article

The Online London Dysmorphology Database

Simon S. Guest, PhD[1], Carl D. Evans, BSc, PhD[1], and Robin M. Winter, BSc, FRCP[2]

A prototype of the Online London Dysmorphology Database is now available for testing purposes. In this paper we describe the new online version of the London Dysmorphology Database (LDDB), which is accessible via the World-Wide Web. The system has been developed primarily to address the information lag associated with the stand-alone version of LDDB (currently 3–15 months) and therefore provides access to the latest data from the master database that resides on a web server. An additional benefit is that, because the Online LDDB is a web-based resource, it will be available to users of both Macintosh and Unix computers as well as PC users.

Key words: *Dysmorphology, malformation, syndrome, database, Internet*

In recent years the ability to use computers to access and search biomedical information has become indispensable to dysmorphologists. Access to such information has generally been provided by "stand-alone" systems, such as the London Dysmorphology Database (LDDB) and POSSUM (Pictures of Standard Syndromes and Undiagnosed Malformations), which provide efficient retrieval and fluid interfaces for accessing data and presenting images, journals references, and abstracts linked to dysmorphic syndromes.[1,2] Although current versions of these systems provide sophisticated, yet friendly search interfaces and store very comprehensive data sets, they do still have some limitations. A key issue, which has provided a major focus for the work reported here, is the problem of "information lag," i.e., the time elapsed between entry of the data into the "master" database and the distribution of those data to registered users of the system. Information lag can often range from several months to more than a year and depends on the frequency with which new versions of the database are made available, which in turn can implicate publishing and distribution costs (currently 3 months to publication and each edition lasts a year). This problem is accentuated by the fact that in addition, new syndromes are being described continuously and the genes for existing syndromes are rapidly being identified and mapped.

The Internet has, in recent years, offered a way around the problem of information lag, and accordingly we have witnessed the emergence of numerous online resources that relate to the biomedical sciences, including genetics.[3] For example, the Human Genome Mapping Project World-Wide Web server in the United Kingdom now provides many links to genome resources, including systems such as Online Mendelian Inheritance in Man (OMIM[4]), the Dysmorphic Human and Mouse Homology Database (DHMHD[5]), and the Mouse Genome Database (MGD[6]). In addition to providing a means to solve the issue of information lag, such systems also implicitly address further problems with stand-alone databases, i.e., their general availability and their "platform-dependence." A web solution will implicitly offer access to the database from any location with an Internet connection and therefore will offer greater convenience to the clinician. Furthermore, stand-alone medical databases are often designed for, and therefore restricted to, one specific type of computer and its associated operating system, e.g., LDDB is currently only available for personal computers that run Windows. Thus, the provision of data via the Internet is an immediate solution to this problem, because Internet based systems can be accessed by virtually all types of computer and operating systems.

Despite these obvious attractions, due to the infancy of the World-Wide Web, the development of Internet-based resources has been restricted in a number of ways. Although an underlying distributed network infrastructure has been in place for a number of years, it is only fairly recently that the software required to develop sophisticated graphical user interfaces (GUIs) and complex querying of remote databases has been available. Before this, systems were restricted in terms of both the query interface and the searching mechanism. Thus, until very recently it has not been possible to develop an online edition of LDDB that would provide anything like the GUI and querying facilities available with the PC version.

This change has come about because of the emergence of programming languages such as Java (Sun Microsystems Inc., Palo Alto, CA). Using Java, sophisticated graphical user interfaces usually associated with professional stand-alone systems are now potentially available to users of Internet-based databases. This report describes the development of such a system, the new online version of LDDB. This new system offers clear advantages over the original stand-alone version. For example, the system has been developed primarily to address the information lag associated with LDDB (currently 3–15 months) and therefore provides access to the latest data from

From the [1]School of Computing Science, Middlesex University; and [2]Mothercare Unit of Clinical Genetics and Fetal Medicine, The Institute of Child Health, London, United Kingdom.

[a]The commercial version of Online LDDB will be hosted by Oxford University Press.

[b]The authors will be pleased to hear from any readers who are interested in joining a trial user group for testing the system.

Address correspondence to: Dr. Carl Evans, School of Computing Science, Middlesex University, Bounds Green Road, London N11 2NQ, United Kingdom.

Received: April 1, 1999

Accepted: June 2, 1999

the master database (a web server). An additional benefit is that the Internet version of LDDB implicitly becomes available for users of Macintosh and Unix computers.

A prototype of Online LDDB has been developed and is available for testing purposes through the Clinical Genetics World-Wide Web pages at the Institute of Child Health, London.[7] This prototype has password protection, and readers are invited to register as part of a trial user group and offer feedback on the usability of the system.

BACKGROUND

Dysmorphology is a field of medicine that has, as one of its concerns, the diagnosis of children born with multiple malformations or variations in quantitative characteristics (e.g., head size) that are outside the normal range. A pattern of signs recognized as occurring together and thought to be pathogenetically related is collectively called a dysmorphic syndrome, or malformation syndrome. Approximately 8 in 1000 children are born with multiple malformations, and in about one-half of these infants the condition is likely to have a genetic component. Most genetic conditions leading to abnormal physical development have not been mapped and the genes responsible have not been identified. All of these genes would be expected to play an important role in normal development. Thus the elucidation of the genetic basis of malformation syndromes will lead to greater understanding of normal embryological development, as well as an explanation for congenital abnormalities and a possible means for their prevention.

The London Dysmorphology Database (LDDB) is a standalone database of more than 3500 nonchromosomal malformation syndromes. The full database is PC based and is published by Oxford University Press (OUP). It contains comprehensive data for each syndrome, including literature references, a detailed abstract, and clinical photographs stored on an associated CD-ROM. The most recent version of LDDB allows syndromes to be linked to the OMIM World-Wide Web database via the OMIM number.

ONLINE SYSTEM DESCRIPTION

The Online LDDB is a "client-server" system. The LDDB web site (the "server") is comprised of the LDDB web pages and the LDDB database. The Online Database contains the

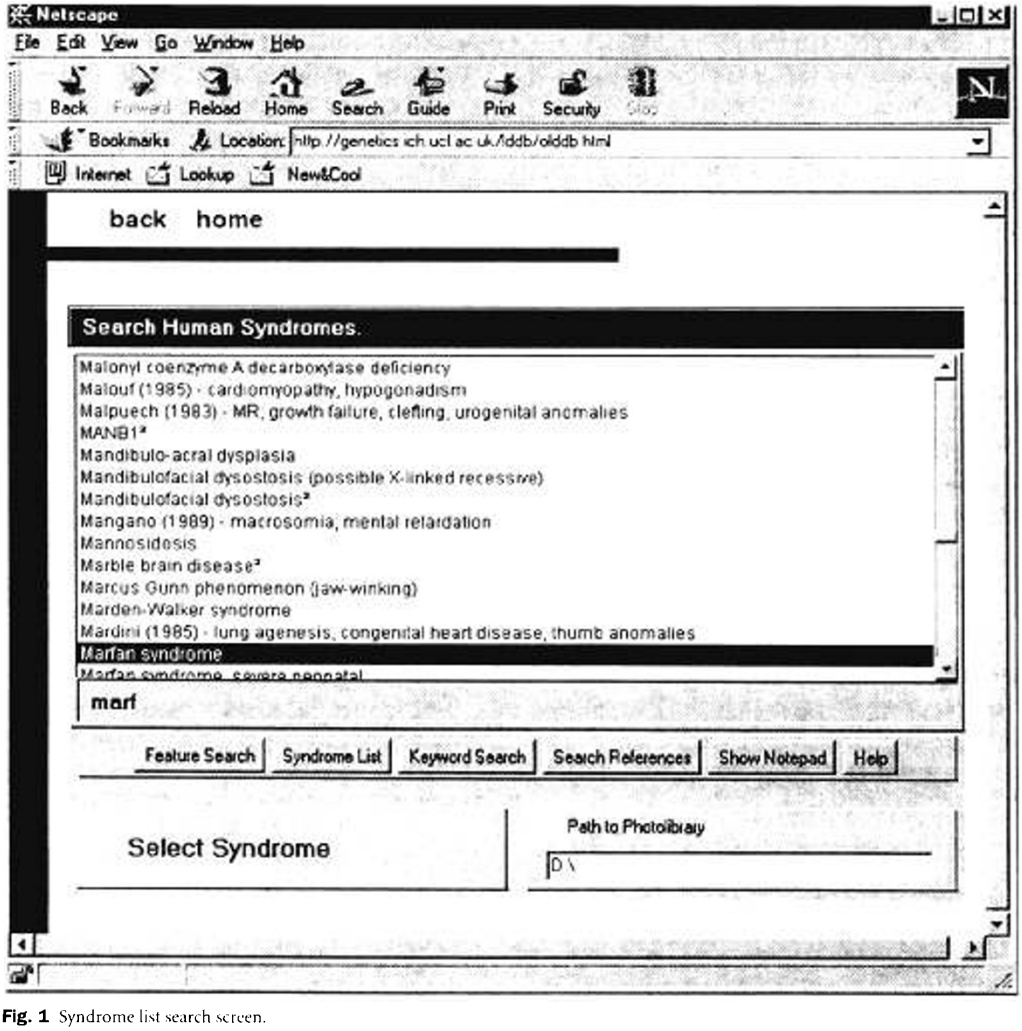

Fig. 1 Syndrome list search screen.

latest syndrome data from the master database that is not included in the CD–ROM version. Access to Online LDDB is secured through password protection. On the user's computer (the "client"), a Java program provides a graphical user interface to the LDDB database. This operates within a suitable Internet browser such as Netscape or Internet Explorer, or any "Java-enabled" web browser. The LDDB Photo Library is not distributed over the Internet at this time. If a user wishes to view reference images, as with the stand-alone LDDB system, s/he is required to have the photo library CD-ROM available at the client computer.

The Online LDDB allows the user to search for information on malformation syndromes and references relating to these syndromes. These searches are similar to those used in the stand-alone CD-ROM version of LDDB. There are three ways to search for syndromes: by a master list of syndromes, by a feature search, and by a keyword search. The results of these searches are returned as a list of selected syndromes. By selecting one of the syndromes presented in the list, all of the details about the syndrome that are stored in the database can be retrieved.

The syndrome list is alphabetically sorted (see Figure 1). The user may scroll down the list to select a specific syndrome, or, by typing the first few letters of a syndrome's name in the relevant text-field, the first syndrome in the list that matches the typed letters will be highlighted. By selecting a specific syndrome from the list, the user is presented with a screen that links to all the details of that syndrome, such as the syndrome abstract, it's list of clinical features, inheritance pattern, and references (see Figure 2). If the LDDB Photo Library is available locally on the client computer, the user can also elect to display the associated images.

The feature search screen (see Figure 3) allows the user to specify a number of clinical features from the master list of features that are presented. These are arranged in a standard "click-and-expand" format that is the same as that used in the stand-alone system, or Microsoft Windows 95/NT. The feature list is categorized so that the highest level lists general parts of the body, e.g., nose, that can be expanded to more specific abnormalities associated with that region, e.g., nasal bridge abnormalities. In all, three levels of feature categories are available, with the lowest level being the most specific, e.g., wide

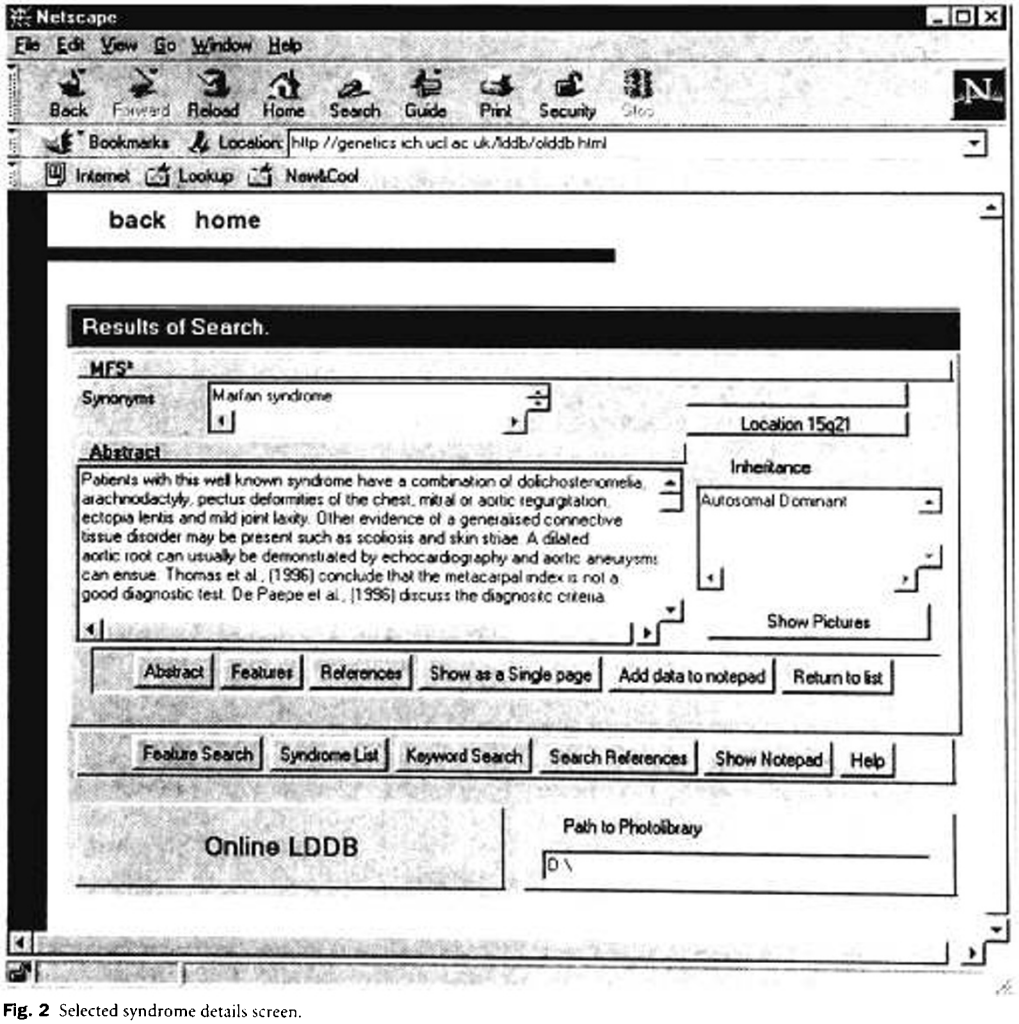

Fig. 2 Selected syndrome details screen.

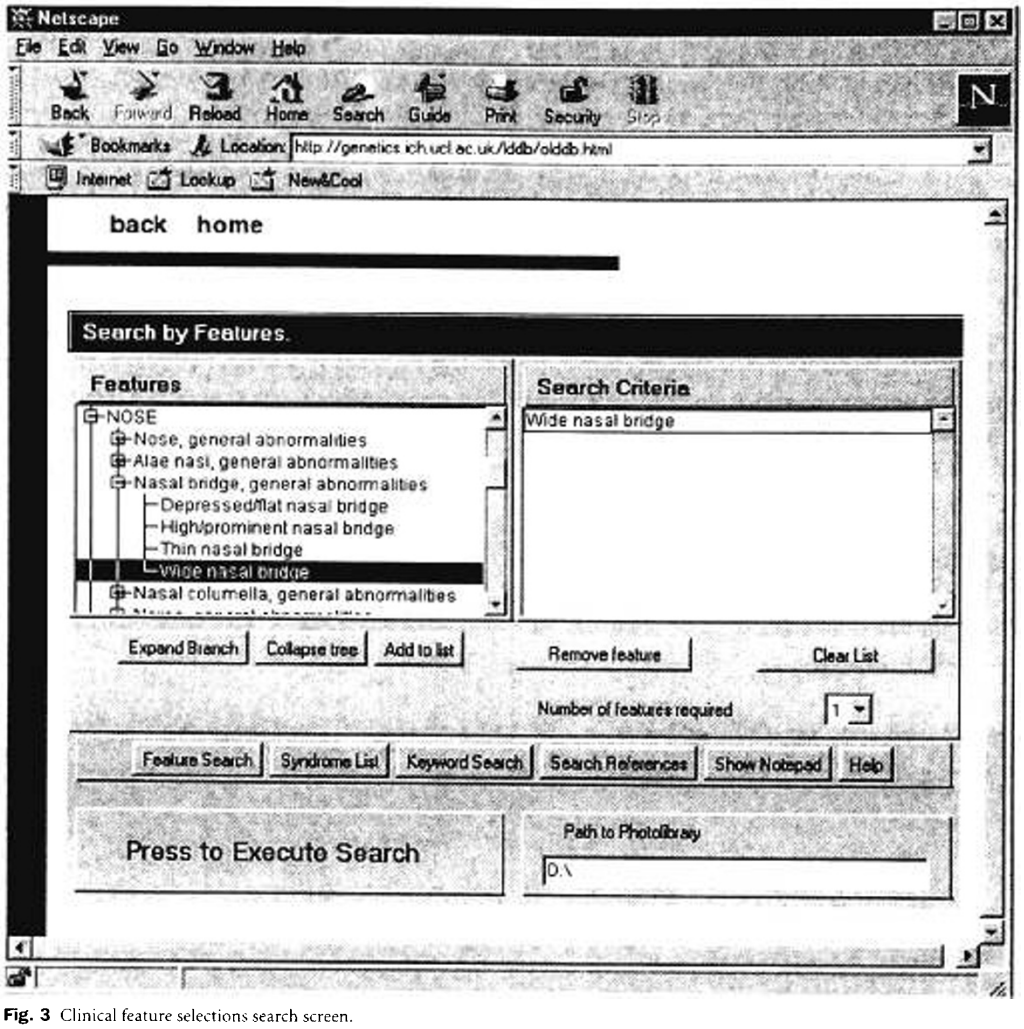

Fig. 3 Clinical feature selections search screen.

nasal bridge. Once a list of selected features is chosen, a search can be performed that retrieves all syndromes that exhibit a specified number of those selected features.

The keyword search facility, illustrated in Figure 4, allows the retrieval of syndromes that match a variety of criteria: chromosomal location, OMIM number, inheritance pattern, keywords in the syndrome name, syndrome abstract, and keywords in the title of references related to the requested syndromes. If more than one criterion is specified, the syndromes that satisfy all criteria are retrieved. Searching for references is performed by specifying keywords that are to be found in the title or abstract, or by specifying an author of the article, the journal of publication, or years in which the articles were published. These criteria can be combined to provide sophisticated access to the references stored in the LDDB database. Once a list of matching references has been retrieved, the details of each reference can be viewed.

As well as searching facilities, the Online LDDB has a number of features to further assist users in achieving maximum use of the system. Due to the security restrictions placed on Java, access to the client computer's hard drive and printers is not permitted. However, the ability to print or save results is invaluable. To provide such a facility, the Online LDDB provides a simple "notepad" into which the results of searches can be copied. This notepad can then be edited in a manner similar to using a simple word processor. To save or print the contents of the notepad, the user is simply required to "copy and paste" the contents of the notepad into a word processor application available locally (i.e., on the client computer) and then to use the local application to print or save the information. Other features include links to other Internet-based resources from the Online LDDB; OMIM, and Entrez[8] databases, based on OMIM numbers and chromosomal locations, respectively, and a simple "context sensitive" help facility that provides information and hints about how to best utilize the interface and search facilities.

DISCUSSION

There are some differences between the Online LDDB and the CD-ROM version. First, there are some minor visual differences in presentation. Second, the Online LDDB does not use standard Windows pull-down menus (at this time), whereas

the original stand-alone version does. These differences are primarily the result of the technical limitations of Java versions that are widely supported at this time. A further issue is that the time required for some of the searches can be considerably longer than that with the stand-alone edition. This aspect is particularly noticeable with searches for references. With the stand-alone database, a reference search takes a few seconds; whereas with the Online LDDB reference, searches can take up to a minute or slightly longer. Although it is hoped that future versions will improve on this performance aspect, at the current time it is a necessary trade off due to the underlying technical differences between the two systems and the fact that the online system queries a remote database rather than one stored locally.

Images could be provided online. However, with the current system, the images are only available to those who have the CD-ROM local to the client and held in a PC/Windows machine. This restriction is partly due to the nature of these case-sensitive data and issues concerning privacy, rights of copyright, and distribution. The system is the first of this type with which Oxford University Press has been involved. Accordingly, Online LDDB serves as a prototype for the OUP Medical Database series, in which copyright and charging mechanisms will be evaluated. It is envisioned that charges for the commercial version[a] will be based on a subscription for registered users. The current prototype is available free for trial purposes[b] and details of user registration are available through the web address listed.[7]

FUTURE WORK

The developers of Online LDDB have been working with a view to future developments within the growing field of Internet-based biomedical research, sometimes referred to as "bioinformatics." In particular, there has been some recent research in developing "distributed" resources, i.e., databases located at different remote locations that can be queried in tandem by one user at the client computer.[10] Use of this technology will enable a number of related genetics resources to be developed with consistent GUI design and to be integrated and made available over the Internet so that a user can query separate remote data sets and cross-reference the results. These resources will include LDDB along with a new edition of the

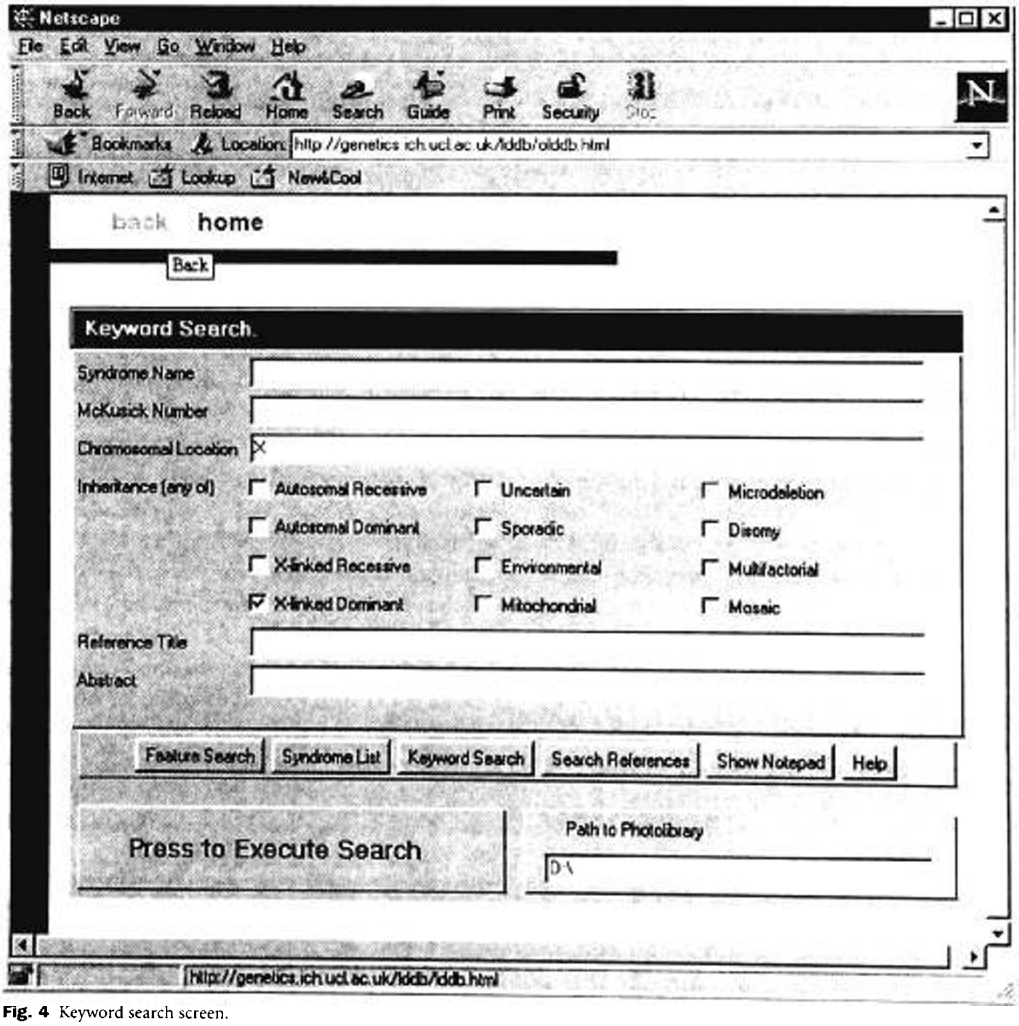

Fig. 4 Keyword search screen.

Dysmorphic Human and Mouse Homology Database,[9] which will be available in 1999, and a revised version of the Dysmorphology Discussion Board.[11] Partly with this in mind, considerable effort has been devoted to ensure that the Online LDDB is independent of the underlying database format. This is important. It means that the online system can utilize the standard database files that are continually compiled at the Institute of Child Health, London, ensuring that the up-to-date data can be distributed easily and regularly, and ensuring that the Internet application will be "future-proof" to changes.

The design philosophy of Online LDDB has been influenced, to a great extent, by developments in Internet technology and the Java programming language. The prototype version has been developed using the Java 1.0 specification, which is well-established and works consistently with current web browsers. Very recent releases of some Internet browsers support the later Java 1.1 specification, and future releases will support the Java 1.2 specifications. These versions of Java provide a means to support, and therefore take advantage of new GUI features, such as "drag and drop" interaction and the facility to print and save on the client machine.

CONCLUSIONS

The Online LDDB provides a graphical user interface with appearance and functionality similar to the current stand-alone version of LDDB, but with a dramatic reduction in information lag. Despite some minor performance issues, the usability of the online system is comparable to the stand-alone version. In addition, the platform independence of the system increases the range of machines that can access the data stored in the LDDB, i.e., any machine that can run Netscape, Internet Explorer, or any Java-enabled Internet browser. In practice, this includes virtually every configuration of desktop computer or workstation.

Acknowledgments

The Online LDDB project is a collaboration between the database compilers at the Institute of Child Health, computer scientists at Middlesex University, and the CD-ROM publishers, Oxford University Press. The project is funded by the Engineering and Physical Sciences Research Council (EPSRC), Grant Number GR/L88863.

References

1. Bankier A, Keith CG. POSSUM: The microcomputer laser videodisk syndrome information system. *Ophtalmic Paediatr Genet* 1989;10:51–52.
2. Winter RM, Baraitser M. The London Dysmorphology Database: A computerised database for the diagnosis of rare dysmorphic syndromes, 5th Edition. Oxford: Oxford University Press, 1998.
3. van Steensel MAM, Winter RM. Internet databases for clinical geneticists-an overview. *Clin Genet* 1998;53:323–330.
4. Online Mendelian Inheritance in Man. http://www.ncbi.nlm.nih.gov/Omim/
5. Dysmorphic Human and Mouse Homology Database. http://www.hgmp.mrc.ac.uk/DHMHD/dysmorph.html
6. Mouse Genome Database. http://www.informtics.jax.org
7. Online London Dysmorphology Database. http://genetics.ich.ucl.ac.uk/
8. Entrez. http://www.ncbi.nlm.nih.gov/Entrez
9. Evans CD, Searle AG, Schinzel AA, Winter RM. The dysmorphic human-mouse homology database (DHMHD): An interactive World-Wide Web resource for gene mapping. *J Med Genet* 1996;33:289–294.
10. Objects in Bioinformatics 1998. August, Hinxton, Cambridge, 1998. http://www.ebi.ac.uk/oib98/
11. Winter RM. The Dysmorphology Discussion Board. *Clin Dysmorphol* 1997;6(1):74. DDB http://genetics.ich.ucl.ac.uk/DDB/ddb.html