Umweltdaten im Internet: Aufbau eines öffentlichen WWW-Servers

Ralf Kramer, Ralf Nikolai
Forschungszentrum Informatik (FZI)
Haid–und–Neu–Str. 10–14, D–76131 Karlsruhe, Germany
{kramer, nikolai}@fzi.de

Horst Spandl
Landesanstalt für Umweltschutz (LfU), ITZ Stuttgart,
Spittlerstr. 8, D–70190 Stuttgart, Germany
Horst.Spandl@lfust.um.bwl.de

1 Einführung

1.1 Ausgangsbasis

Im Umweltinformationssystem (UIS) Baden-Württemberg werden z.Z. in mehreren Projekten die Techniken und Werkzeuge des World Wide Web (WWW) eingesetzt [Mayer-Föll 95]. Mit diesen Techniken können sowohl klassische Berichte zugänglich gemacht als auch online-Zugriffe auf Datenbanken ermöglicht werden. Die entwickelten Systeme werden zum einen im interen Verwaltungsnetz (LVN), dem sog. Verwaltungs-Intranet, zum anderen auch auf dem öffentlich zugänglichen Internet eingesetzt.

Die öffentlich zugänglich zu machenden Informationen können sowohl als statische HTML-Seiten als auch als aus Datenbanken zu generierende (dynamische) HTML-Seiten vorliegen. Wesentliche Anforderungen bei der Einrichtung des öffentlichen WWW-Servers sind

- ein möglichst wartungsarmer Betrieb des öffentlichen Servers,
- die Berücksichtigung organisatorischer Anforderungen (beispielsweise WWW-Server für einzelne Fachabteilungen),
- die Berücksichtigung unterschiedlicher Anforderungen (beispielsweise hinsichtlich Aktualität, Änderungshäufigkeit und Datenvolumen) sowie
- die Beachtung von Sicherheitsanforderungen bei der Internetanbindung.

1.2 Überblick

Das vorliegende Papier basiert auf [Kramer 95b]. Im folgenden Abschnitt 2 wird kurz die Netzstruktur vorgestellt. Das darauf folgende Kapitel 3 zeigt, wie statische HTML-Seiten, die zunächst auf mehreren internen WWW-Servern liegen, auch auf dem öffentlichen WWW-Server verfügbar gemacht werden. Der Abschnitt 4 widmet sich der Frage, wie Datenbanken öffentlich zugänglich gemacht werden. Das Papier schließt mit einem kurzen Resümee.

2 Netzkonfiguration

Das ITZ Stuttgart der Landesanstalt für Umweltschutz betreibt den zentralen Internetzugang für das Ministerium für Umwelt und Verkehr Baden-Württemberg. Das Internet und das interne Netz sind durch eine Firewall getrennt. Die implementierte Sicherheitsphilosophie sieht vor, daß aus Sicherheitsgründen in der Regel keinerlei Verbindungen vom externen Netz in das verwaltungsinterne Intranet zugelassen werden. Ausnahmen müssen bewußt freigeschaltet werden, wie z.B. der Mail- und Newsfeed vom Internet Service Provider. Durch den gezielten Einsatz von Application-Gateways und Proxy-Servern wird die Nutzung von Internetdiensten für Nutzer des UM-Intranet kontrolliert ermöglicht.

In verschiedenen Projekten wurde die Eignung von Internetdiensten, insbesondere die des World Wide Web (WWW), als Zugangsmechanismus für Umweltinformationen gezeigt [Mayer-Föll 95]. Ziel war dabei zunächst der rein verwaltungsinterne Einsatz. Schon bald kam jedoch der Wunsch auf, die zur Veröffentlichung geeigneten Teile auch einer breiten Öffentlichkeit zugänglich zu machen. Hierzu wurde die in Abbildung 1 dargestellte Architektur umgesetzt.

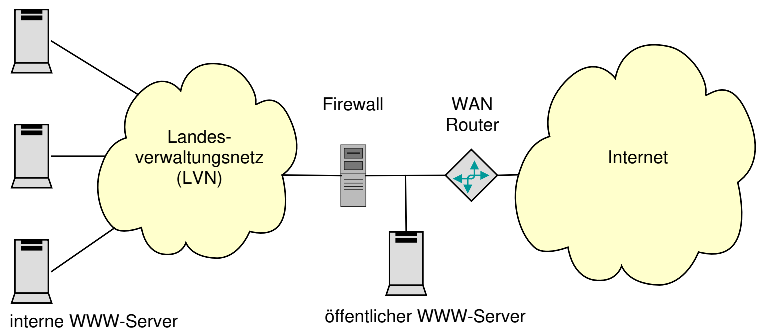

Abbildung 1: Netzkonfiguration, öffentlicher WWW-Server.

Vor der Firewall bedient eine eigene Servermaschine Anfragen aus dem Internet (`http://www.uis-extern.um.bwl.de`). Die dort veröffentlichten Informationen sind Kopien von Daten, die von verschiedenen internen Servermaschinen stammen. Ein Zugriff auf interne WWW-Server durch die Firewall findet nicht statt.

Der Betrieb des öffentlichen Servers erfolgt weitestgehend automatisch. Die zur Veröffentlichung vorgesehenen Informationen werden durch Batchprozeduren auf den externen WWW-Server kopiert. Die Behandlung von Informationen, die als statische HTML-Dokumente vorliegen (siehe Kapitel 3), stellt dabei ganz andere Anforderungen als die Vorgehensweise bei dynamisch generierten HTML-Dokumnten, die dynamisch über Zugriffsprogramme mittels des Common Gateway Interface (CGI) zur Laufzeit erzeugt werden (siehe Kapitel 4).

3 Statische HTML-Seiten

Die für die Publikation im Internet vorgesehenen HTML-Seiten müssen sich auf dem öffentlichen WWW-Server vor der Firewall befinden. Das Problem besteht nun darin, die zu veröffentlichenden Dokumente von mehreren behördeninternen WWW-Servern auf den externen WWW-Server zu bringen und die zunächst im behördeninternen Netz vorgenommenen Änderungen regelmäßig nachzuführen. Hierbei handelt es sich um das Hinzukommen neuer, das Modifizieren sowie das Löschen vorhandener Seiten.

Die drei folgenden Alternativen wurden evaluiert:

- Die zu veröffentlichenden Seiten werden auf den jeweiligen internen Servern in speziellen Verzeichnissen abgelegt und von dort aus periodisch mit dem öffentlichen WWW-Server abgeglichen. Dort erfolgen Überprüfungen und Modifikationen der Seiten.

- Die Seiten der internen Server werden zunächst auf einem internen Master-Server gesammelt und von dort aus wie bei Alternative 1 auf den externen Server gespiegelt.

- In einer Unix-Umgebung wird das eigentlich externe Dateisystem des öffentlichen WWW-Servers vor der Firewall auch nach innen gemountet. Im Gegensatz zu den beiden zuvor skizzierten Alternativen exitieren die HTML-Seiten damit nur einmal.

3.1 Konzept und Realisierung

Nach Abwägung der Vor- und Nachteile (z.B. Datensicherheit, Realisierbarkeit nicht öffentlicher Testläufe, Kopiervorgänge über das Netz, automatische Überprüfbarkeit vor Veröffentlichung auf korrekte HTML-Syntax und Links, notwendige Abstimmung interner Server, Aufwand für Gewährleistung der Konsistenz der internen und des externen Servers, Belastung der Server, einheitlicher Mechanismus) wurde Alternative 1 gewählt und auf Unix-Plattformen mittels automatisierter Skripte realisiert.

Bei Bedarf ist ein stufenweiser Übergang zur Variante 2 jederzeit möglich. Die Alternative 3 wurde vor allem aus Sicherheitsgründen verworfen, da hierbei die Firewall für UDP- bzw. TCP-Pakete des Network File System (NFS) hätte geöffnet werden müssen. Abbildung 2 zeigt die implementierte Struktur.

Auf jedem der internen WWW-Server existiert ein spezielles Directory www_public unterhalb der DOCUMENT_ROOT des WWW-Servers, das die zur

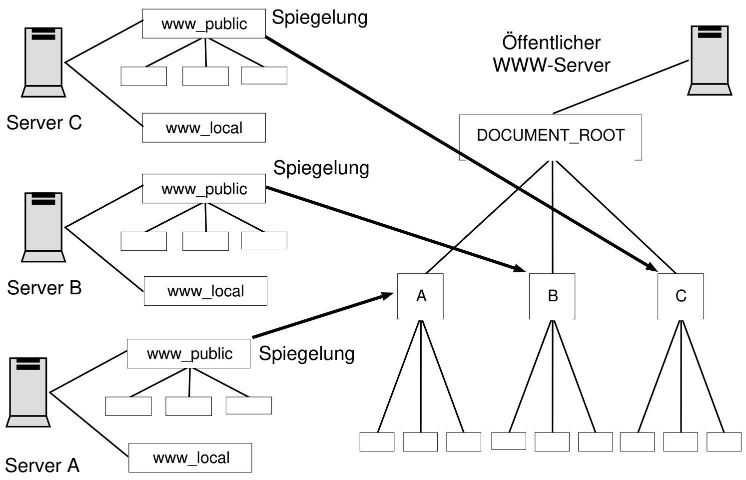

Abbildung 2: Spiegelung statischer HTML-Seiten auf den externen WWW-Server.

Veröffentlichung vorgesehenen Informationen enthält. Informationen, die nicht für die Öffentlichkeit bestimmt sind, müssen in einem anderen Zweig des Dateisystems gehalten werden. Auf dem öffentlichen WWW-Server gibt es ein Directory unter der dortigen DOCUMENT_ROOT für jeden der internen WWW-Server, das den entsprechenden www_public-Teil aufnimmt. Das Konzept sieht vor, daß die Inhalte des externen WWW-Servers ein exaktes Spiegelbild der Informationen der jeweiligen internen Daten darstellen:

- Wird intern eine Datei erzeugt oder modifiziert, wird sie nach außen kopiert.
- Wird intern eine Datei gelöscht, wird sie außen ebenfalls gelöscht.

Mit Hilfe des frei verfügbaren Programms `rdist` werden die HTML-Seiten in ein spezielles Spiegelungsdirectory auf dem öffentlichen WWW-Server kopiert. Gestartet wird die Spiegelung durch einen cron-Job auf dem jeweiligen internen WWW-Server. Um die Erstellungszeiten konsistent zu halten, werden automatische Änderungen an den HTML-Seiten in einem speziellen Testdirectory auf dem externen Server durchgeführt. Dort wird z.B. ein Filterskript gestartet, das automatisch alle URLs in den HTML-Dokumenten an die Situation auf dem öffentlichen Server anpaßt. Des weiteren können HTML-Syntaxüberprüfungen und Konsistenzchecks durchgeführt werden. Erst danach werden die HTML-Dokumente aus dem Testbereich in das via WWW zugreifbare Directory unter der DOCUMENT_ROOT kopiert.

Bei der Veröffentlichung von Umweltinformationen wird das World Wide Web wie jedes andere Medium behandelt. Dadurch greifen automatisch alle behördeninternen Regelungen zur inhaltlichen Vorbereitung und Abstimmung der Inhalte. Die einzelnen Fachabteilungen der LfU haben eigene Teilbereiche im WWW-Baum, die sie eigenverantwortlich mit Inhalten füllen. Das Informationstechnische Zentrum

(ITZ) stellt die Infrastruktur, hilft bei der Erstellung der allgemeinen Gestaltungsrichtlinien (HTML-Seitenrahmen, Sammlung von Icons und Logos etc.) und berät bei der Erstellung der eigentlichen WWW-Seiten (HTML-Editoren, Konvertierungstools für Texte und Bilder etc.). Nur durch diese Aufgabenteilung war es möglich, innerhalb einer relativ kurzen Zeitspanne ein umfangreiches Informationsangebot zu erstellen und zu betreuen.

3.2 Diskussion

Das vorgestellte Konzept ermöglicht einen weitestgehend automatisierten Betrieb eines öffentlichen WWW-Servers, der seine Daten durch eine Firewall hindurch von verschiedenen im Intranet angesiedelten WWW-Servern zugestellt bekommt. Die Erstellung und der Test von Dokumenten kann zunächst lokal erfolgen. Der öffentliche Server kommt systemseitig mit einer minimalen Softwareausstattung aus. In unserem Falle ist z.B. auf dem externen Rechner noch nicht einmal ein C-Compiler installiert. Die Zugangsberechtigungen können äußerst restriktiv gehandhabt werden. Sollte der öffentlich zugängliche Rechner doch einmal korrumpiert werden, so stehen alle Daten nochmal vollständig im internen Netz zur Verfügung.

4 Dynamisch erzeugte HTML-Seiten

Neben Berichten und Dokumentationen existiert eine Vielzahl von Umweltinformationen, die in Datenbanksystemen (DBS) bei der Landesanstalt für Umweltschutz und anderen Behörden gehalten werden. Es werden eine Reihe unterschiedlicher DBS betrieben, die sich unterscheiden u.a. hinsichtlich

- technischer Plattformen (Hardware- und Betriebssystemplattformen, auf denen die DBS installiert sind),
- der unterstützten Datenmodelle und Zugriffsfunktionalitäten (Oracle als relationales, Adabas als prärelationales DBS und Smallworld als Geographisches Informationssystem (GIS)) sowie
- der Datenbestände (Inhalte), die in den einzelnen Systemen verwaltet werden, und damit hinsichtlich der Anforderungen an einen WWW-Zugang.

4.1 Anforderungen

An einen WWW-Zugriff auf Datenbanken werden unterschiedliche Anforderungen gestellt, die sich u.a. in den folgenden Kriterien unterscheiden:

- Aktualitätsanforderungen: Daten, die sich häufig ändern und bei denen zugleich die jeweils aktuellsten Daten öffentlich zugänglich sein sollten, stellen härtere Anforderungen als beispielsweise statistische Jahresmittelwerte, die nur einmal jährlich neu eingespielt werden.

- Zugriffsfunktionalität: Die erforderliche Zugriffsfunktionalität kann je nach Anwendung ganz unterschiedlich sein. Im Fall der Hochwasservorhersage beispielsweise sind für sämtliche Anwender primär die akuellsten verfügbaren Pegelstände sowie Vorhersagen von Interesse. Im Gegensatz dazu wird eine Recherche in den Metadatenbeständen des Umweltdatenkatalogs UDK [Swoboda 95, Lessing 96], in dem beschreibende Informationen über Umweltobjekte gesammelt werden, in der Regel sehr benutzerspezifisch sein.

- Datenvolumen: Datenbanken eines vergleichsweise geringen Umfangs von nur wenigen MBytes wie beispielsweise der UDK können im Gegensatz zu sehr umfangreichen Datenbanken wie der des Meßreihenoperationssystems MEROS, das alle bisher erhobenen Meßdaten aus den verschiedenen landesweiten Meßnetzen enthält, prinzipiell problemlos repliziert, d.h. in einer Kopie, angelegt werden.

- Vertraulichkeit der Daten: Daten in einer Datenbank, die ausschließlich für den verwaltungsinternen Gebrauch bestimmt sind, dürfen nicht frei zugänglich gemacht werden. Für Datenbanken, die neben Daten, die ohne Einschränkungen auch öffentlich zugänglich gemacht werden können, auch vertrauliche Daten enthalten, muß anhand der Dateninhalte differenziert werden.

4.2 Realisierungsalternativen

Diese aufgeführten Beispiele illustrieren, daß unterschiedliche Datenbestände unterschiedliche Anforderungen an einen öffentlichen WWW-Zugang stellen. Für die Realisierung von Zugriffen auf die Inhalte von Datenbanken aus dem (öffentlichen) Internet gibt es im wesentlichen die folgenden Realisierungsalternativen:

1. Die Daten können aus dem Datenbanksystem periodisch extrahiert werden. Auf dieser Basis werden statische HTML-Seiten generiert, mit denen wie in Abschnitt 3 beschrieben verfahren werden kann. Diese Vorgehensweise bietet sich vor allem dann an, wenn Standardfragestellungen formuliert werden können, z.B. die aktuellen Hochwasservorhersagen oder die Ozonwerte der letzten 3 Tage.

2. Es wird ein sicherer Durchgriff auf interne Datenbankserver hinter einer Firewall realisiert. Hierfür existieren u.a. folgende Varianten:

   - Datenbankzugriffsprogramme werden auf dem externen WWW-Server ausgeführt und greifen durch die auf entsprechender Kommunikationsprotokollebene freigeschaltete Firewall auf das interne DBS zu.
   - Ein Proxy-Server vor der Firewall leitet alle Anfragen, die mit Datenbankzugriffen verbunden sind, an einen internen WWW-Server weiter.

3. Ferner besteht die Möglichkeit, zusätzlich zum internen Datenbankserver einen externen Datenbankserver vor der Firewall zu installieren. Die externe Datenbank enthält gezielt erzeugte Replikationen von internen Datenbeständen.

4.3 Umsetzung im UIS Baden-Württemberg

Für unterschiedliche Datenbanken der LfU wurden entsprechend den jeweiligen Anforderungen unterschiedliche Lösungen gewählt und zum Teil bereits auf der Grundlage der Software-Architektur von WWW-UIS [Koschel 95, Koschel 96] umgesetzt.

Beim Betrieb des öffentlichen WWW-Servers wird derzeit die periodische Extraktion von Dateninhalten bevorzugt (siehe auch 4.2). Begünstigt wird dies durch die Möglichkeit, auf vorhandene Programme zurückgreifen zu können, die bisher bereits Informationen für andere elektronische Medien wie BTX oder Videotext aufbereitet haben. Diese Programme müssen lediglich um eine weitere Ausgabeoption erweitert werden. Daß diese Art des Datenzugriffs sich aus Sicht des WWW-Servers nicht von HTML-Seiten aus Textdokumenten unterscheidet, ist ein weiterer Vorteil für den Betrieb angesichts knapper Personalressourcen. Aus Sicherheitsgründen wäre eine umfassende Prüfung der für den öffentlichen WWW-Server vorgesehenen CGI-Programme unausweichlich. Ausgefeilte Datenzugriffe stehen derzeit nur Nutzerinnen und Nutzern innerhalb des Verwaltungs-Intranets zur Verfügung.

Beispiele für die periodische Extraktion und Aufbereitung als statische HTML-Dokumente sowie deren Spiegelung auf den öffentlichen WWW-Server sind Daten der Hochwasservorhersage (`http://www.uis-extern.um.bwl.de/lfu/hvz/welcome.html`) und ausgewählte Luftmeßdaten (`http://www.uis-extern.um.bwl.de/lfu/abt3/luftdat.htm`, `http://www.uis-extern.um.bwl.de/lfu/abt3/ozon.htm`).

Der UDK, der in seiner WWW-basierten Version [Kramer 96a, Kramer 95a, Kramer 96b] (`http://www.uis-extern.um.bwl.de/cgi-bin/www-udk/umlenkung.sh`) zentrales Navigations- und Recherche-Instrument im UIS Baden-Württemberg ist, wurde – wie bereits in Österreich (`http://udk.bmu.gv.at`) – basierend auf einer zusätzlichen Datenbank-Installation auf dem öffentlichen WWW-Server zugänglich gemacht. Eine der WWW-UDK 3.0 Tabellen [Kramer 96b] enthält u.a. URLs. Beim Transfer dieser Tabelle von einem internen WWW-Server auf den öffentlichen WWW-Server ist zu beachten, daß die URLs, die auf interne WWW-Seiten verweisen, zu löschen sind, sofern diese internen Seiten nicht auch auf dem externen WWW-Server verfügbar sind. Sofern die Seiten auch auf dem öffentlichen WWW-Server unter einer anderen URL zur Verfügung stehen, ist eine entsprechende Anpassung erforderlich.

5 Resümee

In diesem Papier wurde der Aufbau eines öffentlichen WWW-Servers für Umweltdaten vorgestellt. Dieser macht sowohl statische HTML-Seiten als auch Datenbanken im Internet verfügbar. Für statische HTML-Seiten konnte eine für alle Seiten einheitliche Vorgehensweise entwickelt werden, durch die WWW-Seiten von internen Servern auf den öffentlichen Server gespiegelt werden. Datenbanken werden mit vom Einzelfall abhängigen Techniken der Öffentlichkeit zur Verfügung gestellt. Die skizzierten Vorgehensweisen für statische und dynamische HTML-Seiten befinden sich bereits im Einsatz.

Die Hauptvorteile des vorgestellten Verfahrens besteht darin, daß der öffentliche WWW-Server weitestgehend automatisch mit geringem Wartungsaufwand betrieben werden kann. Der öffentliche WWW-Server ist zudem durch eine Firewall vom internen Netz getrennt und bekommt seine Daten über Batchprozeduren von internen WWW-Servern übertragen. Es erfolgt kein Zugriff von außen durch die Firewall auf Rechner des Intranet. Die HTML-Dokumenterzeugung und -überprüfung erfolgt zunächst intern. Wird eine Teilpublikation zur Veröffentlichung freigegeben, steht sie automatisch zur internen wie auch zur externen Nutzung zur Verfügung.

Danksagungen: Bedanken möchten wir uns bei Andreas Rüping und Torsten Auerbach, FZI, für die Mitarbeit bei der Veröffentlichung statischer HTML-Seiten.

Literatur

[Huber-Wäschle 95] F. Huber-Wäschle, H. Schauer, P. Widmayer, editors. *Herausforderungen eines globalen Informationsverbundes für die Informatik; 25. GI-Jahrestagung und 13. Schweizer Informatikertag, Zürich, 18.-20. September 1995/GISI 95*, Informatik Aktuell. Springer, 1995.

[Koschel 95] A. Koschel, R. Kramer, R. Nikolai. Architektur des WWW- und CORBA-basierten UIS. In R. Mayer-Föll, A. Jäschke, editors, *Projekt GLOBUS; Konzeption und prototypische Realisierung einer aktiven Auskunftskomponente für globale Umwelt-Sachdaten im Umweltinformationssystems Baden-Württemberg; Phase II 1995*, number FZKA 5700 in Wissenschaftliche Berichte, pages 39–45. Forschungszentrum Karlsruhe Technik und Umwelt, Karlsruhe, Dec. 1995.

[Koschel 96] A. Koschel, R. Kramer, R. Nikolai, W. Hagg, J. Wiesel. A Federation Architecture for an Environmental Information System incorporating GIS, the World-Wide Web, and CORBA. In *Third International Conference/Workshop Integrating GIS and Environmental Modeling*, Santa Fe, New Mexico, USA, Jan. 1996.

[Kramer 95a] R. Kramer, H. Spandl. Metadatenzugriff in Weitverkehrsnetzen: Eine Realisierung am Beispiel des Umweltdatenkatalogs UDK. In Huber-Wäschle et al. [Huber-Wäschle 95], pages 610–617.

[Kramer 95b] Ralf Kramer, Ralf Nikolai, Andreas Rüping. Aufbau eines öffentlichen WWW-Servers beim Umweltministerium Baden-Württemberg, Stuttgart; Organisatorische Rahmenkonzeption. FZI Studie, Karlsruhe, Germany, Dec. 1995.

[Kramer 96a] R. Kramer, T. Quellenberg. Global Access to Environmental Information. In R. Denzer, D. Russel, G. Schimak, editors, *Environmental Software Systems; Proceedings of the International Symposium on Environmental Software Systems, 1995*, International Federation for Information Processing (IFIP), pages 209 – 218, London, 1996. Chapman and Hall.

[Kramer 96b] Ralf Kramer, Ralf Nikolai, Andree Keitel, Rudolf Legat, Konrad Zirm. Enhancing the Environmental Data Catalogue UDK for the World Wide Web. In *10. Symposium Informatik für den Umweltschutz*, Sept. 1996. to appear.

[Lessing 96] H. Lessing, W. Swoboda, O. Günther. UDK: A European Environmental Data Catalogue. In *Third International Conference/Workshop Integrating GIS and Environmental Modeling*, Santa Fe, New Mexico, USA, Jan. 1996.

[Mayer-Föll 95] R. Mayer-Föll, A. Jäschke, editors. *Projekt GLOBUS; Konzeption und prototypische Realisierung einer aktiven Auskunftskomponente für globale Umwelt-Sachdaten im Umweltinformationssystems Baden-Württemberg; Phase II 1995.* Number FZKA 5700 in Wissenschaftliche Berichte. Forschungszentrum Karlsruhe Technik und Umwelt, Karlsruhe, 1995.

[Swoboda 95] W. Swoboda, H. Lessing, P. Grolimund, O. Günther, U. Haas, R. Legat, M. Vogler, K. Zirm. Metadatenklassen im Umweltdatankatalog (UDK). In Huber-Wäschle et al. [Huber-Wäschle 95], pages 601–609.