

# MeSH on Demand as a Tool to Enhance Keywords in Disaster Lit Records

**Sarah Clarke, NLM Associate Fellow 2018-2019**

**December 18, 2018**

**Project Sponsored by Siobhan Champ-Blackwell and Stacey Arnesen, Disaster Information Management Research Center, Specialized Informational Services Division, National Library of Medicine**

## Contents

Abstract.....	3
Background .....	4
Objective .....	4
Methodology.....	4
Limitations.....	5
Results.....	5
Discussion.....	8
Acknowledgements.....	8
References .....	10
Appendix I: Original Project Description.....	11

## Abstract

**Objective:** The purpose of this project was to test whether *MeSH on Demand* would be a useful tool for generating vocabulary to add to *Disaster Lit* records.

**Methods:** Twenty-Five *Disaster Lit* records of varying formats were accessed and sample text for each record was processed through *MeSH on Demand*. *MeSH on Demand* suggestions were saved, and data was entered in an Excel spreadsheet. Data was analyzed to determine whether suggestions accurately characterized the record.

**Results:** *MeSH on Demand* produced an average of 16.68 suggested vocabulary per processed record. On average, the MeSH Analysis score was 2.33, and the PubMed/MEDLINE Similar Citations Analysis score was 2.3. There were 19 instances where *Mesh on Demand* generated good MeSH Vocabulary and PubMed/MEDLINE Similar Citations. The number of suggested MeSH influenced the MeSH Analysis score – the more MeSH suggestions, the higher the score.

**Conclusions:** The use of *MeSH on Demand* to generate vocabulary would be a helpful tool to Disaster Information Management Research Center (DIMRC) staff and will increase the strength of results returned to users when performing a search. The suggested PubMed/MEDLINE Similar Citations also produced by *MeSH on Demand* would additionally enhance *Disaster Lit* records by allowing users to easily navigate to peer-reviewed literature of similar scope.

## Background

*Disaster Lit: Database for Disaster Medicine and Public Health*<sup>1</sup> is a National Library of Medicine (NLM) database intended for professional audiences responsible for public health emergency management.<sup>2</sup> Records link to original sources and provide reports, guidelines, trainings, factsheets, etc. from non-commercial publishing sources such as government agencies, educational institutions, non-government organizations, and associations.<sup>2</sup> *Disaster Lit* is managed by the Disaster Information Management Research Center (DIMRC), part of the Specialized Information Services Division at the NLM.

MeSH (Medical Subject Headings) is the NLM controlled vocabulary used for indexing articles in PubMed.<sup>3</sup> *MeSH on Demand*<sup>4</sup> is a NLM tool, which uses the NLM Medical Text Indexer (MTI) to identify MeSH vocabulary within a selected text.<sup>5</sup> MTI combines human expertise and Natural Language Processing technology to assist in biomedical literature curation.<sup>6</sup> MTI is the result of a close collaboration between the NLM Index Section, Lister Hill National Center for Biomedical Communications, and the Office of Computer and Communications Systems at the NLM.<sup>6</sup> *MeSH on Demand* also produces a list of related PubMed citations, which is generated using NCBI's PubMed Related Citations algorithm.<sup>5</sup>

## Objective

The purpose of this project was to test whether *MeSH on Demand* would be a useful tool for generating keywords to add to *Disaster Lit* records. Currently, record structure in *Disaster Lit* is limited to title, source, date published, format, annotation, URL, authors, publication type, access notes, ID, and whether the resource include research tools.<sup>2</sup> Subsequently, these are also the searchable fields for the record.

## Methodology

A sample of 25 recently added *Disaster Lit* records (as of October 24, 2018) were selected to be processed through *MeSH on Demand* for this project. This sample contained the most common format types within the database: PDF, video or multimedia, and text.

Each record was accessed in the database and the source opened. *MeSH on Demand* allows for up to 10,000 characters to be entered for processing,<sup>5</sup> therefore text to be processed in *MeSH on Demand* needed to be selected from the source. In most cases, the text selected for processing was an executive summary or a website summary (9 instances) or the document in its entirety was processed (7 instances). The remaining text selections pulled from places reflecting the unique sources: detail of outcome (1 instance), overview and webinar objectives (1 instance), introduction (1 instance), background (1 instance), discussion and public health implications (1 instance), purpose of the project (1 instance), first page of document (1 instance), or annotation in *Disaster Lit* (2 instances).

An Excel spreadsheet was created to collect the following information: *Disaster Lit* ID#, title, URL, format, text submission, suggested MeSH vocabulary (list), and number of suggested MeSH. Additionally, three columns are included in the spreadsheet scoring the suggested MeSH vocabulary, the suggested PubMed/MEDLINE similar citations, and the average of the two. And finally, there is a column for notes for each record.

Scoring the suggested MeSH vocabulary and the suggested PubMed/MEDLINE similar citations was based off a scale of 1-3 was used: (1) poor, (2) acceptable, and (3) good. MeSH Analysis scores were determined by the number of suggested MeSH vocabulary, the specificity of the vocabulary, and

whether suggested vocabulary were currently located in the annotation or title of the record in *Disaster Lit*. Suggested PubMed/MEDLINE similar citations scores were determined by manually scanning the titles to determine their relevance to the *Disaster Lit* record.

### Limitations

The sample did not include two of the format types: image and presentation slides. Only one individual was involved in scoring suggested MeSH vocabulary and the suggested PubMed/MEDLINE similar citations.

### Results

The 25 sample *Disaster Lit* records were processed through *MeSH on Demand* and individually produced a range of 4 to 40 suggested vocabulary per record. The overall average of number of suggested vocabulary per processed record was 16.68.

Processed records that resulted in few, generic, or duplicative *Mesh on Demand* suggested vocabulary received a low MeSH Analysis score. Processed records that *Mesh on Demand* suggested off-topic PubMed/MEDLINE Similar Citations, received a low PubMed/MEDLINE Similar Citations score. On average, the MeSH Analysis score was 2.33 and the PubMed/MEDLINE Similar Citations Analysis score was 2.3.

As seen in Figure 1, the number of suggested MeSH influenced the Average Analysis Score – the more MeSH suggestions, the higher the score.

Of the six records with ten or fewer MeSH suggestions produced, three were video or multimedia format – the only three records of that format in the sample (see Figure 2).

The results show that of the 25 sample *Disaster Lit* records processed through *MeSH on Demand*, 19 produced good MeSH Vocabulary and PubMed/MEDLINE Similar Citations; 4 produced acceptable MeSH Vocabulary and PubMed/MEDLINE Similar Citations; and 2 produced poor MeSH Vocabulary and PubMed/MEDLINE Similar Citations (see Figure 2).

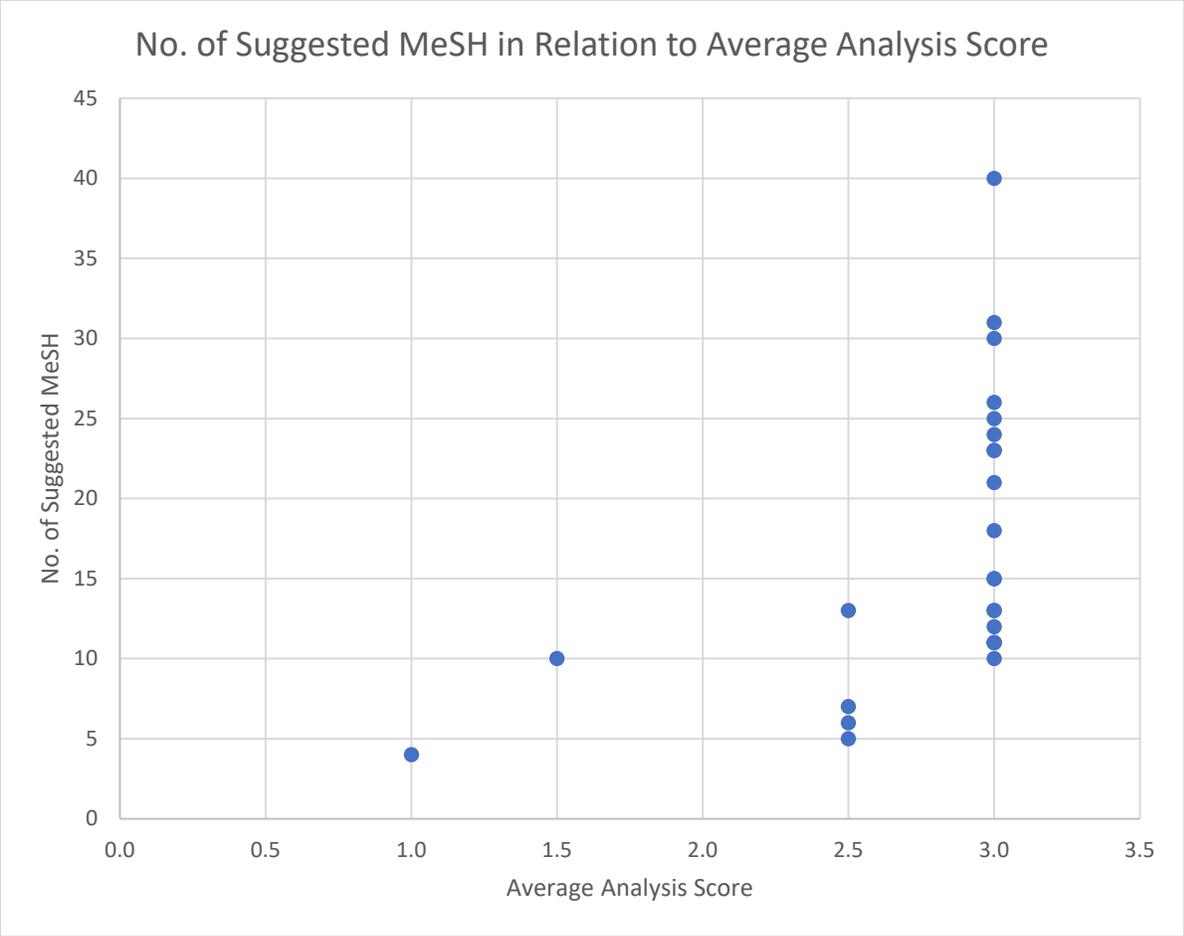


Figure 1: No. of Suggested MeSH in Relation to Average Analysis Score

### Text Input and MeSH on Demand Results

Disaster Lit ID#	Format	Text Submission	No. of Suggested MeSH	Average Analysis Score
18015	Video or Multimedia	Annotation in Disaster Lit	4	1.0
17968	Video or Multimedia	Annotation in Disaster Lit	10	1.5
18019	PDF	First Page of Document	5	2.5
17948	PDF	Background	6	2.5
18090	Video or Multimedia	Overview & Webinar Objectives	7	2.5
18096	PDF	Exec. Summary	13	2.5
18099	PDF	Detail of outcome	10	3.0
18077	PDF	Exec. Summary	11	3.0
18013	Video or Multimedia	Website Summary	11	3.0
18008	PDF	Purpose of the Project	11	3.0
17997	PDF	Exec. Summary	12	3.0
18059	Text	Entire Doc	13	3.0
18055	Text	Intro	13	3.0

Disaster Lit ID#	Format	Text Submission	No. of Suggested MeSH	Average Analysis Score
17987	PDF	Exec. Summary	15	3.0
18009	PDF	Website Summary	15	3.0
18095	PDF	Website Summary	18	3.0
18020	PDF	Entire Doc	21	3.0
18068	PDF	Entire Doc Except End Notes	23	3.0
17950	PDF	Discussion & Pub. Health Implications	23	3.0
17953	PDF	Entire Doc	24	3.0
18012	PDF	Entire Doc	25	3.0
18075	Text	Entire Doc	26	3.0
18014	PDF	Exec. Summary	30	3.0
18018	PDF	Entire Doc	31	3.0
18080	PDF	Exec. Summary	40	3.0

Figure 2: Text Input and MeSH on Demand Results

## Discussion

The MeSH vocabulary suggested through *MeSH on Demand* for each of the sample records produced relevant keyword suggestions. One benefit is the addition of countries, which can assist users to finding specific information for countries of interest.

*Disaster Lit* users would benefit from having MeSH vocabulary added to the records. Currently, searches are limited to text in the title and annotation fields of each record. By adding a record field for MeSH vocabulary, users can search an additional set of relevant keywords. *MeSH on Demand* produced a list of all the countries found in the text submitted for processing, so the addition of MeSH would be very helpful for any users looking for guidance or information pertaining to a specific country. Similarly, *MeSH on Demand* returned specific viruses located in the text submitted for processing that were not mentioned in the record annotation.

NLM Director, Dr. Patti Brennan, blogged about the goal of “One NLM” – aligning and integrating NLM towards a shared vision.<sup>7</sup> With the integration of *MeSH on Demand* for Disaster Lit records, PubMed/MEDLINE Similar Citations can also be utilized, and users could be bridged to additional NLM resources of relevance. This implementation would meet the goal of “One NLM” to accelerate discovery of biomedical knowledge in assisting in addressing global health needs by bringing together NLM’s various resources.

According to Stacey Arnesen at DIMRC, it currently takes staff an average of 20 minutes to enter a new record (including citation and annotation) into the database.<sup>8</sup> It takes *MeSH on Demand* 30 to 45 seconds to process text and provide suggestions,<sup>5</sup> therefore including the *MeSH on Demand* as part of the process for entering new records into the database would take additional minimal time and would produce an average of 16.68 suggested vocabulary and ten PubMed/MEDLINE Similar Citations to add to the record. DIMRC staff would need to select which information within the resource best describes the resource and copy and paste it into *MeSH on Demand* and then add the suggested MeSH vocabulary into the record. The process of adding suggested PubMed/MEDLINE Similar Citations may take a little more time as they would need to be entered and linked to their source. If DIMRC staff decide to pursue these recommendations, they could work with the MTI team at Lister Hill National Center for Biomedical Communications to determine ways to move forward with batching the *MeSH on Demand* processing to further automate the process.

The use of *MeSH on Demand* to generate vocabulary would be a helpful tool to DIMRC staff and will increase the strength of results returned to users when performing a search. The suggested PubMed/MEDLINE Similar Citations also produced by *MeSH on Demand* would additionally enhance *Disaster Lit* records by allowing users to easily navigate to peer-reviewed literature of similar scope.

## Acknowledgements

The author would like to thank the project sponsors, Siobhan Champ-Blackwell and Stacey Arnesen, for their guidance; NLM Associate Fellowship Program Coordinator, Dr. Kathel Dunn, for her continued support; James Mork for his time discussing MTI and *MeSH on Demand*, fellow associate Amelia Llorens for assistance with Excel, and fellow associate Cecelia Vetter for proofreading.

This research was supported in part by an appointment to the NLM Associate Fellowship Program. This program is administered by the Oak Ridge Institute for Science and Education

through an interagency agreement between the U.S. Department of Energy and the National Library of Medicine.

## References

1. Disaster Information Management Research Center, Specialized Information Services, National Library of Medicine. *Disaster Lit* (database). Accessed at <https://disasterinfo.nlm.nih.gov/disaster-lit>
2. Disaster Information Management Research Center, Specialized Information Services, National Library of Medicine. (2018, June 29). About *Disaster Lit*. Retrieved from <https://disasterinfo.nlm.nih.gov/disaster-lit-about>
3. National Library of Medicine. *MeSH on Demand* (tool). Accessed at <https://meshb.nlm.nih.gov/MeSHonDemand>
4. National Center for Biotechnology Information, National Library of Medicine. MeSH. Retrieved from <https://www.ncbi.nlm.nih.gov/mesh>
5. National Library of Medicine. *MeSH on Demand* Frequently Asked Questions. Retrieved from <https://meshb.nlm.nih.gov/ModFAQ>
6. National Library of Medicine. (2018, April 13). Indexing Initiative. Retrieved from <https://ii.nlm.nih.gov/MTI/index.shtml>
7. Patti Brennan. (2018, January 3). NLM Musings from the Mezzanine. Retrieved from <https://nlmdirector.nlm.nih.gov/2017/01/03/one-nlm/>
8. Stacey Arnesen. (2018, November 20). Personal Communication.

## Appendix I: Original Project Description

### MeSH on Demand as a tool to enhance keywords in Disaster Lit records

This is a short project (2 weeks?) to test whether or not MeSH on Demand would be a useful tool for generating keywords to add to Disaster Lit records. The Associate would develop a methodology for this testing, conduct the testing, and report on their findings and recommendations. One possible methodology would be to use a selected sampling of Disaster Lit titles and annotations on a range of topics to generate suggested MeSH vocabulary. The Associate would then determine whether or not the suggested terms accurately characterize the document and would be a useful addition to a record. Disaster Lit documents are not indexed and do not have added keywords. Retrieval is based on searching words in the content-rich fields of each record. Would the use of MeSH on Demand generate additional terms that would enhance the retrieval of records? I used MeSH on Demand, <http://www.nlm.nih.gov/mesh/MeSHonDemand.html>, to generate subject headings for this paragraph; it came up with: Disasters, MEDLINE, Medical Subject Headings, and Names.

The Associate will provide a written report of observations and recommendations of the usefulness of MeSH on Demand for enhancing the description and retrieval of documents in Disaster Lit.