese nicht nur netzwerkfähig, sondern zudem plattformunabhängig bereitgestellt werden, so dass nun ein ortsunabhängiger Zugriff von unterschiedlichen Rechnersystemen aus ermöglicht wird. Ferner kann eine entsprechende Schnittstelle je nach Bedarf entweder weltweit oder in abgrenzbaren Teilnetzen wie Extranets oder Intranets eingesetzt werden; ein weiterer wesentlicher Vorteil ist die Möglichkeit einer zentralen Bereitstellung, die die Wartbarkeit dieser Zugriffsschnittstellen entscheidend erleichtert.

Eine derartige Nutzung des World Wide Web wird immer wichtiger - nach [Röwekamp, Roßbach 2000] können sich nur solche Unternehmen auf dem Markt behaupten, die auf geänderte Kundennachfragen innerhalb kürzester Zeit mit einem entsprechenden Angebot im World Wide Web reagieren; den Autoren zufolge kann es sich kaum noch ein Unternehmen leisten, nur statische Webseiten bereitzustellen. [Plessl, Wilde 2001] konstatieren entsprechend einen Trend, das World Wide Web als Plattform für die Entwicklung von Applikationen zu nutzen, um bestehende IT-Ressourcen einzubinden. Hiermit korrespondieren wiederum die von [UC SIMS 2003b] für 2002 abgeschätzten Datenvolumina (vgl. Tab. 5.1) – während für das aus statisch bereitgestellten HTML-Seiten bestehende Surface Web 167 Terabyte veranschlagt werden, wird für die dynamisch aus Datenbanken erzeugten HTML-Seiten des Deep Web mit 91.850 Terabyte ein um den Faktor 550 größerer Wert angegeben.

Dennoch darf nicht verkannt werden, dass das World Wide Web nur eine der Möglichkeiten darstellt, das Internet zum Zugriff auf verteilte Ressourcen zu nutzen. So berichten Nielsen NetRatings, dass im November 2003 bereits drei Viertel der Internetnutzer zuhause oder am Arbeitsplatz über sog. *non-browser based applications*, insbesondere über Windows Media Player, AOL Instant Messenger und RealOne/RealPlayer, auf das Internet zugriffen [NNR 2003]. Auch das nachfolgend vorgestellte Konzept der Grids verdeutlicht, dass sich die Potentiale des Internet nicht mit dem World Wide Web erschöpfen.

## 5.3 Grids

### 5.3.1 Ansatz

Seit einigen Jahren wird in internationalen Bestrebungen unter dem Namen *Grid* der Versuch unternommen, basierend auf dem Internet eine bedarfsgetriebene Kopplung und Synergie vielfältiger Ressourcen zur Wissensgewinnung aus digitalen Daten zu ermöglichen. Dabei wird nicht weniger als eine weitere informationstechnologische Revolution adressiert, wie der Homepage des britischen GridPP Project zu entnehmen ist:

- ▶ „*The next IT revolution will be the Grid. While the Web is aimed mainly at the exchange of information, the Grid is concerned with the exchange of computer power, data storage, and access to large databases, without users searching for these resources manually. Computer services will become a utility like electricity, supplied via a Grid.*“

*Once connected to the Grid, the end user will see it essentially as one large computer system. The Grid is a practical solution to the data-intensive problems that must be overcome if the computing needs of many scientific communities and industry are to be fulfilled over the next decade.*" [GridPP 2003]

Das Konzept des Grid soll es ermöglichen, zur Lösung komplexer Fragestellungen, die nicht mit lokal vorhandenen Ressourcen bewältigt werden können, die hierfür erforderlichen weltweit verteilten informationstechnologischen Ressourcen je nach Bedarf verbinden und wie ein einziges virtuelles Computersystem einsetzen zu können. Dies umfasst zählen Rechnerressourcen, die von vielen tausend PCs und Workstations bis hin zu den besten und teuersten Höchstleistungsrechnern reichen können, ferner kostspielige Instrumente wie Weltraumteleskope oder Satelliten, ebenso Datenbanken mit hochwertigen Datenressourcen zu allen relevanten Wissensgebieten aus Naturwissenschaft, Geisteswissenschaft, Wirtschaft etc. sowie aufwendige Softwareressourcen, bspw. Simulationsmodelle zu unterschiedlichsten Fragestellungen.

### 5.3.2 Definition

Der Begriff *Grid* wurde Mitte der 90er Jahre des 20. Jahrhunderts geprägt, um eine angestrebte verteilte Rechner- und Dateninfrastruktur für anspruchsvolle Anforderungen aus Wissenschaft und Ingenieurwesen zu beschreiben. Angestrebt wird dabei

- ▶ „[...] *the synergistic use of high-performance networking, computing, and advanced software to provide access to advanced computational capabilities, regardless of the location of users and resources.*" [Foster, Kesselman 1998]

Als Data Grid wird eine Spezialisierung und Erweiterung des Grid verstanden, die speziell die effiziente Nutzung großer und verteilter Datenmengen adressiert [Chervenak et al. 2001]. Grids sollen über ein weltumspannendes Netzwerk überall und jederzeit Rechenleistung gegen Entgelt zur Verfügung stellen – ähnlich wie ein Stromnetz, dass Zugriff auf Energie über standardisierte Steckdosen erlaubt [Hoffmann et al. 2002, 3]. Die vom *GPP Project* gezogene Parallele zwischen Grid und Elektrizität ist also nicht zufällig. [Foster, Kesselman 1998] vergleichen die Situation computerintensiver Anwendungen mit dem Status der Nutzung von Elektrizität um 1910. Zu dieser Zeit konnte bereits elektrischer Strom erzeugt werden; auch darauf basierende Geräte wurden bereits entwickelt. Ihre Verwendung wurde jedoch durch dem Umstand beeinträchtigt, dass jeder potentielle Nutzer hierfür einen eigenen Generator betreiben musste. Entsprechend konstatieren die Autoren, dass die wahre revolutionäre Entwicklung nicht die Elektrizität, sondern das elektrische Stromnetz (engl. electrical power grid) im Verbund mit zugehörigen Übertragungs- und Verteilungstechnologien gewesen sei, da auf diese Weise ein universeller, verlässlicher und preiswerter Zugriff auf Energie realisiert werden konnte. Zur Bezeichnung der avisierten neuen Infrastruktur wählen sie daher den Begriff *Computational Grid* in bewusster Analogie zu elektrischen Stromnetzen und deren Vorteilen. Ein Computational Grid ist danach

- ▶ „[...] *a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities.*" [Foster, Kesselman 1998]

Dabei steht den Autoren zufolge die Bezeichnung *infrastructure* für eine Verteilung der Rechenlast auf unterschiedliche Computer; mit der Bezeichnung *dependable* wird die als fundamental für ein Grid eingestufte Anforderung nach Zuverlässigkeit im Bezug auf vorhersagbare Performance für oft hohe Ansprüche adressiert. Die Bezeichnung *consistent* bezieht sich auf die ebenfalls zentrale Forderung nach Konsistenz des Grid bezüglich Schnittstellen, bereitgestellten Diensten etc.; die Bezeichnung *pervasive* hingegen zielt auf



eine angestrebte weitreichende Verfügbarkeit des Grid. Die Bezeichnung *inexpensive* schließlich steht für die Forderung, dass die Nutzung eines Grid preiswert im Vergleich zum durch dieses erzielbaren Nutzen sein muss, um eine breite Akzeptanz und Verwendung zu erreichen.

[Hoffmann et al. 2002, 3f.] verweisen darauf, dass die vielfach aufgegriffene Analogie von Grid und Stromnetz allerdings nur bedingt trägt. Anders als beim Stromnetz, bei dem nur wenige Parameter wie Spannung, Stromstärke und Frequenz zu berücksichtigen sind, müssen für die Kopplung verteilter Rechnerressourcen weitaus mehr Aspekte einbezogen werden, die von der Architektur der Rechner über verschiedene Softwareebenen bis zur Anbindung an des Netzwerk reichen; hier sind den Autoren zufolge noch nicht alle Probleme vollständig untersucht und gelöst.

### 5.3.3 Architektur

Die Kopplung der verteilten Ressourcen in einem Grid stellt besondere Anforderungen, die über typische Client-Server-Architekturen hinausgehen. In solchen Architekturen stellt ein Server über vordefinierte Schnittstellen bestimmte Dienste bereit, die von Clients abgerufen werden können. Ein Beispiel hierfür ist das Zusammenspiel von Web-Servern und Web-Browsern im World Wide Web; hier können die Clients (Web-Browser) über standardisierte Schnittstellen HTML-Dokumente bei einzelnen Web-Servern abrufen. Während bei Client-Server-Architekturen meist zunächst die Schnittstelle des Servers festgelegt wird und so die möglichen Interaktionen der Clients definiert werden, ist dies im Grid nicht möglich. Hier sollen unterschiedliche, autonome Kommunikationspartner in potentiell beliebigen Konstellationen miteinander in Beziehung treten, die sich zudem durch Hinzukommen neuer oder Wegfall bisher integrierter Ressourcen dynamisch über die Zeit verändern können [Hoffmann et al. 2002, 5]. Ein Architektorentwurf zur Realisierung von Grids (vgl. Abb. 5.4) wird in [Foster et al. 2001] vorgestellt. Dabei werden insgesamt fünf Schichten unterschieden (vgl. auch [Hoffmann et al. 2002, 5ff.]):

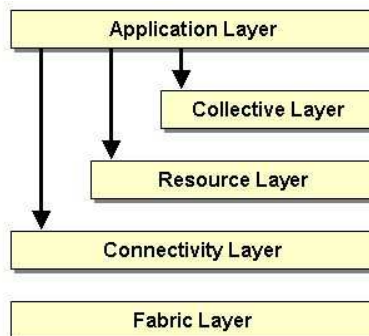


Abb. 5.4 - Bestandteile einer Grid-Architektur<sup>183</sup>.

- **Hardware und Software-Ressourcen – der *Fabric Layer***

Der *Fabric Layer* bildet die unterste Ebene der Grid-Architektur; hier werden die unterschiedlichen Ressourcen (Rechner, Speichersysteme, Sensoren, Netzwerke etc.) bereitgestellt. Auf dieser Ebene sind bereits eine Vielzahl von Aspekten zu berücksichtigen, insbesondere die Heterogenität der beteiligten Komponenten, Anpassungen des Systems an lokale Eingriffe, geeignete Reaktionen auf eventuelle Systemausfälle etc.

- **Verbindungsebene – der *Connectivity Layer***

Aufgabe der als *Connectivity Layer* bezeichneten zweiten Schicht der Grid-Architektur ist ein einfacher und sicherer Zugriff auf die Ressourcen über Netzwerkverbindungen, der

<sup>183</sup> Abbildung erstellt in Anlehnung an [Foster et al. 2001, Fig. 2].

Austausch von Daten zwischen den Ressourcen ebenso wie die sichere Identifikation von Nutzern und Ressourcen. Auch auf dieser Ebene sind vielfältige Aufgaben zu lösen. Hierunter fallen bspw. Aspekte wie *Single Sign On* oder *Delegation* – Nutzer sollten sich nur einmal an das System anmelden müssen, um danach Zugriff auf die verteilten Ressourcen zu erhalten, ohne dass jeweils eine erneute Anmeldung erforderlich ist; ferner sollen Softwareprogramme, die von einem Nutzer im Grid gestartet werden, dessen Zugriffsrechte übernehmen und ihrerseits an weitere Programme übertragen können.

- **Zugriff auf die Ressourcen – der *Resource Layer***

Die dritte Ebene der Grid-Architektur, der *Resource Layer*, dient zur Initiierung und Kontrolle des Zugriffs auf einzelne Ressourcen. Hierunter fallen Aufgaben wie die Abfrage ihres aktuellen Zustands, die Definition von Anforderungen und ihre Aushandlung mit den lokalen Systemen, sowie Kontrolle und Steuerung des Programmablaufes. Diese Ebene ist ausschließlich mit individuellen Ressourcen befasst; ihre Koordination wird von der vierten Schicht - dem *Collective Layer* - übernommen.

- **Kooperatives Ressourcen-Management – der *Collective Layer***

Der *Collective Layer* baut auf dem *Resource Layer* auf und übernimmt die Koordination der globalen Aspekte der Ressourcenverwaltung. Hierunter fallen bspw. die Suche nach verfügbaren Ressourcen anhand von Namen oder Eigenschaften, Reservierungen oder die zeitliche und örtliche Planung der Durchführung komplexer Anforderungen. Ferner soll diese Schicht Aufgaben wie Datenreplikation zur Erhöhung der Performance übernehmen sowie Dienste bereitstellen, die es bspw. erlauben, die jeweils am besten geeigneten Softwareimplementierungen und Ausführungsplattformen für ein konkretes Problem zu identifizieren.

- **Anwendungsebene – der *Application Layer***

Der *Application Layer* bildet die oberste Ebene der Grid-Architektur. Diese Schicht umfasst die Anwendungen, die auf den von den tieferen Schichten bereitgestellten Funktionen und Diensten aufbauen. [Foster et al. 2001] verweisen darauf, dass hier ihrerseits komplexe Technologien<sup>184</sup> zum Einsatz kommen können.

### 5.3.4 Anwendungsgebiete

Die Einsatzpotentiale von Grids sind vielfältig. [Foster, Kesselman 1998] unterscheiden fünf hauptsächliche Formen möglicher Anwendungen mit jeweils spezifischen Herausforderungen:

- |                                   |  |
|-----------------------------------|--|
| <i>Distributed Supercomputing</i> | <ul style="list-style-type: none"> <li>▸ Grids sollen die Aggregation von verteilten Computerressourcen ermöglichen, um so rechenintensive Probleme anzugehen, die von einem einzelnen System nicht bewältigt werden können. Eine solche Aggregation kann beispielsweise im Zusammenschluss der Mehrzahl aller Supercomputer eines Landes, aber auch in der Kopplung aller Workstations eines Unternehmens bestehen. Als Anwendungsbeispiele nennen die Autoren sowohl Distributed Interactive Simulation (DIS) für Training und Planung im Militär wie die Simulation komplexer Prozesse bspw. in den Bereichen Klimamodellierung oder Chemie. Herausforderungen liegen hier bspw. in der effizienten Verknüpfung der in der Regel</li> </ul> |
|-----------------------------------|--|

---

<sup>184</sup> So wird etwa die Verwendung von CORBA (Common Object Request Broker Architecture), einer Infrastruktur zur sprachunabhängigen Kopplung verteilter heterogener Softwareanwendungen, im Rahmen von Grid-Entwicklungen diskutiert (vgl. bspw. [Denis et al. 2001] [Verma et al. 2001] [Pérez et al. 2002]).

- knappen und kostspieligen Ressourcen oder im Erzielen hoher Performance trotz der Einbeziehung heterogener Systeme.
- High-Throughput Computing* ▶ Grids sollen die Verteilung einer großen Zahl von lose gekoppelten oder unabhängigen Einzelaufgaben ermöglichen. Als Ziel wird dabei die Ausnutzung brachliegender Prozessorleistungen (Idle Times) genannt; mögliche Anwendungen<sup>185</sup> werden u.a. in der Lösung kryptographischer Probleme oder im Design von Computerchips gesehen. Die Autoren gehen hier nicht auf spezifische Herausforderungen ein.
- On-Demand Computing* ▶ Grids sollen für kurzzeitige Anforderungen Ressourcen verfügbar machen, die lokal nicht kosteneffizient oder nicht in geeigneter Weise bereitgestellt werden können. Diese Ressourcen können sowohl in Rechenleistung, Software, Daten oder Geräten wie Sensoren bestehen. Herausforderungen sind hier in der dynamischen Natur der Anforderungen sowie der potentiell hohen Anzahl von Anwendern wie Ressourcen begründet; sie umfassen bspw. Aspekte wie die Lokalisierung der Ressourcen, Konfiguration, Fehlertoleranz oder Mechanismen zur Kostenabrechnung.
- Data-Intensive Computing* ▶ Grids sollen die Synthese neuer Informationen aus geographisch verteilt gehaltenen Datenquellen wie Datenbanken oder digitalen Bibliotheken ermöglichen. Anwendungen werden in der Auswertung großer Datenmengen aus Teilchenphysik, Astronomie oder Meteorologie gesehen; Herausforderungen bilden hier den Autoren zufolge die Organisation und Konfiguration komplexer und sehr umfangreicher Datenströme.
- Collaborative Computing* ▶ Grids sollen die Kommunikation oder die Zusammenarbeit zwischen verschiedenen Teilnehmern unterstützen; bspw. durch eine geteilte Nutzung von Datenarchiven und Simulationen. Herausforderungen liegen hier in Echtzeitanforderungen sowie der Vielzahl möglicher Interaktionen.

Rund drei Jahre später stellen [Foster et al. 2001] fest, dass der Begriff Grid nicht einheitlich wahrgenommen und mit einer gewissen Beliebigkeit verwendet werde. Sie betonen, dass als das spezifische durch Grids adressierte Problem das koordinierte Teilen von Ressourcen sowie das Lösen von Problem in dynamischen, multi-institutionalen sog. *virtuellen Organisationen* anzusehen sei, wobei dem Zusammenspiel jeweils festgelegte Regeln zugrunde liegen sollen. Eine virtuelle Organisation in diesem Sinne kann den Autoren zufolge beispielsweise ein Industriekonsortium sein, das eine Machbarkeitsstudie für ein neues Überschallflugzeug durchführen soll und für eine hierfür erforderliche hochgenaue multidisziplinäre Simulation unterschiedliche Softwarekomponenten sowie Rechner- und Datenressourcen benötigt, oder ein Krisenstab, der auf verschiedenen Datenbanken und Simulationssysteme zugreift, um den geeigneten Umgang mit einer Notfallsituation zu planen.

---

<sup>185</sup> Ein Variante dieses „Durchsatzrechnens“, die allerdings nicht zum Grid-Computing im engeren Sinne zu rechnen ist, ist das *Internet-Computing*, wie es beim populären Projekt SETI@home angewendet wird ([www.seti.org](http://www.seti.org)). SETI steht für Search for Extraterrestrial Intelligence; durch Herunterladen eines Bildschirmschoners wurden auf einer Vielzahl verteilter PCs Teleskopdaten nach außerirdischen Signalen abgesucht. Das Programm wird aktiv, wenn der PC des Anwenders nicht genutzt wird; dabei gelang es, kostenlos ansonsten brachliegende Rechenkapazitäten in Größenordnungen von mehreren Teraflop/s zu aktivieren [Hoffmann et al. 2002, 4].

Da erkannt wurde, dass die umfangreichen Aufgaben, die bei der Realisierung von Grids aufgeworfen werden, nur in internationalen Kooperationen bewältigt werden können [Hoffmann et al. 2002, 10], etablierten sich unterschiedliche Foren wie das Global Grid Forum<sup>186</sup> und das europäische E-Grid-Forum<sup>187</sup>. Grids sind Thema internationaler Workshops; die entsprechenden Sammelbände geben einen Eindruck von der Vielzahl unterschiedlicher Aspekte und Herausforderungen, die im Rahmen von Grid-Entwicklungen behandelt werden (vgl. [Buyya, Baker 2000] [Lee 2001] [Parashar 2002]). Umsetzungen von Grids werden von einer Vielzahl internationaler Projekte adressiert (vgl. Tab. 5.2).

Name	Kurzbeschreibung	URL
Crossgrid	europäisches Forschungs- und Entwicklungsprojekt	<a href="http://www.eu-crossgrid.org/">http://www.eu-crossgrid.org/</a>
Data Grid	EU-Projekt unter der Leitung des CERN	<a href="http://eu-datagrid.web.cern.ch/eu-datagrid/">http://eu-datagrid.web.cern.ch/eu-datagrid/</a>
EuroGrid	EU-Projekt	<a href="http://www.eurogrid.org">www.eurogrid.org</a>
GridPP	The Grid for UK Particle Physics, britisches Grid-Projekt	<a href="http://www.gridpp.ac.uk/">http://www.gridpp.ac.uk/</a>
GriPhyN	Grid Physics Network, US-amerikanisches Grid-Projekt	<a href="http://www.griphyn.org">www.griphyn.org</a>
INFN Grid	Grid-Projekt des italienischen Istituto Nazionale di Fisica Nucleare	<a href="http://server11.infn.it/grid/">http://server11.infn.it/grid/</a>
iVDGL	International Virtual Data Grid Laboratory, US-amerikanisches Grid-Projekt	<a href="http://www.ivdgl.org/">http://www.ivdgl.org/</a>
LCG	Das LHC (Large Hadron Collider) Computing Grid-Projekt des CERN	<a href="http://lcg.web.cern.ch/LCG/">http://lcg.web.cern.ch/LCG/</a>
PPDG	Particle Physics Data Grid, US-amerikanisches Grid-Projekt	<a href="http://www.ppdg.net">www.ppdg.net</a>
UNICORE	Verbundprojekt des BMBF unter Leitung des Forschungszentrum Jülich	<a href="http://www.unicore.org">www.unicore.org</a>

Tab. 5.2 - Grid-Projekte (Auswahl).

Zu den derzeit prominentesten Grid-Projekten zählt das im September 2001 gegründete LCG (für LHC Computing Grid)<sup>188</sup> [Hoffmann et al. 2002] [Butler 2003], das in Zusammenarbeit mit anderen Grid-Projekten die verteilte Verarbeitung der erwarteten enormen Datenmengen des Large Hadron Collider (LHC)<sup>189</sup> des CERN ermöglichen soll. Das CERN, bereits an der Initiierung des World Wide Web maßgeblich beteiligt, wird 2007 mit dem LHC den weltweit leistungsfähigsten Teilchenbeschleuniger<sup>190</sup> in Betrieb nehmen. Er soll als Forschungs-werkzeug von vier großen Zusammenschlüssen von Physikern genutzt werden; insgesamt werden mehr als 5.000 Physiker aus über 50 Ländern, davon mehr als 300 Wissenschaftler aus 25 deutschen Einrichtungen, beteiligt sein<sup>191</sup>. Der LHC wird pro Jahr Daten in der Größenordnung von 12 bis 14 Petabyte erzeugen, deren Analyse das Äquivalent von rund 70.000 der heute schnellsten PC-Prozessoren erfordern wird<sup>192</sup>.

Um die Verarbeitung dieser Datenmengen zu ermöglichen und zugleich einer weltweit

<sup>186</sup> [www.Gridforum.org](http://www.Gridforum.org)

<sup>187</sup> [www.Egrid.org](http://www.Egrid.org)

<sup>188</sup> <http://lcg.web.cern.ch/LCG/>

<sup>189</sup> <http://lhc-new-homepage.web.cern.ch/lhc-new-homepage/>

<sup>190</sup> Der LHC wird Proton-Proton-Wechselwirkungen erzeugen, bei denen Protonenbündel bis zu 40 Millionen mal pro Sekunde kollidieren werden; dabei werden sich jeweils ca. 25 Wechselwirkungen mit einer Produktion von durchschnittlich 100 neuen Teilchen ergeben. Der LHC wird ferner Wechselwirkungen erzeugen, bei denen viele tausend Teilchen pro Ereignis registriert werden [Hoffmann et al. 2002, 8].

<sup>191</sup> Angaben aus [Hoffmann et al. 2002, 7ff.]; die Homepage des LCG-Projektes spricht von „some 6.000 people from universities and laboratories all around the world“ (<http://lcg.web.cern.ch/LCG/>).

<sup>192</sup> Angaben von der Homepage des LCG-Projektes (<http://lcg.web.cern.ch/LCG/>).

verteilten Nutzergemeinde zugänglich zu machen, wurde das LHC Computing Model entwickelt. In diesem hierarchischen System sind verschiedenen Ebenen (sog. Tiers, vgl. Abb. 5.5) vorgesehen.

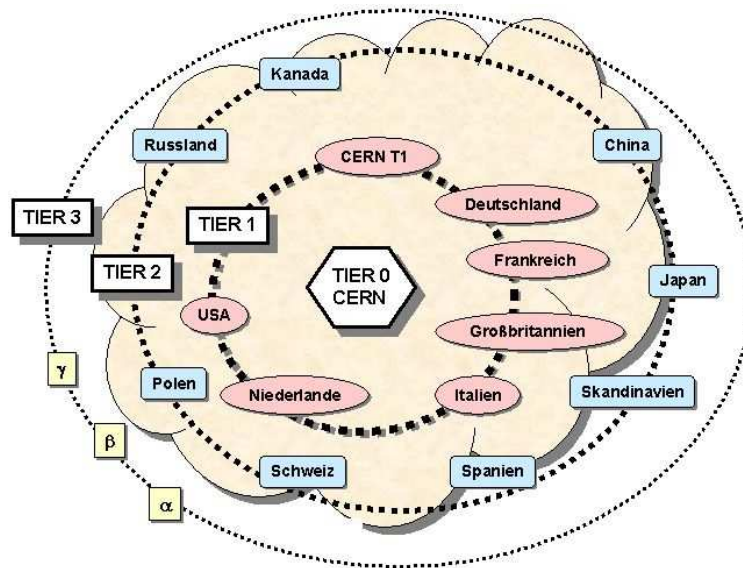


Abb. 5.5 - Das LHC Computing Model<sup>193</sup>.

Dabei soll das CERN als Tier-0-Zentrum fungieren und die Speicherung der Rohdaten und ihre Rekonstruktion übernehmen; die weiteren Ebenen sollen von regionalen Tier 1-Zentren, nationalen Tier-2-Zentren, Institutsrechnern (Tier 3) sowie lokalen Arbeitsplatzrechnern (Tier 4) gebildet werden (vgl. [Hoffmann et al. 2002]). Ziel des LCG-Projektes ist es, die hierbei auftretenden Herausforderungen durch Nutzung des Grid-Konzeptes zu bewältigen und die Kapazitäten von über Europa, Amerika und Asien verteilten wissenschaftlichen Rechenzentren zu einem virtuellen Rechenzentrum zu integrieren.

### 5.3.5 Bisheriger Stand

Trotz bisheriger Bestrebungen und Entwicklungen konnte die mit dem Konzept des Grid verbundene Vision einer unkomplizierten Verfügbarkeit verteilter IT-Ressourcen bisher noch nicht verwirklicht, sondern bestenfalls angenähert werden. [Chetty, Buyya 2002] analysieren den bisher erreichten Stand durch den naheliegenden Vergleich des Computational Grid mit dem elektrischen Stromnetz (vgl. Kap. 5.3.2). Sie konstatieren, dass die Anwendung von Grids für die Lösung von Realwelt-Problemen bisher auf Versuchslabore und eine hochspezialisierte, von Regierungen geförderte wissenschaftliche Anwendergemeinde limitiert ist und dass für eine umfassende Etablierung noch erhebliche Anstrengungen auf verschiedenen Ebenen erforderlich sind. Die Autoren stufen das Computational Grid als neues und sich entwickelndes Gebiet ein, dessen gegenwärtiger Zustand allerdings mit dem des elektrischen Stromnetzes vor annähernd einem Jahrhundert vergleichbar sei.

Wenn es gelingt, die anfallenden Herausforderungen zu bewältigen, könnte die Grid-Technologie in Zukunft so weit entwickelt werden, dass sie zunehmend transparent eingesetzt werden kann:

- ▶ „Then, eventually, in a next (and final?) step, the central Grid Engine will disappear, partly as an integrated component of the local operating systems, and partly being replaced by intelligent mobile agents, which enable a universal and self-healing environment with potentially infinite compute power available on-demand, and as easily ac-

<sup>193</sup> Abbildung erstellt in Anlehnung an [Hoffmann et al. 2002, Abb. 5].

*cessible as our today's electricity, telephony, roads and water infrastructures."*  
[Gentzsch 2000].

- ▶ „Grids will have succeeded when they become invisible to the user; when we stop talking about them” (David Williams, CERN, zitiert nach [Butler 2003]).

Wann - und ob überhaupt - eine so weitreichende Fortentwicklung erreicht werden kann, bleibt abzuwarten. Einen Prüfstein für den in naher Zukunft erreichbaren Stand von Grids wird ihr Einsatz für die rechen- und datenintensive Auswertung der LHC-Experimente ab 2007 bilden.

## 5.4 Fazit

Mit dem Internet steht eine mächtige weltumspannende Infrastruktur zur Vernetzung von Rechnern und zum Austausch von Daten zur Verfügung, die bis heute allerdings international uneinheitlich genutzt wird. Internet und World Wide Web können dabei als Annäherungen an einen „virtuellen” Computer aufgefasst werden, bei dem eine Vielzahl verteilter Rechner im Prinzip wie ein einziger Computer angesprochen und verwendet werden kann. So wurde mit dem Internet die Möglichkeit geschaffen, auf entfernte Computer zuzugreifen, ihre Rechenkapazität bei Bedarf von einem anderen Rechner aus zu verwenden sowie Daten zwischen nahezu beliebigen Rechnern auf der Erde auszutauschen. Nutzbar wurde dies durch technologische Fortschritte, die es dem einzelnen Anwender ermöglichen, von den unterliegenden Details von Verbindungsaufbau, Kommunikation zwischen Rechnern, Datentransfer etc. abzusehen und sich auf seine Aufgaben zu konzentrieren - in gewisser Weise können die über das Internet verbundenen Rechner wie ein einziges virtuelles Computersystem genutzt werden. Deutlich komfortabler wurde dieses System durch die Etablierung des World Wide Web, das als graphische Schnittstelle zum Internet aufgefasst werden kann und den Abruf weltweit verteilter, über Hyperlinks miteinander verbundener Informationsressourcen ermöglicht. Wiederum kann als einer der Erfolgskriterien ausgemacht werden, dass der Anwender nicht wissen muss, wie dies im Detail funktioniert – das World Wide Web, das ja *de facto* erst durch ein Zusammenspiel einer gigantischen Anzahl weltweit verteilter Computer, Softwaresysteme und Datenbanken entsteht und eine hochkomplexe informationstechnologische Infrastruktur voraussetzt, kann im Prinzip wie ein einziges Computersystem genutzt werden.

Eine neue Dimension der Kopplung verteilter Daten-, Hard- und Software-Ressourcen zeichnet sich gegenwärtig durch die Entwicklung des auf dem Internet aufsetzenden Grid ab, das sich allerdings noch im Entwicklungsstadium befindet. Wenn es gelingt, die vielfältigen Herausforderung, die hierdurch aufgeworfen werden, zu bewältigen, könnte so in naher Zukunft eine internetbasierte Infrastruktur bereitstehen, über die alle erforderlichen informationstechnologischen Bestandteile für die Lösung spezifischer Fragestellungen nach Bedarf flexibel bereitgestellt und verbunden werden können. Von einer solchen Entwicklung können sowohl individuelle Unternehmen wie Forschungseinrichtungen profitieren, die so Zugriff auf lokal nicht verfügbare Ressourcen erhalten können; ferner wird auf diese Weise ein anforderungsgetriebener, temporärer Zusammenschluss „virtueller” Organisationen zur Bewältigung spezifischer Aufgabenstellungen unterstützt. Es darf allerdings nicht verkannt werden, dass auch durch solche visionären Konzepte die grundlegende Herausforderung einer Integration heterogener Datenressourcen (vgl. Kap. 2) bestehen bleibt – eine Einbeziehung verschiedener Quellen setzt voraus, dass diese auch in sinnvoller Weise miteinander kombiniert werden können. Es ist daher zu erwarten, dass auch die über künftige Grids erschließbaren Datenressourcen zunächst auf solche Quellen beschränkt bleiben werden, für die eine hinreichende inhaltliche Konsistenz und Qualität sichergestellt werden kann.