

Informationsinfrastruktur im Wandel

Peter Scherber

Daten, Fakten und Software

Ein Serverkonzept für die Wissenschaft

Es wird eine heterogene Architektur von Informationsservern vorgestellt, mit der bei der GWDG Daten und Dokumente unterschiedlicher Art und Organisation für die Wissenschaften bereitgestellt werden.

Data, facts and software

A heterogenous architectural model of information servers is presented, which is under construction at the EDP center for the Science (GWDG) in Goettingen. In it we will spread multiple kinds of data and documents among the scientific community.

Données, faits et software

La contribution présente une architecture hétérogène de serveurs d'information à l'aide desquels la Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen met à la disposition des chercheurs des données et des documents.

Einleitung

Die Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen ist eine je zur Hälfte vom Land Niedersachsen und von der Max-Planck-Gesellschaft (MPG) getragene Einrichtung, die ein Hochschulrechenzentrum betreibt und sich als Kompetenzzentrum für die wissenschaftliche Datenverarbeitung versteht.

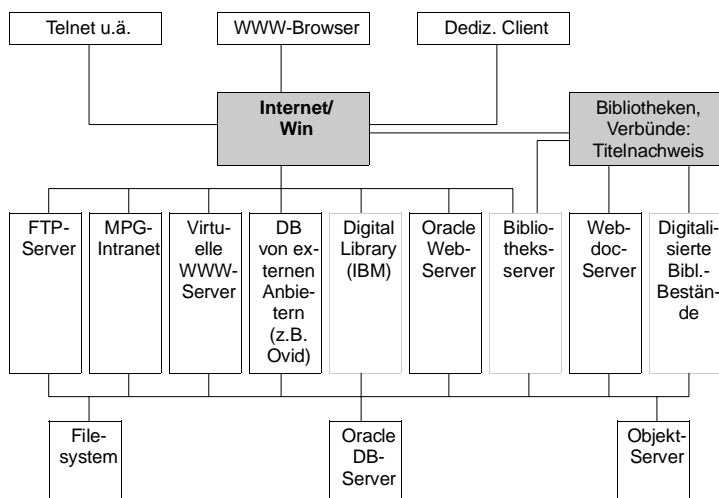
Im folgenden Beitrag wird von einem bei der GWDG im Aufbau befindlichen Serverprojekt Bericht gegeben, in dem vor allem entfernt liegenden Max-Planck-Instituten der Zugriff auf zentral vorgehaltene wissenschaftliche Information ermöglicht werden soll. Die Ausrichtung auf die Institute der Max-Planck-Gesellschaft erklärt sich daraus, daß einige dieser Serversysteme zu 100% von dieser finanziert wurden, bzw. daß auf ihnen Informationen angeboten werden, welche auf Grund eines eigenen Softwareabkommens mit einem Anbieter allein von der MPG genutzt werden dürfen. Ein ähnliches Informationsangebot kann von den lokalen Göttinger Anwendern auch über Rechner der Staats- und Universitätsbibliothek (SUB) abgerufen werden. Bei dem im folgenden vorgestellten Serverkonzept wird daher insbesondere auf das überörtlich verfügbare Informationsangebot Bezug genommen.

Angestrebt ist hierbei die Umsetzung neuer Paradigmen im Informationsgeschäft, die anderenorts eingehend diskutiert wurden¹, beispielsweise:

- Alle Informationen sollen vom PC des wissenschaftlichen Anwenders direkt erreichbar sein.

- Die Integration von Informationsnachweis und-lieferung, ohne Medienbruch.
- „just in time“, als Metapher für das verlässliche Vorhalten der Information für den Zeitpunkt, an dem sie vom Wissenschaftler gebraucht wird.
- Die Systeme sind endbenutzerorientiert, der Wissenschaftler sucht sich seine Information selbst; der Informationsvermittler als bisheriger Mediator der Recherche, wird andere Aufgaben finden müssen, z.B. als Information Broker.
- Auflösung der traditionellen Rollenverteilung (Anbieter werden zu Abnehmern und umgekehrt).
- Verteilte Information, zentraler Nachweis.

¹ Vgl. unter vielen anderen: M. Bangemann et alii: Recommendations to the European Council. Europe and the global information society (Bangemann-Bericht), Brussels 1994; R. Bayer: Plädoyer für eine Nationale Informations-Infrastruktur. In: Informatik-Spektrum (1994) 17: 302-308; A. Brüggemann-Klein et alii: Informatik und die Informationsgesellschaft der Zukunft. In: Informatik-Spektrum (1995) 18: 25-30; R. Kuhlen: Informationsmarkt – Chancen und Risiken der Kommerzialisierung von Wissen, Konstanz 1995; Digital Libraries – Communications of the ACM, April 1995; A. Brüggemann-Klein: Wissenschaftliches Publizieren im Aufbruch. In: Informatik Forschung und Entwicklung (1995) 10: 171-179.



Vorstellung des Modells (vgl. Abbildung)

Auf der in der Modelldarstellung obersten Ebene werden die Zugriffsprogramme des Anwenders nach drei Klassen unterschieden:

- Standard IP-Clients
- Dedizierte Clients
- Standardbrowser

Die Reihenfolge ist auch hier eine chronologische und der Trend, immer mehr Informationsangebote über die Standard-Internetbrowser anzubieten, wird sich wohl nicht mehr umkehren.

Auf der Anwendungsebene laufen einige der hier gezeigten Anwendungen, wie zum Beispiel der FTP-Server oder die Referenzdatenbanken des Anbieters OVID, seit mehreren Jahren mit großem Erfolg. Weitere Projekte, u.a. die Bibliotheksdatenbanken für Horizon, befinden sich in der Implementierungsphase, während die Anwendungen für „Digital Libraries“ derzeit noch in Planung sind.

Die unterste Ebene ist die der Speichermedien. Neben der Unterbringung der Dateien auf einfachen, standardisierten Filesystemen oder in speziellen sog. Objektservern steht ein leistungsfähiger Datenbankserver zur Verfügung, auf dem derzeit eine Oracle – Serverdatenbank (in Kürze wird hier auch eine weitere Sybasedatenbank eingerichtet) Dienste für die darüberliegende Applikationsebene bereitstellen kann.

Zu den einzelnen Anwendungen

Der *FTP-Server*, auch dieser hat sich längst als unverzichtbares Informationsangebot in den Hochschulen eingeführt, hält hinsichtlich Gesamtkapazität (90 GB) und monatlichem Durchsatz (ca. 300 GB) eines der umfangreichsten und vielseitigsten Angebote in Deutschland an Software und Dokumenten vor. Ca. 70% hiervon ist den beiden Unix-Plattformen Linux und FreeBSD gewidmet. Die überwiegende Menge des Angebots wird mit Hilfe von Spiegelungen der jeweiligen weltweit als einschlägig anerkannten Primärserver täglich aktuell gehalten. Daneben bietet eine Anzahl örtlicher wissenschaftlicher Institutionen auf diesem Wege ihre eigenen Materialien an.

Das *MPG-Intranet* ist ein im Aufbau befindliches Informationssystem, in welchem die Kommunikation der

MPG-Generalverwaltung mit den einzelnen Instituten, insbesondere auf dem Gebiet der Verwaltung neu organisiert werden soll. Hierbei steht die Distribution von Dokumenten im Vordergrund, z.B. von Rundschreiben. Daneben geht es um die zentrale Pflege und Fortschreibung wichtiger in den Verwaltungen konsultierter Regelwerke (Personalhandbuch, Einkaufshandbuch u.a.). Natürlich haben hier moderne Sicherheitskonzepte, wie elektronische Zugangskontrollen und ganz allgemein die Wahrung von Vertraulichkeit einen hohen Stellenwert.

Virtuelle WWW-Server. Hierbei handelt es sich um eine Art Einstiegsdienst in das World Wide Web (WWW). Instituten, die bislang, aus welchen Gründen auch immer, noch kein eigenes WWW-Angebot auf einem lokalen Rechner anbieten, wird auf einer Workstation bei der GWDG Platz für eine Webpräsentation zur Verfügung gestellt und dazu

eine sog. „virtuelle“ auf das Institut verweisende symbolische Adresse vergeben. Dem betreibenden Institut wird in der Regel, d.h. wenn es möchte und dazu in der Lage ist, dauernder Zugriff auf seinen virtuellen Server eingeräumt. Wir denken, daß die meisten Benutzer dieses Dienstes mit der Zeit so selbständig werden, daß sie nach einiger Zeit diesen Dienst auf ihre lokalen Rechner migrieren werden.

Datenbanken von externen Anbietern (z.B. OVID). Nach einem erfolgreichen bundesweiten Test vor drei Jahren werden vom amerikanischen Anbieter OVID Technologies Inc. die Datenbanken Medline, Biosis, und Current Contents unter einer modernen, auf dem Standard Z39.50 basierenden Rechercheoberfläche bereitgestellt. Im Jahre 1998 treten hierzu noch eine Zahl weiterer international renommierter Referenz-Datenbanken. An einem Stichtag Anfang September 1997 wurden im Ovidserver die folgenden Benutzungswerte gemessen:

1900 Benutzer aus 130 Instituten und Arbeitsgruppen der MPG haben Zugriffsberechtigung.

50-60 sind gleichzeitig im System aktive Benutzer.

Es fanden 747 Recherchesitzungen statt, davon

67 – d.h. unter 10% – über den Standard-Telnet-Client,

260 – d.h. mehr als 30% – über einen von OVID bereitgestellten Windows-Client,

420 – d.h. mehr als 50% – über den WWW Standardbrowser, obwohl dieser nicht den vollen Recherche-Komfort bietet wie der Windows-Client.

In der Planung sind verschiedene Datenbanken auf Basis der von IBM angebotenen *Digital Library*. Dieses Programmsystem ist auf den Plattformen Windows NT und AIX einsetzbar und eignet sich im besonderen Maße für multimediale Objekte. Diese werden in einem besonderen Objektserver gespeichert und können auch in Zusammenarbeit mit einer Oracle-Datenbank verwaltet werden. Geplant sind hier Bilddatenbanken für verschiedene Archive.

Unter *Diverse Kundendatenbanken* verstehen wir mehrere im Aufbau bzw. noch in der Planung befindliche kleinere und mittlere Vorhaben, in denen die GWDG als Betreiber eines Servers tätig werden kann. Hierunter verbergen sich so heterogene Projekte wie z.B.:

- Publizieren aus einer Dokumentdatenbank unter SGML: Dabei sollen unterschiedliche Dokumentrepräsentationen generiert werden (CD, Buch, Webseiten).

– Weltweite Bereitstellung von Experimentdaten in der Biologie.

– Zeitschriftennachweis in den Bibliotheken der MPG. Ein *Bibliotheksserver* für das Bibliotheksverwaltungssystem Horizon wird gerade aufgebaut. In diesem Rahmen werden anfänglich etwa 15 über die Bundesrepublik verteilte Institutsbibliotheken ihre Bestände auf einem Server bei der GWDG halten und pflegen.

Auch der von uns in Kooperation mit der SUB Göttingen betriebene Dokumentserver im *WebDoc*-System, über den an anderer Stelle berichtet wird, findet hier seinen Platz im Kreis der Informationsserver.

Schließlich werden auch *digitalisierte Bestände* von Bibliotheksmaterialien auf Servern der GWDG vorgehalten. Ein bereits begonnene Projekt erfaßt einen rechts-historischen Bestand von ca. 4000 Dokumenten aus dem Privatrecht des Max-Planck-Instituts für europäische Rechtsgeschichte. Die Daten werden als Volltext-Images erfaßt und in den Katalogen der Bibliotheksverbände nachgewiesen. Bei den Katalogsnachweisen finden sich Links auf die Serveradressen, unter denen die Bestände auf unserem Server im Volltext und über das WWW einzusehen sind.

Allgemeine Rahmenbedingungen des Serverkonzepts

Die Daten der hier vorgestellten Informationsserver werden überwiegend dezentral durch die Benutzer selbst oder durch Dritte (z.B. kommerzielle Anbieter) bereitgestellt. So gesehen ist die GWDG niemals Produzent, sondern immer nur der Distributor der Informationen.

Dies differenziert sich für Erschließung und Nachweis der Informationen, an denen praktisch alle Beteiligten teilnehmen können. Je nach Ausrichtung des Angebots sind ein oder mehrere Instanzen hieran beteiligt:

- der kommerzielle Anbieter;
- das Institut oder der Wissenschaftler als Initiator oder Betreiber der Datenbank;
- Bibliotheken und Verbände;
- allgemeine Internet-Nachweisinstrumente, z.B. Suchmaschinen;
- die GWDG als Distributor, bei der eigene serverspezifische Suchhilfen eingesetzt werden.

Die Bereitstellung für und Nutzung durch den Wissenschaftler muß sich natürlich auch an vertraglichen Regelungen und an den Wünschen der Informationsanbieter ausrichten. So sind als Rezipienten im allgemeinen zwar alle Wissenschaftler weltweit vorgesehen, die über einen Standard-WWW-Browser das Angebot wahrnehmen können. Es ist aber in einzelnen Fällen aus lizenzrechtlichen und anderen Gründen vorzusehen, daß nur bestimmte Rezipientengruppen auf die für sie bestimmten Informationen zugreifen können:

- im Intranet der MPG sollen auf die Kernbereiche nur die einzelnen Max-Planck-Institute zugreifen;

– auf den OVID-Server ist aus lizenzrechtlichen Gründen der Zugriff nur Wissenschaftlern der MPG nach persönlicher Anmeldung über die GWDG erlaubt;

– einige Angebote sind nur für die lokalen wissenschaftlichen Kunden der GWDG (Universität Göttingen und Göttinger Max-Planck-Institute) kostenfrei erhältlich (z.B. bestimmte Online-Angebote kommerzieller Zeitschriftenverlage in WebDoc);

– einige Bestände sind nur für die Angehörigen eines einzelnen Instituts bestimmt (z.B. Horizon-Bibliotheksdaten).

Was die Archivierung und Bestandssicherung angeht, sind derzeit nur EDV-typische Sicherungs- und Backupverfahren im Einsatz. Über geeignete Archivierungskonzepte, insbesondere zur langfristigen Bestandssicherung muß noch nachgedacht werden.

Die GWDG wird – und dies ist schon anhand der vorgestellten heterogenen Systemarchitektur nicht zu leisten – nicht Informationsvermittlung alter Art anstreben. Wir denken, daß wir statt dessen im wesentlichen auf drei Gebieten aktiv tätig sein werden:

- in der Unterstützung – ggf. auch nur während der Initialphase – der einzelnen Projekte durch Anwendungsprogrammierung und Bereitstellung von Kompetenz im Web-Publishing;
- bei der Systemschulung, sofern dies nicht von Dritten ausreichend und kostengünstig angeboten wird;
- im Hineinwachsen in Broker- und „Pfadfinder“-dienste, die vermutlich in enger Zusammenarbeit mit den Bibliotheken zu entwickeln sind.

Abschließend sollte auf mögliche Einwände eingegangen werden, der Betrieb derartiger zentraler Server könne in einer Zeit verteilter Systeme auch als eine Art Rückschritt, ja eine Rückkehr zu einem neuen Zentralismus bedeuten.

Diese Dienste bei der GWDG *können* genutzt werden, wenn ein zentrales Anbieten der Information z.B. wegen günstiger Lizenzabkommen oder wegen fehlender Kompetenz vor Ort wirtschaftlich geboten erscheint. Dabei ist uns klar, daß Dienste, die der Markt nicht annimmt, keine oder keine lange Lebensdauer haben werden. Wir denken aber, daß gerade die Vielfalt des derzeit vorgesehenen Angebots und die Bereitschaft, flexibel dieses Angebot nach den Erfordernissen des Informationsmarktes auszuweiten, auch in einer sehr dynamisch sich entwickelnden Informationslandschaft Chancen haben wird.

Anschrift des Autors:

Prof. Dr. Peter Scherber
Gesellschaft für wissenschaftliche
Datenverarbeitung mbH
Postfach 28 41
D-37018 Göttingen
E-Mail: pscherb@gwdg.de