
Modeling public health interventions for improved access to the gray literature

By Anne M. Turner, MD, MPH, MLIS
amtturner@washington.edu
Clinical Assistant Professor

School of Public Health and Community Medicine
University of Washington
Box 357660
Seattle, Washington 98195

Elizabeth D. Liddy, PhD, MLS
liddy@mailbox.syr.edu
Professor

Jana Bradley, PhD, MLS, MA, FMLA
janabrad@email.arizona.edu
Professor

Joyce A. Wheatley, MS, MLS
jawheatl@syr.edu
Research Analyst

Center for Natural Language Processing
School of Information Studies
Syracuse University
Syracuse, New York 13244

Objective: Much of the useful information in public health (PH) is considered gray literature, literature that is not available through traditional, commercial pathways. The diversity and nontraditional format of this information makes it difficult to locate. The aim of this Robert Wood Johnson Foundation–funded project is to improve access to PH gray literature reports through established natural language processing (NLP) techniques. This paper summarizes the development of a model for representing gray literature documents concerning PH interventions.

Methods: The authors established a model-based approach for automatically analyzing and representing the PH gray literature through the evaluation of a corpus of PH gray literature from seven PH Websites. Input from fifteen PH professionals assisted in the development of the model and prioritization of elements for NLP extraction.

Results: Of 365 documents collected, 320 documents were used for analysis to develop a model of key text elements of gray literature documents relating to PH interventions. Survey input from a group of potential users directed the selection of key elements to include in the document summaries.

Conclusions: A model of key elements relating to PH interventions in the gray literature can be developed from the ground up through document analysis and input from members of the PH workforce. The model provides a framework for developing a method to identify and store key elements from documents (metadata) as document surrogates that can be used for indexing, abstracting, and determining the shape of the PH gray literature.

INTRODUCTION

Much useful information in public health is generated through governmental and nonprofit organizations in the form of program reports, meeting notes, data sets, policy briefs, and other formats. These publications typically do not find their way into established commercial outlets for publication and, as a consequence, are not consistently indexed in MEDLINE and other established tools for locating public health information. As a result, public health documents often fall into the category of "fugitive" or gray literature, literature that is difficult to locate because it is not available through traditional commercial pathways [1–4].

Gray literature is important to public health professionals for a variety of reasons.

- It enables them to learn from and build on the activities of others working in the field.
- It enables them to provide examples of successful assessments or interventions to their stakeholders and constituents.
- The ability to access, review, and organize gray literature materials provides support for policy and decision making.

From the information science perspective, the creation of a model for locating public health gray literature could lead to standards for reporting key elements of these documents and could facilitate users' retrieval and use of this information.

THE ACCESS PROBLEM

The proliferation of such non-indexed, potentially valuable information has stirred interest among those involved in the collection of, organization of, and access to this body of public health literature [1–3]. In 1997, at a forum sponsored by the National Library of Medicine, "Accessing Useful Information: Challenges in Health Policy and Public Health," experts in public health information called for a more systematic way of organizing the gray literature of public health [2].

Improving access to this literature faces a number of considerable problems. The sheer volume, diversity, and nontraditional formats make gray literature difficult to locate. Notable collections of gray literature exist, such as the New York Academy of Medicine's (NYAM's) Gray Literature Report, but they represent a small fraction of potentially useful gray literature. Although gray literature documents have become increasingly available on the Internet, collocation, cross-linking of materials across sites and regions (county-to-county, state-to-state, county-to-state), is negligible.

The two established models for systematic access to literature in many academic domains are limited solutions to the access problem for the gray literature in public health. The model of controlled vocabulary indexing by human experts, successful in the relatively circumscribed world of the literature published through commercial and vetted venues, has difficulty scaling up to accommodate the quantity and variety

of gray literature documents. Although search engine technology is useful in locating some relevant documents, it is restricted in its coverage by problems of collocation across keyword texts and the absence or choice of metadata that exist. Thus, the public health professional who is looking for information about a particular public health problem, population, geographic region, and so on faces a formidable challenge in finding pertinent gray literature in the public health domain. This paper reports on an exploratory project to develop an alternative response to the access problem in public health gray literature.

The natural language processing (NLP) techniques developed at the Center for Natural Language Processing at Syracuse University have been used to identify, mark, and extract key elements of text for summarizing and indexing a sample collection of digital documents submitted to the National Science Foundation Digital Library Project (NSFDL) in the education domain [5]. This project investigates the application of these tested NLP techniques to improve access to public health gray literature.

A brief conceptual description of the NLP approach is useful at this point. The application of NLP depends on human conceptual identification of the key elements in an event or situation that would constitute a useful summary, or surrogate, of a document. These key elements can be considered to be a model of the event or situation. Humans then write machine-processing rules for identifying these elements in free text and for extracting them as metadata into a database. The extracted elements can then be displayed as desired—as text, tables, or graphics—and searched. If the original documents exist on the Internet, links from the database to the original documents deliver the document to the user.

The first step, then, in exploring the usefulness of NLP techniques in addressing the problem of access to gray literature in public health is to identify the key conceptual elements of public health events or situations that would comprise a useful surrogate for public health documents. This paper describes the process of developing a model of such elements, based on two complementary methodologies:

- an analysis of the content and format of a collection of public health gray literature documents (literary warrant)
- the expert input of public health professionals (expert user warrant)

METHODS

The authors' methodology is best described in stages:

- document collection
- document analysis and preliminary model development
- user survey
- integrated model development

Table 1
Distribution of training set documents

Source	Uniform resource locator (URL)	#
New York Academy of Medicine (NYAM) Gray Literature Report		120
August 1999, volume 1(1)	www.nyam.org	39
November 2001, volume 3(4)	www.nyam.org	81
State health department Websites		100
Minnesota Department of Health	www.health.state.mn.us	45
Georgia Department of Public Health	www.ph.dhr.state.ga.us	27
North Carolina Center for Health Statistics	www.schs.state.nc.us	28
County health department Websites		100
Hennepin County Community Health Department	www.co.hennepin.mn.us	59
Kent County Department of Health	www.accesskent.com	21
Lake County Health Department	www.co.lake.il.us	20

Document collection

The first task was to develop a collection of gray literature documents in the public health domain. In the first phase of our project, reported in this paper, this collection formed the basis for a literary analysis to identify preliminary key elements of our model. In later phases, a large percent of the documents will be used as a training collection for writing the rules to extract key elements. A smaller percent, set aside at the start of the project, will constitute a test collection and will be used to test the resulting system.

The first conceptual problem the authors encountered in document selection was defining the scope of public health content to which we were seeking to provide improved access. For this exploratory project, we chose to focus on the public health intervention as a focal event or situation in public health and thus to collect documents relevant to public health interventions. This decision to focus on interventions was later confirmed as worthwhile by the user study.

The following broad definition of an intervention guided the selection of documents for this collection: "Any strategy, procedure, therapy, approach, method, or technique that changes, stops, deters, or interacts with a problem, disorder, disease, or disability of a patient, group, or community" [6]. Each document included in the collection had to relate to some aspect of a public health intervention. A document was considered gray literature if it was not available in a commercial format.

A representative sample of 365 gray literature documents relevant to public health interventions was collected from 7 Websites that offered a large number of public health publications. Approximately one-third of the documents were from the NYAM Gray Literature Report Website ($n = 135$), one-third from 3 state public health department Websites ($n = 115$), and one-third from 3 county public health department Websites ($n = 115$). Of the 365 documents collected, 320 were used for the training collection and 45 documents (15 from each set) were set aside to use in testing the system once it was developed. Table 1 provides a list of the organizations, uniform resource locators (URLs), and number of training documents from each site.

Document analysis and preliminary model development

The first part of a two-pronged approach to developing a model of key elements related to public health interventions was to apply standard content analysis techniques to identify commonly occurring text elements across documents and to develop a preliminary model of these elements.

User survey

The second part of our two-pronged approach to model development was to seek the opinion of expert users. The goal was to determine what key elements public health professionals perceived as important to include in a summary of a document. The Institutional Review Board (IRB) of Syracuse University granted approval to perform these studies.

Participants were provided with copies of four public health documents chosen randomly from the training set. Three different approaches were used to elicit important elements. One approach asked participants to rank a list of standard bibliographic elements, modeled after those commonly found in citations for documents published through commercial channels (title, publisher, subject heading, etc). A second approach requested participants underline elements in the text that they felt were important to help public health professionals assess the usefulness of a document [7]. The third approach asked participants to write an abstract of the length and content necessary to determine if a document would be useful in their work.

In addition, the survey sought to get an indication of the types of documents that public health professionals found useful by asking for two examples of documents used at work in the last month. Demographics were also collected on age, gender, degrees, position, and public health experience.

Public health professionals were recruited for the survey via four professional email discussion lists:

- PHLNut for public health nutritionists
- PHNurses for public health nurses
- PHSocialWork for public health social workers
- PHAdm for public health administrators

We sought a sample of interested practitioners, dis-

tributed across a range of public health occupations and positions. The self-selected nature of the sample was viewed as a virtue: because we wanted user opinion, we reasoned that interested volunteers were more likely to both take the time to respond and to have opinions about the subject at hand.

Thirty participants were chosen to take part in this survey. Selection was based on the order of their responses, as well as professional background and work experience, to reflect, as much as possible, the professional make-up of the public health workforce [8]. Twenty-three of the selected participants agreed to participate. Participants included public health nurses, administrators, policy makers, and educators. Three participants were selected to pretest the survey. After obtaining informed consent, twenty participants were sent the revised survey and hard copies of four public health documents from the training set. Fifteen documents in all were evaluated. At least three participants reviewed each document, and each participant evaluated at least one NYAM, one state, and one county document. No two people evaluated the same set of documents. Fifteen of the twenty participants completed the study (75%).

Integrated model development

The results from the user survey were integrated into a comprehensive model of key elements.

RESULTS

Document analysis and preliminary model development

Our analysis of 320 documents resulted in a detailed picture of the characteristics of public health gray literature, as well as a preliminary model of elements in the documents relating to public health interventions.

The gray literature documents in our training collection presented broad and variable ranges of format, level of content, and subject matter. The collection included newsletters, guidelines, annual reports, policy statements, fact sheets, and data sets. The length of many documents ranged from a single page to more than 100 pages. Often a report consisted of multiple electronic files. In fact, 14% of the documents consisted of multiple files. These electronic document "bundles" often included statistical data and graphic images with bibliographic information, such as publishing agency. Of the single electronic file documents, 27% were in hypertext markup language (HTML) format and 57% were published in portable document format (PDF) (2% other formats). A large number of documents were text reports with some figures and tables.

As we examined the documents, we looked for the presence of metadata—descriptive information about the documents such as title, creator, and subject. We distinguished between two kinds of metadata, "formal" metadata and metadata "in situ." Formal metadata are elements that have been assigned by the documents' creators and placed in the document header. Metadata in situ are descriptive elements about the

document found in the document itself. We found almost no formal metadata in the documents in the training collection. These documents also had widely varying levels of metadata in situ, located in widely varying positions in the documents. The potential of NLP techniques to locate metadata elements in situ is one of the strong reasons for exploring this methodology.

User survey

Identification of key elements. The primary purpose of the user survey was to elicit user opinion concerning the key metadata elements relating to public health interventions that would be useful to public health practitioners. This information was gathered using three approaches, each embodying a different elicitation technique.

The first question asked participants to choose and rank, from a list of standard bibliographic elements, the three elements that were most helpful in determining whether or not to review the whole document. Of the elements listed for ranking, six described characteristics of the document (e.g., title, date, etc.) and two described the content (subject headings and table of contents). A place was also provided for "other" elements of the user's choice.

"Title," a bibliographic descriptive element, was listed as most important, followed by "publication date," also a bibliographic descriptive element, and "subject headings," a content element. Although "abstract" was not included in the survey's list of bibliographic elements, three participants entered it under "other" as the most important element. Thus, users, choosing from a list of elements familiar from the established approach to literature retrieval, chose both descriptive and content elements as most important.

The second approach asked participants to underline or highlight the information that they felt was most important. Interestingly, participants primarily highlighted or underlined text that provided specific details rather than more abstract summary statements. The third approach asked participants to write abstracts of the articles, including key information necessary to determine if the article would be useful to their work. The participants wrote abstracts that ranged from two to seven sentences in length. The concepts listed in the abstract were generally at a higher level than the body of the text. When statements were taken directly from the text, they were generally drawn from the executive summary or conclusions. Although participants underlined and highlighted very specific details in the print text, the abstracts were written at a higher level of abstraction and tended to summarize or generalize from the more specific.

Although abstracts created for the same article varied from participant to participant, some notable trends guided us in model development and in the later task of assigning priorities to elements.

1. All of the participants' abstracts included a "problem statement" with a description of the public health problem or issue addressed in the document.

Table 2
Document resources recently used at work

Source	Peer-review/ commercial	Gray literature	Total
Print	6 (19%)	5 (16%)	11 (35%)
Internet	7 (22%)	11 (36%)	18 (58%)
Not provided	1 (3%)	1 (3%)	2 (6%)
Total	14 (44%)	17 (55%)	31

- All abstracts provided a "description of the intervention" or purpose of the report.
- Most abstracts mentioned the "document type," such as policy brief, progress report, or update.
- When articles included demographic parameters, such as "target population," these were included in the abstracts.
- When articles included "results," they were summarized in the abstract.

Document use. The survey also sought an indication of the extent to which this sample of public health professionals used the gray literature. In response to a request to name 2 documents used in the last month that were important to their work, participants provided 31 document titles and sources. Over half (55%) of the 31 documents listed were gray literature. The Internet was a frequent source of both journal and gray literature documents. Nearly two-thirds (65%) of the named documents were obtained through the Internet, and the source of 11 of 17 of the gray literature documents (65%) was the Internet. These results are summarized in Table 2.

Frequently cited Internet sources for gray literature documents included the Websites for the Centers for Disease Control and Prevention (CDC), the United States Department of Health and Human Services (DHHS), and state departments of health (DOH). Meeting notes, local county reports, and legal documents from print sources were also listed.

One-third of the named documents were from peer-reviewed journals such as the *New England Journal of Medicine (NEJM)*, the *Journal of the American Medical Association (JAMA)*, and *Morbidity and Mortality Weekly Reports (MMWR)*. One textbook was listed. A full list of the reported documents is provided in Table 3.

Integrated model development

The results of the content analysis and of the user survey were integrated into a model of key elements in gray literature documents relating to public health interventions (Figure 1). Elements identified by the users as important are marked with an asterisk. Thus, users' input helped prioritize elements for automatic extraction from the document text, using NLP rules, to form a document surrogate that would be part of a searchable database of such records. The priorities for NLP rule writing were:

- title
- description of the problem or issue

- description of the intervention
- results
- target population
- geographic location
- document type
- date of publication

DISCUSSION/CONCLUSIONS

Our approach to developing a model of key elements relating to public health interventions was to work from the data up, with input from the intended users, rather than the approach used in other metadata projects where a top-down model was imposed [9]. An analysis of a relatively large document collection provided us with an initial framework for a model, and the user study provided guidance on the key elements for extraction from the text and use in creation of a document surrogate. The gray literature for public health is vast and diverse, and, although our evaluation was by no means exhaustive, it was large enough to provide a sense of the domain.

The participants in the user study were educationally and academically diverse, consistent with what is known about the public health workforce [8]. While the size of the user group was small, it reflected the relevant work force. Participants were a highly educated and varied group professionally, who were interested enough in public health literature to participate in a study requiring a considerable amount of effort and who, the study showed, used gray literature documents to do their work.

The results of our analysis of 320 documents confirms that the gray literature in the domain of public health is both broad in subject matter and variable in format. The user study supports the premise that gray literature is important to the everyday work of public health professionals. While the Internet is an increasingly important source of both peer-reviewed journal articles and gray literature, traditional print sources continue to be important.

A central finding from the document analysis was that public health gray literature can be described using the concept of the public health problem and elements relating to interventions used to address that problem. This approach to document description, identified during the analysis phase, was confirmed by the study participants, who indicated that description of the public health problem and description of the intervention were essential elements in determining if the full document is useful. Additionally, the participants indicated that title, abstract, document type, target population, and geographic location are important.

The underlying purpose of this first phase of our project was to identify key elements to automatically extract from the text of the documents and use them to create a document surrogate. An example of such a surrogate, based on the elements prioritized so far, appears in Figure 2.

It is important to note that the surrogate description proposed above, like established bibliographic tradi-

Table 3
Documents used for work by participants in the last month

Title	Produced by	Print/source	Gray literature
Postpartum Exercise and Food Intake: The Importance of Behavior Specific Self-efficacy	<i>Journal of the American Dietetic Association</i>	Print/personal subscription	
Diabetes Education for African American Women	<i>Diabetes Educator</i> (journal)	Print/from medical library	
Surgeon General's Call to Action to Prevent and Decrease Overweight and Obesity	Centers for Disease Control and Prevention (CDC)	Print/sent by the CDC	Gray
Guidelines for Comprehensive Programs to Promote Healthy Eating and Physical Activity	Department of Health and Human Services (DHHS), Nutrition and Physical Activity Work Group	Print/sent by the CDC	Gray
2000 Census: Population Statistics	US Census Bureau	Website	
Tuberculosis (TB) Core Curriculum	CDC	Website	
TB pamphlets, multiple languages	Minnesota Department of Health (DOH)	Website	Gray
Diabetes Prevention Program Research Group	<i>New England Journal of Medicine (NEJM)</i> (journal)	Website/links from PH.Nutrition email list	
Community Nutrition Institute (CNI) Newsletter	<i>Nutrition and Clinical Care</i> (journal)	Website/found by Internet search	
Environmental Nutrition			
Ross Mother's Survey	Ross	Website	Gray
Health People 2010		Print/personal subscription	
School Nutrition on Dietary Assessment Study II—Summary of Findings	US Department of Agriculture (USDA) Report	Website	Gray
Physical Activity for Children	National Association for Health, Physical Education, and Fitness	Print	Gray
Guidelines for School Health to Promote Lifelong Healthy Eating	<i>Morbidity and Mortality Weekly Reports (MMWR)</i>	Website	
National Institute on Disability and Rehabilitation Research (NIDRR) Disability Statistics	NIDRR	Website	
MMWR General Recommendations on Immunizations	CDC	Website/CDC	
Content and Design Attributes of Anti-Vaccine	<i>Journal of the American Medical Association (JAMA)</i> (journal)	Website/CDC	
Writing Tips	The Fog Index	Website <www.fpd.finnop.umn.edu>	Gray
The Health of Washington State	Washington DOH	Website/CDC	
CDC Synergy Health Communications/Infections	CDC	Website/CDC	
Georgia Epidemiology Report Prenatal Care Provider Survey: Screening and Treatment of Infections During Pregnancy	Georgia Department of Health	Website/found through Google search	Gray
HIV in the United States at the Turn of the Century: An Epidemic in Transition	<i>American Journal of Public Health</i>	Print/provided by boss	
Documents on Training of the Public Health Workforce		Website <www.trainingfinder.org>	Gray
Women's Health Initiative	<i>JAMA</i> (journal)		
Banning Artificial Nails from Health Care Settings	<i>American Journal of Infection Control (AJIC)</i> (journal)	Print/personal subscription	
Antimicrobial Resistance in Pierce County	Tacoma-Pierce County Health Department (TPCHD) Antibiotic Task Force	Print/Proceeds of June meeting	Gray
Early Statistical Detection of Anthrax Outbreaks by Tracking Over the Counter Medication Sales	<i>Proceedings of the National Academy of Science (PNAS)</i> (journal)	Website/email list	
Regional Water Association of Pierce County Meeting Minutes	Pierce County	Print/provided by coworker	Gray
The Future of Public Health	Institute of Medicine	Print	
Oregon Reportable Disease Statistics	Oregon Department of Human Services (DHS)	Website/Oregon DHS	Gray
The Model State Emergency Health Powers Act	Center for Law and the Public's Health Georgetown/Johns Hopkins University		Gray

tion, contains both descriptors of the document for identification purposes (title, publication date) and descriptors of the content (target population, health problem). The descriptive document identifiers, culled from an analysis of gray literature documents, align with standard metadata schemes.

The surrogate above, however, represents a departure from established bibliographic tradition in the approach to the description of content. Instead of the use of human-assigned, controlled vocabulary subject terms, the surrogate provides structured content elements, based on those that public health professionals have identified as useful.

In the next phase of our project, the model will be

used to guide the development of NLP algorithms that automatically extract these elements from text and create a rich and yet concise document surrogate.

FUTURE DEVELOPMENT

We have characterized our approach to the problem of access to public health gray literature as exploratory. It is, in fact, exploratory along several dimensions. In the phase of the project reported here, we have investigated and proposed an alternative approach to describing the content of public health gray literature. The approach has been developed from a combination of literary and user warrant.

Figure 1
Elements in gray literature documents relating to public health interventions

INTERVENTION ELEMENTS			
Problem/issue*			
Background information (reports/statistics/guidelines/protocols/recommendations)			
Description of intervention*			
Organizations	Intervention type	Methods	Outcomes
Sponsoring/funding/affiliated Governmental: ■ Federal ■ State ■ County Nongovernmental: ■ For-profit ■ Nonprofit	■ Education ■ Prevention ■ Treatment ■ Surveillance	Date/duration Setting ■ Individuals ■ Practitioners ■ Clinics ■ Hospitals ■ Institutions ■ Community Target population* ■ Age ■ Ethnicity ■ Gender ■ Employment ■ Geographic Location* ■ Status: Socioeconomic, insurance	■ Results/findings* ■ Knowledge increase ■ Behavioral change ■ Health Status change ■ Recommendation or guideline ■ Evaluation
DOCUMENT ELEMENTS			
Document type*		Bibliographic elements	
■ Guidelines ■ Newsletter ■ Program reports ■ Meeting notes ■ Policy brief ■ Statistics/data ■ Fact sheet	■ Title* ■ Creator ■ Publishing agency ■ Publishing date* ■ URL ■ Length of documents		

* Elements identified by the users as important.

The next exploratory stage involves the automatic extraction of the elements identified in the model, both content and descriptive identifiers. This stage will encompass adapting NLP systems already in use to the model described in this paper. The system will then be used to extract these elements from the reserved test collection of documents. Public health professionals can then evaluate the resulting surrogates for usefulness.

Once successfully completed, these two phases will constitute proof of concept for this alternate approach to the problem of access in public health. At this point, empirical exploration and testing will have other issues, which can be raised only briefly here. One issue is the identification of pertinent public health gray literature documents over the vast range of the Internet as an automated way to build a sufficiently large base of documents from which to create surrogates. It is likely that the model, and resulting rules, developed here can provide the framework for augmented searching for such documents, but that is a further exploratory project.

Another area for research is an effective means of searching and displaying document surrogates. Dis-

plays of search results could conceivably go beyond retrieval of individual document records to listing by key element. For example, one might search and retrieve a list of problem areas sharing certain characteristics or a list of interventions in a particular problem area. Data might also be displayed textually or graphically. Such capabilities would be able to reveal aspects of the surrogates in the database as a whole, showing, for example, the distribution of intervention elements across certain demographic features. Thus, the system has the potential for revealing the shape of the gray literature collectively, as well as document by document.

Looking even further ahead, if a demonstration system is developed with sufficient identification and coverage of gray literature documents, it will be possible to investigate, in a more definitive way than previously possible, which gray literature, if easily available to public health professionals, is most useful to them. At the heart of this issue is really whether the diversity and variety of documents outside commercially established outlets, with their processes for review and quality control, will benefit public health professionals. Only when the range of such documents can be

Figure 2

Sample document surrogate based on natural language processing (NLP) extraction of key elements

Title	Health Privacy Principles for Protecting Victims of Domestic Violence
Problem	Advocates have long emphasized the importance of detailed documentation of abuse in medical records and the importance of maintaining the confidentiality of all the records of battered women.
Description	Health information regarding victims of domestic violence should respect patient autonomy and confidentiality and serve to improve the safety and health status of victims of domestic violence. This paper describes this fundamental principle and related issues in more detail and offers advocates, providers, and policy makers specific privacy principles to help maximize the privacy of sensitive health information for victims of domestic violence.
Target population(s)	Battered women Victims of domestic violence Policy makers
Geographic focus	National: United States
Document type(s)	Guideline Report
Publication date	October 2000
Document address	www.healthprivacy.org/usr_doc/37713.pdf

searched and retrieved in a sufficiently targeted fashion can the issue of usefulness be thoroughly investigated.

ACKNOWLEDGMENTS

The Robert Wood Johnson Foundation Grant number 04668 supported this work. The authors express spe-

cial thanks to Sue Corieri for preparing and organizing this manuscript.

REFERENCES

1. LASKER RD. Challenges to accessing useful information in health policy and public health: an introduction to a national forum held at the New York Academy of Medicine, March 23, 1998. *J Urban Health* 1998 Dec;75(4):779-84.
2. LASKER RD. Strategies for addressing priority information problems in health policy and public health. *J Urban Health* 1998 Dec;75(4):888-95.
3. O'CARROLL PW, CAHN MA, AUSTON I, SELDON CR. Information needs in public health and health policy: results of recent studies. *J Urban Health* 1998 Dec;75(4):785-93.
4. COMMITTEE ON ENVIRONMENTAL EPIDEMIOLOGY, NATIONAL RESEARCH COUNCIL. *Environmental epidemiology: use of gray literature and other data in environmental epidemiology*. v.2. Washington, DC: National Research Council and National Academies Press, 1997.
5. LIDDY ED, SUTTON S, PAIK W, ALLEN E, HARWELL S, MONSOUR M, TURNER AM, LIDDY J. Breaking the metadata generation bottleneck: preliminary findings. *International Conference on Digital Archives. Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries*, 2001. New York, NY: ACM Press, 2001.
6. TIMMRECK TC. *Dictionary of health services management*. 2nd ed. Owing Mills, MD: National Health Publishing, 1987.
7. BRADLEY, J. Applied information quality: a framework for thinking about the quality of specific information. *J Urban Health* 1998 Dec;75(4):864-77.
8. HEALTH RESOURCES SERVICES ADMINISTRATION (HRSA). *The public health workforce: enumeration 2000*. New York, NY: Center for Health Policy, Columbia University School of Nursing, 2000.
9. YILMAZEL O, FINNERAN CM, LIDDY ED. MetaExtract: an NLP system to automatically assign metadata (2004). *ACM/IEEE-CS Joint Conference on Digital Libraries. Proceedings of the 4th ACM-IEEE-CS Joint Conference on Digital Libraries*. New York, NY: ACM Press, 2004.

Received November 2004; accepted December 2004