

# The Virtues of the Virtual World

Genetics professionals rely on the Internet as a source of information at the point of care. Accessing OMIM, PubMed, GeneTests (formerly Helix), and an array of online patient resources is part of the daily routine in the evaluation and genetic counseling of patients and their families. Guest et al.<sup>1</sup> describe the process of adding another Internet resource to this growing list and allude to another view of the Internet as a "virtual database" of networked diagnostic and management tools. Medical Genetics, the specialty that encompasses thousands of human diseases, with rapidly evolving information about its tens of thousands of genes, lends itself more than any other field of medicine to this networked world, which integrates important but fragmented and ever-changing information. In the developing world of virtual databases, a group with a specific interest designs and maintains its own database to store information that can then be networked with other special interest databases in one of two ways.

In the simplest model, one database contains pointers (links) to the appropriate datafield in another database [e.g., an OMIM search result returns both OMIM content and appropriate pointers (links) to genomic databases; a GeneTests (formerly Helix) search result includes links to OMIM and GeneClinics; a search result of the online London Dysmorphology Database (LDDDB) includes links to OMIM]. These links are "man made"; they are created and maintained in the database by a person who makes the decision about the most appropriate link between information resources. This model requires an internally consistent representation of information and a human to make translations between databases.

A more complex approach permitted by newer databases is to pose a query at the "front end" of one database (the interface that the user sees with the database, which is usually a "search screen") and to provide a search result that is based on queries of multiple databases, which have occurred without any additional instructions from the user. A simple example is provided by the web search engine Metacrawler ([www.metacrawler.com](http://www.metacrawler.com)), which takes a user query and executes it against a series of other web search tools (with some query reformulation or translation as needed). In order to provide maximal benefit in the medical domain, these "virtual databases" require interdatabase consistency of data representation; i.e., a standardized vocabulary and/or nomenclature system as well as a common data model (i.e., a common understanding of the underlying medicine or biology).

For a stand-alone database to be poised to become part of a virtual database, it must first be converted into a network accessible form. Tarczy-Hornoch et al.<sup>2</sup> have described the

process for Helix (now GeneTests); Guest et al.<sup>1</sup> describe the process for LDDDB.

The next step is more complex. The model of virtual or networked databases is being used by basic science researchers in the field of human genetics. For example, a scientist can query multiple different genomic databases via the Entrez system ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), which permits a search for a gene product to be executed against protein, sequence, and locus databases. A comparable model does not yet exist for clinicians, but the tools and interest do. The tools required are a set of databases [rather than text files such as a word processed document or a plain web page (i.e., an HTML file)], standardized nomenclature, and/or vocabulary, and data models that are as flexible as possible in order to accommodate conceptual changes that will alter how data elements need to relate to each other. In this virtual world, any interested professional or lay person could query multiple databases through a single search mechanism using clinical findings; the search would be executed against a diagnostic database such as LDDDB, which in turn would launch searches against its networked databases. The result would be a list of diagnostic possibilities from the LDDDB and for each diagnostic possibility, the list of links to the relevant clinical databases [e.g., OMIM ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), PubMed ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), GeneTests ([www.genetests.org](http://www.genetests.org)), GeneClinics ([www.geneclinics.org](http://www.geneclinics.org))] which has been generated on the fly rather than from "man made" static links as described above.

The day is coming in which the clinician will need to be as comfortable in the use of such databases as in the use of print resources. A commitment of the editorial board of *Genetics in Medicine* is to provide information on clinically relevant databases for the benefit of the readers of *Genetics in Medicine*.

**Roberta A. Pagon, MD**

*Division of Medical Genetics  
Children's Hospital/Medical Center  
Seattle, Washington*

**Peter Tarczy-Hornoch, MD**

*Division of Neonatology  
Department of Pediatrics  
University of Washington  
Seattle, Washington*

## References

1. Guest SS, Evans CD, Winter RM. The Online London Dysmorphology Database. *Genetics in Medicine* 1999;1:207-213.
2. Tarczy-Hornoch et al. Creation and maintenance of Helix, a Web-based database of medical genetics laboratories, to serve the needs of the genetics community. *Am Med Inform Assoc Fall Symposium Suppl* 341-345, 1998.