# Building Infrastructures for Digital Libraries

A. Bartelt[1], D. Faensen[2], L. Faulstich[2], E. Schallehn[3], C. Zirpins[1]

**Abstract**

Digital Libraries today are often monolithic systems. In the future, they will dissolve into collections of electronic services. The challenge will be to provide an infrastructure that supports the user in dealing with this multitude of services. Such an infrastructure should offer integrated access to the combined contents of multiple services, it should provide active dissemination of new contents, and it needs to support the users in locating and combining the services most suitable to their needs.

In the *Global Info* program *Infrastructures for Digital Libraries* components of such an infrastructure are being developed. The federated query service *DEMETRIOS* and the alerting service *HERMES* are both integration services that combine underlying services, i.e., heterogenous information sources. The *GIBRALTAR* portal provides a meta-service that supports the user in locating and applying various Digital Library services.

## 1 Introduction

We expect that Digital Libraries of the future will be comprised of a large number of heterogeneous services offered by various parties involved in the production, distribution, and archival of digital contents. For the user it becomes increasingly difficult to deal with a multitude of heterogeneous services with different user interfaces and access mechanisms, and requiring separate registrations. Hence the goal is to facilitate the interaction of the user with a Digital Library consisting of a set of heterogeneous services offered by independent providers. In particular, the user needs support in three fields: First, users have to locate and combine appropriate Digital Library services, including trading and negotiation of service properties such as prices, licenses etc. Second, users want to search and query over multiple information sources. Third, users want to stay up-to-date on relevant content of the Digital Library. In this paper, we present three projects that address these issues. These projects constitute the program *Infrastructures for Digital Libraries* within the German Digital Library initiative *Global Info* [Glo01].

The *DEMETRIOS* federation service for bibliographical metadata is being developed at University of Magdeburg. It supports the flexible integration of bibliographical data from multiple information sources of various types. The unified query service offered by DEMETRIOS can be utilized by other Digital Library services such as HERMES.

The *HERMES* project at Free University Berlin has developed an alerting service for Digital Libraries [FFS+01]. Researchers who have entered their profile of interests into the HERMES

[1] {bartelt,zirpins}@informatik.uni-hamburg.de
[2] {faensen,faulstic}@inf.fu-berlin.de
[3] eike@iti.cs.uni-magdeburg.de

system (http://hermes.inf.fu-berlin.de) are notified about matching bibliographical references harvested from a variety of heterogeneous underlying information sources. This means that researchers can delegate the tedious task of continuously scanning all available information sources for relevant publications to HERMES.

Within the *GIBRALTAR* project at the University of Hamburg, a comprehensive user-centered integration platform for electronic publication services is being created. The GIBRALTAR project provides fundamental infrastructure mechanisms comprising facilities for typing, trading, negotiation, and integration. These mechanisms constitute a broker for Digital Library services such as HERMES or DEMETRIOS, that are made available to the user through this portal [ZWBW01].

The developed functionalities may be used as infrastructures for various cases in the domain of digital libraries. Together, they constitute an integration environment for multiple aspects of information services. To give an introductive example, consider a scholar who needs information on a certain research topic. Entering the GIBRALTAR portal, (s)he decides to search catalogs and selects the service type catalog. The portal suggests a set of various catalogs. Since a federated query service such as DEMETRIOS offers the functionality of a catalog, it is included as well. The researcher chooses DEMETRIOS because of its wide coverage and is thus forwarded to a search interface for DEMETRIOS. After posing his/her search requests and viewing the results of DEMETRIOS, the researcher finds a topic of interest about which (s)he wants to stay informed, so (s)he asks for an alerting service that covers the information sources relevant to this topic. The portal suggests the HERMES alerting system. After entering an appropriate search request, HERMES will start watching various sources for matching publications and will notify the researcher about positive results.

In the remainder, the interoperability issues within the projects DEMETRIOS, HERMES and GIBRALTAR arising from this scenario will be outlined.

# 2 Interoperability Issues

The domain of electronic publishing, as well as its sub-domain of virtual digital libraries, consists of many different application-level services including document repositories, catalogs, retrieval engines and alerting facilities. Today such types of services are mostly accessed via the Internet, where typically each one is offered by a different provider.

To constitute the functionality of a Digital Library in a multi-service environment, a specific set of service types has to be chosen, for which the best possible providers must be found and homogeneously integrated. While the identification of a set of interoperable services is done by the GIBRALTAR web portal, integration of different data sources is provided by services such as HERMES or DEMETRIOS.

## 2.1 Federating Bibliographical Databases

The DEMETRIOS federation service [EHS+00] mediates queries on various bibliographical data sources by offering extended query processing capabilities on a global view over the integrated data. It can be used to support simple search scenarios on publications, as well as other services requiring access to integrated, more complex data, like citation indexes [SES00].

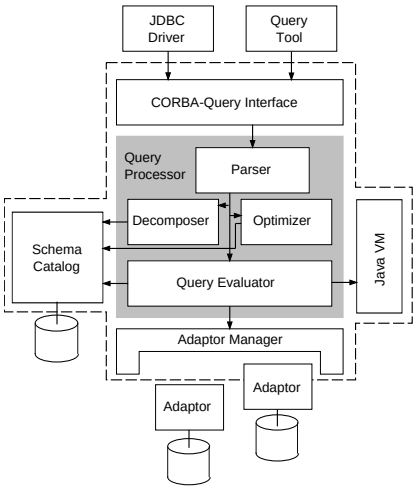


Figure 1: Architecture of the DEMETRIOS federation service

DEMETRIOS promotes the interoperability with its clients by offering a standard JDBC interface that allows to post standard SQL queries against the integrated data. For this purpose it deals with different data representations (model, schema) among the sources, data inconsistencies, like overlapping data sets, and possibly limited query capabilities of the source systems. These aspects are flexibly configured for a given application by describing the underlying sources and their interdependencies.

To achieve interoperability with the sources, easily configurable light-weight wrappers called adapters are used. Each adapter covers a class of data sources, e.g. RDBMS, XML, HTML, Z39.50. For instance, we use the XML adapter to integrate the XML version of the DBLP bibliography [Ley01].

## 2.2 Harvesting Bibliographical Information

The main interoperability issue for the HERMES alerting system [FFS+01] depicted in Fig. 2 is to interface with various heterogeneous providers of bibliographical data by providing appropriate wrappers. We characterize providers along two dimensions: active vs. passive, and proprietary vs. standard. Active providers notify the alerting system on new available publications, typically by e-mail. Passive providers are monitored on a regular basis to detect new publications, for instance by using HyperView wrappers for their Web sites [FS00]. A promising standard for passive providers is the OAI metadata harvesting protocol. HERMES includes a generic wrapper for arbitrary OAI compliant information providers. Another HERMES wrapper is responsible for harvesting the bibliographic databases offered by the DEMETRIOS federation service, thus covering all bibliographic data sources integrated by this service. An example for a mixed type is our project partner Springer Verlag, who notifies via e-mail about new bibliographies becoming available at their FTP server.

Although most information providers use proprietary formats, some adhere to standards. For instance, the Majour SGML format [Eur91] is a quasi-standard used by publishers such as Springer to describe articles in scholarly journals. Internally, we use a slightly modified XML version of Majour to describe articles. HERMES uses various format conversion techniques,

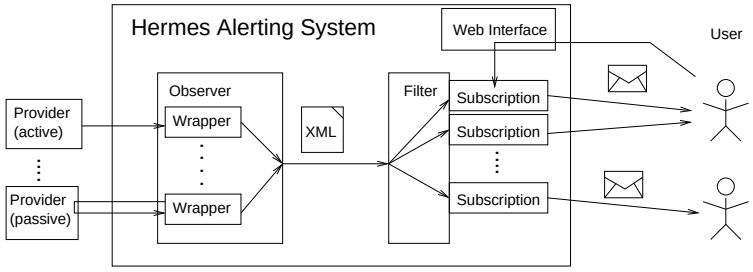


Figure 2: HERMES architecture

ranging from ad-hoc implementations to rule-based methods such as XSLT and HyperView.

## 2.3 Brokering Digital Library Services

The GIBRALTAR project provides several system-support-mechanisms that dynamically broker electronic publishing services, as well as an advanced user interface in the form of a web portal(figure 3).

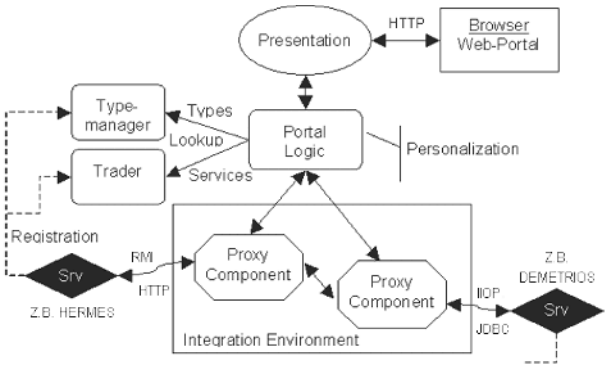


Figure 3: GIBRALTAR architecture

The fundamental infrastructure consists of a type manager and a trader component. Together they implement the system-level functions needed for service mediation. The type manager incorporates a sophisticated component-based type system [GZMW01] and allows for the handling of publication service specifications. The trader dynamically maintains a repository of current offers and matches requests for services in an optimizing fashion.

Services of remote providers are integrated through proxies, which are based on an extensible set of adapters for different communication protocols. The GIBRALTAR portal application uses the system-level components to provide orientation, mediation, combination and utilization to end users in a personalizable fashion.

# 3 Conclusion

We have presented the Digital Library services being developed in the *Global Info* program *Infrastructures for Digital Libraries*. In particular, we have focused on the interoperability

issues both within and among these projects.

Currently we are working on making HERMES and DEMETRIOS available through the GIBRALTAR portal. This includes specification of service descriptions and their implementation in form of APIs for a middleware such as CORBA.

Further open issues include a unified mechanism that transparently manages user accounts and preferences for multiple services, support for adding new services, and user assistance in combining services to a new customized services.

# References


[EHS+00] M. Endig, M. Höding, G. Saake, K. Sattler, and E. Schallehn. Federation Services for Heterogeneous Digital Libraries Accessing Cooperative and Non-cooperative Sources. In *Proceedings of Kyoto International Conference on Digital Libraries: Research and Practice*. IEEE Computer Society Press, 2000.

[Eur91] European Workgroup on SGML. *MAJOUR Header Manual*, 1.0 edition, 1991. Available at http://link.springer.de/author/sgml/help-sgml.html.

[FFS+01] D. Faensen, L. Faulstich, H. Schweppe, A. Hinze, and A. Steidinger. Hermes – a notification service for digital libraries. In *ACM/IEEE Joint Conference on Digital Libaries*, Roanoke, Virginia, USA, June 24-28 2001.

[FS00] Lukas C. Faulstich and Myra Spiliopoulou. Building HyperView wrappers for publisher web-sites. *International Journal on Digital Libraries*, 3(1):3–18, 2000.

[Glo01] Global Info Initiative. Globale Elektronische und Multimediale Informationssysteme für Naturwissenschaft und Technik - Förderprogramm des BMBF. http://www.global-info.org, 2001.

[GZMW01] F. Griffel, C. Zirpins, and S. Müller-Wilken. Generative softwarekonstruktion auf basis typisierter komponenten. In U. Killat and W. Lamersdorf, editors, *Proceedings 12. GI-Fachkonferenz Kommunikation in Verteilten Systemen (KiVS)*, Informatik aktuell, pages 325–338, Berlin Heidelberg New York, Feb 2001. Springer-Verlag.

[Ley01] Michael Ley. DBLP – Digital Bibliography & Library Project. http://dblp.uni-trier.de, 2001.

[SES00] E. Schallehn, M. Endig, and K. Sattler. Citation Linking in Federated Digital Libraries. In M. Roantree, W. Hasselbring, and S. Conrad, editors, *Proc. 3nd Int. Workshop on Engineering Federated Information Systems, EFIS'00, Dublin, Ireland, June*, pages 53–60, Berlin, 2000. Akadem. Verlagsgesellschaft.

[ZWBW01] C. Zirpins, H. Weinreich, A. Bartelt, and W. Lamersdorf. Advanced concepts for next generation portals. In *Proc. First International Workshop on Web Based Collaboration WBC'01*. IEEE, 2001.