Hardware and Software

The DiVA system is built on two servers: one database server and one web server. The web server calls the database server via Java servlets and builds XML pages of the incoming data. These XML pages are transformed with XSL style sheets into HTML and PDF pages.
XML is also used for the submission and updates to the database as a part of application.

The Database Server

The database is an Oracle 8i database installed on MS Windows 2000 Advanced Server operating system. This server is also responsible for the backup of the database and the web server.

Hardware

The server machine contains the following components:

- Intel dual processor motherboard with two Pentium III 900 MHz
- 2 GB RAM
- 2 hot-swap power supplies
- RAID controller with one RAID5 array and one RAID1 array
- DLT I tape streamer from Tandberg (see http://www.tandberg.com/)

Software

On the database server the following software is installed:

- Windows 2000 Advanced Server
  http://www.microsoft.com/windows2000/advancedserver/default.asp
- Oracle 8i RDBMS (relational database management system)
  http://technet.oracle.com/products/oracle8i/content.html
- Veritas Backup Exec 8.5
  http://www.veritas.com/products/category/ProductDetail.jhtml?productId=bent2000&_requestid=51298

Security

A firewall is installed on this server to block all network traffic that does not originate from the web server or the internal network. All files are backed up nightly.

The Web Server

The web server relies entirely on open source software. All web pages are built dynamically.

Hardware

The web server contains the following components:

- Dual processor motherboard with two Pentium III 1.3 GHz
- 2 GB RAM
- 2 hot-swap power supplies
- RAID controller with one RAID5 array

Software

On the web server the following software is installed:

- Red Hat Linux 7.2 (see http://www.redhat.com)
- Apache 1.3.x web server (see http://httpd.apache.org/)
- Tomcat 4.1.x servlets container (see http://jakarta.apache.org/tomcat)
- Java 2 Platform Standard Edition J2SE 1.4.x (see http://java.sun.com/j2se/)

Security

A firewall, which blocks all network traffic to the web server, is installed. Only ports 80 (HTTP) and 443 (HTTPS) are open. Another program searches the system and web server log files for known attacks. It controls also if any file on the system has been changed and creates a report over this changes. For more information see http://www.redhat.com.
A system to generate checksums for the full text files is under development. All files on the web server are backed up nightly and are also stored at the national library.

The Database

Currently we use a database to save bibliographical, administrative and archival metadata and abstracts.

These were the decisive factors when the RDBMS was chosen:

1. independence of operating system
2. support for Unicode
3. support for XML
4. support for full text searching and ability to search for diacritic characters with their base characters (for example: search for ç, ć, ĉ, č with a c)
5. support for sorting and dates depending on the [“you might want to support for sorting of dates with language-dependent formats” and then give an example]
6. fast and easy extraction of data

7. support of indexing external files
8. well supported and well known database

The Database Management System

The Oracle 8i relational database management system (RDBMS) was chosen at the beginning of the project (the current version is 8.1.7.3). A migration to Oracle 9i (9.2.0.1) is planned for the future. More information about the RDBMS can be found here:

- Oracle 8i
- Oracle 9i

Unicode and National Language Support

The RDBMS supports the storing of all information in Unicode (http://www.unicode.org). It is important for storing texts in different languages and with different types of characters. The National Language Support (NLS) makes it possible to sort according to different languages rules and to create dates in a format appropriate to a given languages. This makes it easier to create user interfaces for different languages.

Oracle Text

The RDBMS has its own index and search engine. It is possible to create a full text index over the entire database and XML files, which are stored or linked to the database. Searching for words that contain diacritics is then possible with their base characters (å is searchable with a). This feature is necessary for users who are unable to create the diacritics with their keyboard.
Another useful feature is the ability to create thesauri for indexing. The documents could be indexed according to a known vocabulary; then the search results would be more specific than a pure free text search.
More information can be found on http://otn.oracle.com/products/text/content.html.

XML Development Kit

The XML development kit (XDK) is available for different platforms (http://technet.oracle.com/tech/xml/xdkhome.html). In the DiVA project the Java servlets implementation is used. The use of the XDK makes it possible to handle the databases input and output via XML and XSL. This made it possible to develop the system in a way that will facilitate migration to a XML based database.

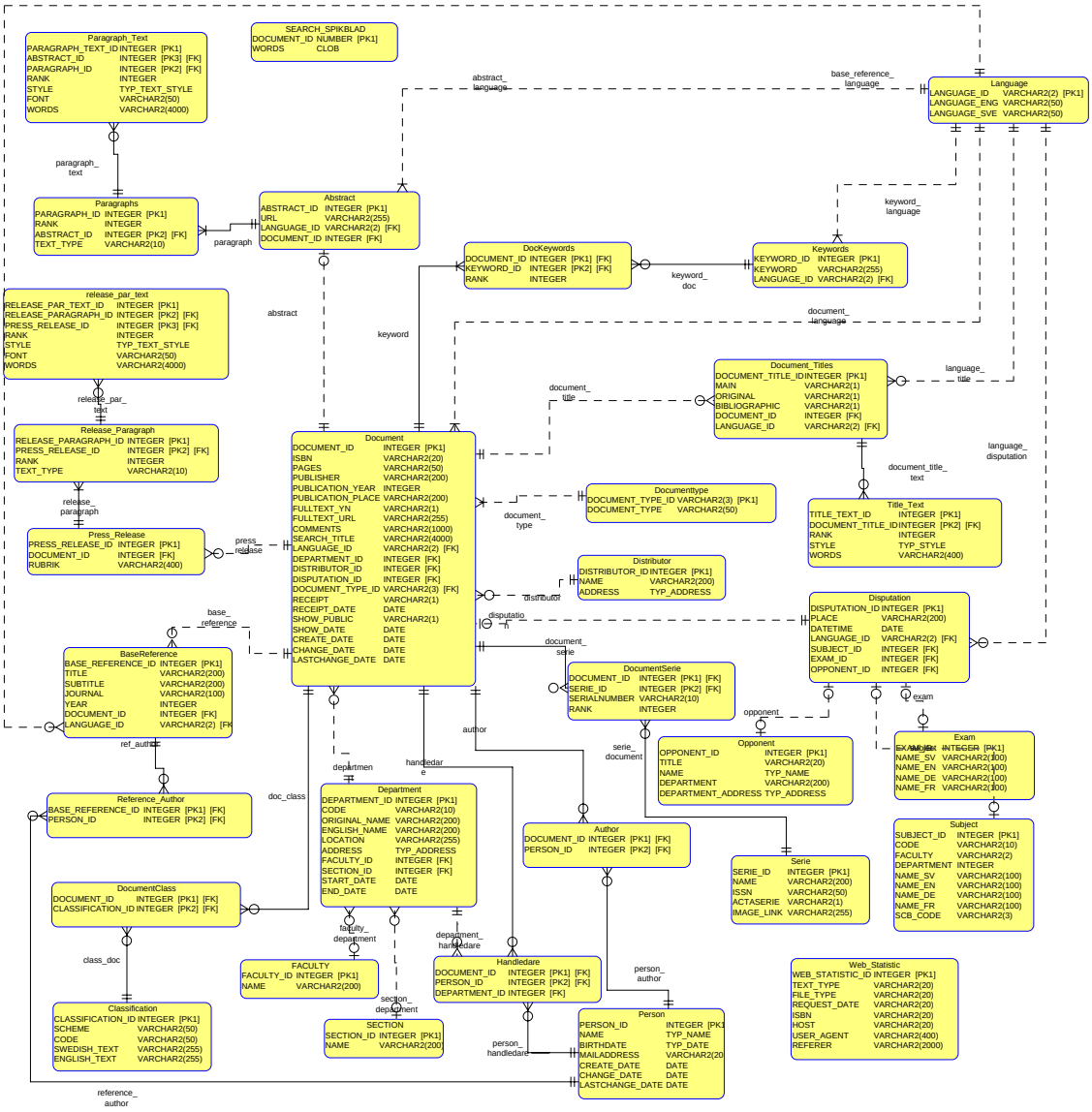
To store a document in the database, it is only necessary to create a XML-file and to send it to the database via the XDK. In the database the XML-file is converted to a set of rows, which contain the data. This 'rowset" is stored via standard SQL in the database tables.

To get the data of document, it is only necessary to create an SQL query. The XDK converts the result set to well formed XML, with the column names of the result set as the XML-tags. This XML data can then be used to create HTML or PDF pages for the user interface.

The Database Schema

The first version of the database was created more than a year ago and contains all the basic metadata, some administrative metadata, and abstracts.

(See the picture).



In our current system, the database must be redesigned whenever the XML schema is changed. The migration towards a database, which stores the static details of an item as pure XML and the authorized data and changeable data in database tables, is planned. The database will also be used to index and search the XML data.