

The Database of Online Health Statistics: a new search tool²

Trish Chatterley and Liz Dennett

Introduction

Health statistics are notoriously difficult to locate, even for the most competent health librarians. However, as statistics often play a vital role in healthcare policy, planning, services, and evaluation, finding them is a necessity. This article reviews the difficulties in finding health statistics and describes the creation of the Database of Online Health Statistics at the Institute of Health Economics in Edmonton, Alberta, Canada. This resource was designed to simplify the search for statistics by librarians and the general public.

For decades, the hunt for health statistics has been described as a frustrating and (or) feared task by the general public and information specialists alike [1–5]. The challenges involved are numerous. Government agencies, international organizations, associations, research institutes, corporations, universities, and interest groups, among others, all generate health statistics. However, access provided by these groups differs. Governments often make the statistics they gather freely available on the Internet and researchers publish their figures within journal articles, whereas corporations may not provide public access at all. This plethora of creators and publication methods, coupled with the lack of a central location to house the statistics, often makes finding a particular figure difficult. The format of the publications may create impediments; a statistic may be represented by one or more sentences buried within an article or report. The searcher must then know what terminology to use to locate the numerical data within the document. Alternatively, the statistics could be featured in image formats such as figures or tables, which often precludes searching within them [5]. New technologies, such as geographical information systems (GIS), are being employed to provide the public with interactive graphical displays of datasets generated in real-time that, while useful for manipulating variables to get at the

statistics one desires, prohibit searching via online search engines [1]. Also, drop-down menus are not crawled by search engines, and are often of broad categories rather than the specific variables of interest.

The various creators of statistics also use different terminology, definitions, collection and analytical methods, and date ranges for their data collection, making comparisons complicated [1]. This leads to fragmentation of data, and thus the need to look in multiple places to find answers to different elements of the same question. Many countries have national statistical agencies that serve as repositories for statistical publications. While many provide free access via websites to these statistics, their sites are often difficult to navigate. Coverage is also inconsistent as not all countries, states, or provinces are represented.

The information specialist has the additional challenge of trying to interpret the patron's request. Reference queries are not always expressed clearly at the best of times, so when the confusing world of statistics is thrown in, the patron's needs often become even less apparent. The National Library of Medicine's tutorial *Finding and Using Health Statistics* [6] and the book *Introduction to Reference Sources in the Health Sciences* [7] provide useful tips for the reference librarian in assessing needs and formulating a search strategy.

Finally, librarians or information specialists are rarely asked to find the same statistics more than once, so it is difficult to gain confidence, knowledge of specific resources, and subject mastery in this area.

Description

At the Institute of Health Economics (IHE), we faced all of the above issues in locating health statistics. The IHE publishes a number of quick-reference booklets that summarize interesting healthcare statistics for policy makers.

T. Chatterley.¹ Public Services Librarian, 2K3.28 John W. Scott Health Sciences Library, Walter C. Mackenzie Health Sciences Centre, University of Alberta, Edmonton, AB T6G 2R7, Canada.

L. Dennett. Institute of Health Economics Research Librarian, Institute of Health Economics, 1200, 10405 Jasper Ave, Edmonton, AB T5J 3N4, Canada.

¹Corresponding author (e-mail: patricia.chatterley@ualberta.ca).

²Based on presentations at: Northern Alberta Health Libraries Association TRENDS mini-conference, Edmonton, AB, 23 April 2010; Canadian Health Services and Policy Research Conference, Calgary, AB, 12 May 2009; Health Technology Assessment International Conference (poster presentation), Montreal, QC, 8 July 2008; Canadian Health Libraries Association Conference, Halifax, NS, 29 May 2008.

Each year, we spent time retracing our footsteps to find these statistics. We thought it would save time if we compiled all the sources we used into a database that we could refer to the following year. We quickly realized that external information specialists, researchers, and even the general public might find our database useful. We therefore decided to make the Database of Online Health Statistics freely available (<http://www.ihe.ca/publications/health-db/>) and to expand our efforts in locating statistics to capture a wide swath of the freely available online health statistics.

At the same time we decided to create this database, the IHE website was being redesigned by a professional web design company that agreed to include the infrastructure and design of the database for a relatively small additional cost.

To locate statistics beyond those known to us, we mined health statistics resource guides from health sciences libraries and government agencies. An Excel spreadsheet of guides and their URLs was maintained to keep track of visited sites. Once the resource guides were exhausted, we continued to find useful health statistics by searching the websites of statistical agencies, governments, health professional organizations, and interest groups. Currently we primarily add publications and resources that we discover during our regular duties or that are suggested by colleagues.

To supplement the URL of each statistical resource and to improve retrieval, we added descriptive metadata to each record created. Because statistics are almost always tied to a particular geographic region, we assigned a geographical location to each record. We also assigned at least one broad healthcare category (e.g., cancer, mental health, or chronic diseases). The full list of geographic locations and broad health topics are presented on the main page of the database to allow browsing by geography or topic. If one browses by geography, the results are sorted by health topic and if one browses by health topic, the results are sorted by geographic location.

Each record includes the title and creator of the resource, and to provide further entry points for searching, we also included a description field and tags. We felt that creating a controlled vocabulary would be too time consuming considering the limited time we were able to give this project, so we decided to use free text tagging instead. In the individual record display, the tags provide a link to a list of the other records that have the same tag. To ensure consistent tagging and eliminate multiple variations of similar tags, a matching system was built into the backend. As letters were entered in the tag field, a list of previously used tags popped up from which we could select tags for the new record.

The search interface on the main page of the Database of Online Health Statistics uses a Google search box to search all the records in the database. We are hoping to upgrade this search engine to one that allows for more control (e.g., Boolean searching), but this will depend on available funding.

The web design company created a backend to the database that allows us to add, edit, manage, and delete records efficiently, which helps to keep the costs of maintaining this database low. Each record is assigned an expiry date at creation and the backend allows us to review expired records quickly, update URLs if needed, and reinitialize the records.

The initial populating of the database took 8 to 10 days of librarian time. The budget for this project was limited, but we were given internal funding to hire a contract librarian (a recent library school graduate) for 100 hours to locate and add additional resources. A practicum student also spent about half of her 100 hour placement doing the same. An administrative assistant at the IHE helped with identifying and updating dead links, though we are hoping to further automate this process in the future as we do not believe manual checking is sustainable as the database continues to grow.

Outcomes and discussion

Paula Berinstein outlined a three-step strategy to finding statistics: (1) select the right place to look, (2) choose an effective strategy, and (3) evaluate the data [5]. The Database of Online Health Statistics provides assistance with the first two steps. The best places to look for health statistics have been identified and direct links to the resources provided. While this sounds similar to what one would find on a traditional library guide, the item records contain major categories, annotations, and tags that provide more information than a title alone. This enables the database to be searched for topics of interest, in addition to the more typical browsing by category or geographical region. However, Berinstein's third step, evaluating the data and statistics and determining relevance, is still the responsibility of the user.

Although we have not promoted the database extensively, beyond presentations and (or) posters at four conferences, it appears to meet a need in the web searching community. The database's records have been indexed by major search engines, which has increased the IHE's website usage statistics immensely. People are also beginning their search at the home page of the database, as that page is one of the most used pages on the IHE website. According to Google analytics, the database home page has been viewed approximately 3900 times in the past year (1 November 2009 – 31 October 2010). Total number of views for all pages in the database is approximately 27 900. It is important to keep in mind that someone might click through three or four pages on their way to finding a useful resource. While page views do not tell us if people found relevant information, it does indicate that people are finding their way to the database.

We have received a number of anecdotal reports from information specialists and researchers letting us know how useful they have found the database. We have also found the database beneficial in answering questions at the reference desk and helping University of Alberta students find health statistics resources in our roles as academic health sciences librarians.

There are a number of limitations to the database. It is obviously not comprehensive, nor can we reasonably expect it to be. There are many health statistics that will evade our notice no matter how intensively we search for them; too many statistics are hidden deep in reports. We have only tried to find reports and resources that are rich in statistics. It is too much work to catalogue sources that offer single or limited statistics. However, we hope to reach and maintain a level of comprehensiveness that offers

access to many of the key resources in the health field. Our geographic focus is undoubtedly English-speaking Canada, but we are making an effort to be aware of major resources in the rest of the English-speaking world. Unless our funding is substantially increased, we are unlikely to be able to provide access to materials in other languages. At present, the database does not index open access journal literature, although this is undoubtedly a rich source of free online health statistics.

The stability of URLs is a larger problem than we had originally anticipated. We mistakenly assumed that since the majority of websites that we were linking to were owned by large and well-funded organizations, the URLs would be fairly stable. Instead we have been surprised by the number of dead links or redirected links that have occurred since we released the database in January 2009. At the moment we rely heavily on our users to notify us when links are not available as we provide an e-mail link to do so.

Both the functionality and the content of this database will remain a work in progress.

We invite and encourage our fellow health sciences information colleagues to contribute links to our database in the hope that it can be a useful resource for all of us.

Acknowledgements

This work was made possible as part of a larger capacity building grant from Alberta Health and Wellness to the Institute of Health Economics. We wish to thank Andrea

Emberley and Ashley Eisen for their assistance with updating and maintaining database content.

References

1. Rankin JA, Burgess ML. Medical and health statistics. In: Huber JT, Boorkman JA, Blackwell J, editors. Introduction to reference sources in the health sciences. 5th ed. New York: Neal-Schuman Publishers, Inc.; 2008. pp. 249–302.
2. Bexon N. Brief communication. Finding health statistics: making a difficult task easier. *Health Inf Libr J*. 2001 Mar;18(1): 54–6.
3. Snow B. Finding medical and health care statistics online. *Online*. 1988 Jul;12(4):86–95.
4. DeGeorges KM. Internet working. Numbers, I need numbers! finding health statistics on the web. *AWHONN Lifelines*. 1999 Apr/May;3(2):49–50.
5. Berinstein P. Finding statistics online: how to locate the elusive numbers you need. Børner S, editor. Medford, New Jersey: Information Today; 1998.
6. National Library of Medicine. Finding and using health statistics [Internet]. Bethesda, Maryland: National Institutes of Health; 7 Oct 2008 [cited 10 Aug 2010]. Available from: <http://www.nlm.nih.gov/nichsr/usestats/index.htm>.
7. Huber JT, Boorkman JA, Blackwell J, editors. Introduction to reference sources in the health sciences. 5th ed. New York: Neal-Schuman Publishers, Inc.; 2008.