**Chapter 6 – Disaster Publications and Bibliographic Databases**

**Introduction**

Disaster-related events have a long history but the scientific study of these required advances in physical and social sciences. Newer technologies and study methods were introduced in the last 60 years resulting in increasing numbers of studies dealing with these issues. Disaster-related events marshal the efforts and energy from all aspects of society. This multidisciplinary effort implies that the scientific results will be entered into the literature from a variety of sources and stored in numerous bibliographic databases. One search of the term – disaster, using a multi-base search strategy, showed that over 90% of the bibliographic repositories were health-related. This raised the question of ‘best sources’ of scientific literature dealing with disaster events. That question was investigated and the findings reported in this chapter.

**Bibliographic Databases**

Computerized databases containing descriptors and copies of scientific reports were introduced following World War II and by the late 1970’s, a large number of such repositories were in existence. These databases were designed to include documents dealing with specific topics. However, with time, the number of documents and the topics considered have increased exponentially. In addition, a document may contain information dealing with more than one topic. As a result, it could be included in different theme databases making the selection of those containing the most information, relevant to a topic, more difficult.

Table 1 shows the number of articles involving the term – disaster – in selected scientific databases. The period considered was 1990-2011 since the bulk of the work was reported during that time. PubMed contained over 26 thousand documents and the Web of Science followed with over 18 thousand. Vendors of bibliographic databases are known to ‘package’ different databases in order to make subject retrieval more efficient. The database, Illumina, is an example of that format. While this mega-search process is desirable, there are two obvious flaws. The first is duplication of the same document in different databases. The second is the incompleteness of any particular database or package.

**Table 1. Bibliographic Databases and Frequency of Disaster-related Documents.**

|  |  |
| --- | --- |
| **Bibliographic Databases** | **Disaster-related** |
| Emerald | 5804 |
| Illumina | 13016 |
| PubMed | 26294 |
| Sociological Abstracts | 3848 |
| Web of Science | 18359 |

The number of documents identified by a given search statement (e.g., ***disaster, 1990-2011***) ranged, in Table 1, from 3848 to 26,294. Assuming that manual analysis of each required about 15 minutes, the 67,321 documents would take about 16,830 hours. This time estimate does not take into account the fatigue and increased error as this manual process is performed. As such, the search engine solved the challenge of ***retrieving*** information. Monumental searches can be performed in milliseconds. The challenge of ***using*** the information, once retrieved, has not yet been consistently considered. The bulk of the effort continues to be primarily manual.

In choosing databases for study of disaster-related events, an appropriate starting point would seem to be the online version of MedLine, i.e., PubMed. This repository contains over 22 million documents and 26,289 dealt with disaster during the period, 1990-2011. Disaster-related events also have been studied by social scientists and others. Illumina, Sociological Abstracts, and Web of Science underscore the interest in the word – disaster. A search statement containing the term does not retrieve only documents describing natural events or man-made ones. In addition to those categories are documents describing treatment errors, observation errors, and a myriad of other descriptions where the author included the term disaster.

Typically, this search and retrieval problem is addressed by adding more terms in the search statement. An example might be ***disaster AND earthquake***. These Boolean statements, generated by the user, are attempts to match the authors’ use of the terms. Since the user is guessing about the combinations used by the authors, this type of searching is called ***a posteriori*** indicating that the relevance of the retrieved documents can be determined only ***after retrieval***. Search statements of this type would retrieve only those containing each combination. Changing the search statements to cover the various possibilities can result in retrieval of the same document multiple times. While characterizing documents in terms of themes is a popular approach, the fact is that a document may contain numerous topics.

As such, the capture of all of the correct documents and the elimination of all of the incorrect documents can not be guaranteed by the ***a posteriori*** search process. The only way to determine the relevant documents is to look at them all. With the appropriate processing method, this approach might be feasible. If so, the authors’ Boolean combinations could be used in the search statements. This ***a priori*** strategy ensures retrieval of documents containing those terms used in a contextually relevant way. The authors’ Booleans eliminates the problem of the search software arbitrarily creating links where none was implied by the authors.

**Databases Studied:** In considering suitable databases, the Emerald database was a possibility. It was purported to be a leading resource for social science and business. This repository was initiated by University of Bradford and contained 5,804 documents dealing with disaster-related events in the time period considered. Accordingly, this report considered a large (i.e., PubMed) and a moderate sized (i.e., Emerald) concentration of scientific documents dealing with the topic of disaster.

**Analytic Approach:** The Idea Analysis was used to identify, extract, organize, and store essential data from each scientific publication. The resulting data can be used for a variety of purposes. One is the ***degree of agreement between the databases*** in describing disaster-related events. A second is the ***matching of an expert consensus*** dealing with disaster and health. A third is a ***transparent approach*** to duplicating an expert-derived creative result.

**Ideas as Building Blocks:** These analyses use ***ideas as the building blocks***. Ideas have been a potent force in enhancing progress throughout history. The definition of this important notion is surprisingly vague and refers to something formed in the mind. This non-specificity refers to the act of ***creating*** an idea rather than ***recognizing*** an existing one. This distinction is important. Once the author has provided the idea in the form of a combination of informative terms within a sentence, software can identify and extract it. The procedures include pattern recognition, editing functions (i.e., sorting, copying), and data processing functions (i.e., storing, retrieving, and analyzing). These tasks can be performed by software without fatigue or error, thus, transforming a labor-intensive manual process into one that is rapidly accomplished. The manual processing time of 15 minutes per document can be reduced to about 0.3 minute, with over 95% accuracy and completeness.

The identified ideas can be used to build descriptions of the topic as well as to develop strategies leading to new research studies. The ideas can be classified into dimensions or groupings representing the main components of the topic. These dimensions can be populated by specific ideas and procedures. The combination of major groupings and inherent details enhances the students’ understanding of the topic and enables professionals to rapidly identify gaps and inconsistencies. These serve as a stimulus for new research.

**Disaster Ideas:** The disaster-related literature, entered into PubMed and Emerald during 1990-2011, was analyzed using the idea analysis approach. There were four objectives. The first was to illustrate ***algorithms*** which facilitate separation of the two major phases inherent in knowledge utilization. Those are the technical and intellectual. The former involves capture and organization of the essential data presented in text and provides a transparent, quality-controlled approach to that clerical and mechanical process. The intellectual phase represents the application of the higher cognitive functions and enables the individual to focus on developing appropriate measures, criteria, and decisions necessary in accomplishing those functions.

A second objective is the development of an ***information structure*** representing the terms and relationships in the topic of disaster. The advantages associated with this structure include recognition of frequently cited relationships as well as those that are infrequent or missing. The latter may lead to new research strategies.

A third objective is the ***construction of a resource*** consisting of the vocabulary, relationships and algorithms for processing them. This resource would eliminate the clerical/mechanical component of search and retrieval. It would enhance the individual’s focus on the higher cognitive functions.

Finally, to determine if the ***experts’ results could be duplicated*** using a formal approach to identification, organization and utilization of essential data.

**Comparing Bibliographic Database Contributions:** The contributions from PubMed and Emerald were compared for selected terms. The general term, disaster, gives a gestalt view, while earthquake and tornado show findings dealing with specific components. The term -- disaster – was linked with related terms to form 25,776 ideas in the PubMed set. There were 9,980 ideas involving disaster in the Emerald set. The ideas were ranked from most to least frequently used and divided into quartiles.

**Table 2. Number of Ideas and Quartile Ranks for Informative Terms Each Linked with the Term – *Disaster* – in PubMed and Emerald Databases.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Term** | **PubMed** | **PubMedR** | **Emerald** | **EmeraldR** |
| health | 812 | 4 | 35 | 4 |
| response | 507 | 4 | 158 | 4 |
| mental | 448 | 4 | 33 | 4 |
| medical | 447 | 4 | 6 | 2 |
| provide | 381 | 4 | 97 | 4 |
| management | 381 | 4 | 96 | 4 |
| emergency | 374 | 4 | 18 | 4 |
| experience | 285 | 4 | 68 | 4 |
| nation | 282 | 4 | 28 | 4 |
| natural | 264 | 4 | 235 | 4 |
| person | 262 | 4 | 28 | 4 |
| hospital | 227 | 4 | 4 | 2 |
| planning | 225 | 4 | 151 | 4 |
| problem | 222 | 4 | 9 | 3 |
| public | 211 | 4 | 61 | 4 |
| worker | 205 | 4 | 24 | 4 |
| cause | 201 | 4 | 42 | 4 |
| training | 195 | 4 | 36 | 4 |
| symptom | 195 | 4 |  |  |

Etc. Etc.

Table 2 shows an excerpt of the file containing the informative terms linked with the term – disaster – in sentences from PubMed and Emerald reports. These terms were ranked and divided into quartiles based on frequency of ideas. The frequencies (i.e., number of times authors cited the idea -- disaster 🡪 related term) are shown in columns labeled with the name of the database. The ranks are shown in columns indicated by the name plus an R.

Disaster 🡪 Health ideas were most frequently used in the PubMed database. That same idea in the Emerald database also was in the highest quartile. The ranks offer the opportunity to compare the two data sets even when the frequencies of use seem dissimilar. In the excerpt of ideas shown in Table 2, all of the terms except for symptom occurred in the two databases. Most ideas from Emerald involving these terms were in the fourth quartile.

The distribution of these higher frequency terms is shown in Table 3 with the number of ideas shared in both databases and those only in the PubMed repository.

**Table 3. Comparison of Higher Ranked Terms Linked with Disaster from PubMed and Emerald.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Groups** | **Total Terms** | **PubMed** | **Percent** |
| Environment | 9 | 0 | 0 |
| Events | 14 | 1 | 7 |
| Health | 27 | 10 | 37 |
| Management | 33 | 2 | 6 |
| Personal | 12 | 3 | 25 |
| Research | 8 | 0 | 0 |
| **Total** | **93** | **16** | **17** |

There were 93 terms in the highest quartile group and 16 of those were contributed only by the vocabulary used in the PubMed literature. The remaining terms were contributed by both databases.

**Earthquake Comparisons:** There were 5174 ideas in PubMed and 1551 in Emerald involving earthquake and a related term. Table 4 shows the number of terms linked with earthquake in each major grouping or dimension. There was a total of 83 terms identified from the two databases. PubMed contributed 38 percent of the terms linked with earthquake in the health grouping and 37.5 percent in the people grouping. Emerald contributed 33 percent in the environmental grouping, 41 percent in the management set, 12.5 percent in the people group, and 25 percent in the research group.

**Table 4. Number of Terms Linked with Earthquake from PubMed and Emerald.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Group/Dimension** | **Total** | **PubMed** | **Percent** | **Emerald** | **Percent** |
| Disaster | 10 | 1 | 10 | 0 | 0 |
| Environment | 9 | 1 | 11 | 3 | 33 |
| Health | 26 | 10 | 38 | 0 | 0 |
| Management | 22 | 1 | 4.5 | 9 | 41 |
| People | 8 | 3 | 37.5 | 1 | 12.5 |
| Research | 8 | 0 | 0 | 2 | 25 |

**Tornado Comparisons:** There were 615 ideas in PubMed and 58 in Emerald involving tornado and a related term. Table 5 shows the number of terms linked with tornado in each major grouping or dimension. There was a total of 24 terms identified from the two databases. PubMed contributed 92 percent of the terms linked with tornado. The percent contribution ranged from 50 percent in the research group to 100 percent in the environment, health, people, and management groupings. Emerald contributed no additional terms.

**Table 5. Number of Terms Linked with Tornado from PubMed and Emerald.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Group** | **Total** | **PubMed** | **Percent** |
| Disaster | 6 | 5 | 83 |
| Environment | 3 | 3 | 100 |
| Health | 6 | 6 | 100 |
| Research | 2 | 1 | 50 |
| People | 3 | 3 | 100 |
| Management | 4 | 4 | 100 |

**Idea Graphs:** While comparison of frequency of using ideas is relevant, a better understanding of similarity and difference can be accomplished by considering the ideas and their organization. The idea map shows that. The terms linked with the central one are shown on the periphery and are classified into groups representing major components of the topic. This idea structure shows major categories and the ideas that serve to enhance meaning or function. These ideas can be assigned importance using frequency of use (i.e., a consensus among experts) or perceived contribution to understanding the topic (i.e., specificity).

**Figure 2. Terms Linked with Tornado from PubMed and Emerald, 1990-2011, In an Idea Map Format.**

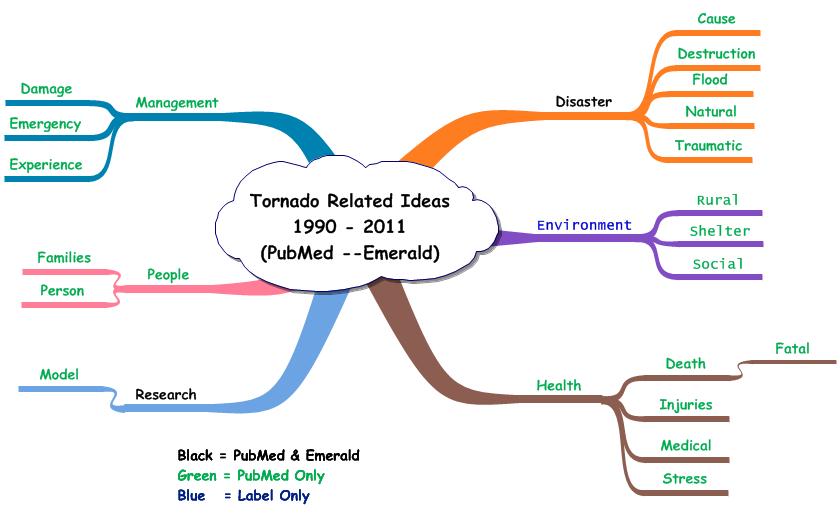
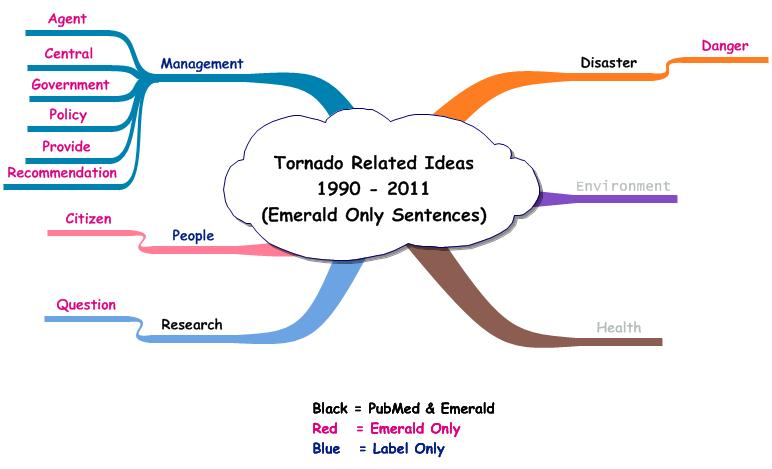


Figure 2 shows the terms linked with tornado from the two databases. The terms in both are denoted in black and those from PubMed alone in green. The findings showed that if Emerald ideas were the only ones considered in describing tornado events, the ideas would be restricted to tornado, respectively, with disaster and research. Table 6 shows the idea records from the Emerald database for the sentences containing either tornado 🡪 disaster or tornado 🡪 research.

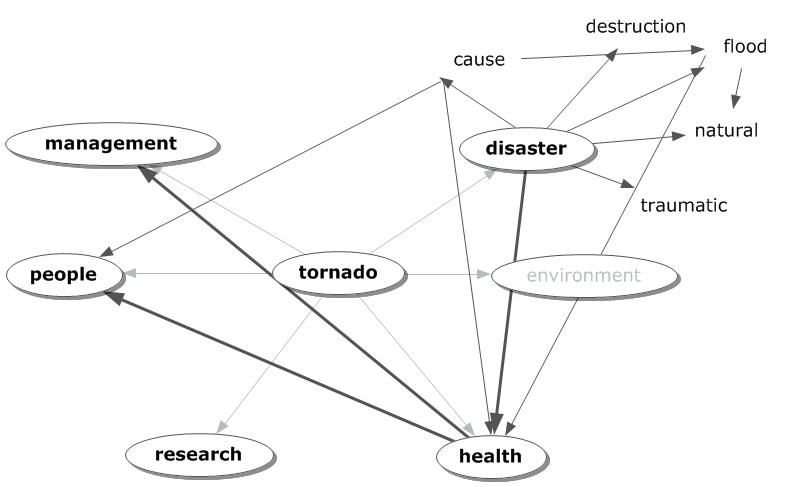
Figure 3 shows the idea map representing the ideas dealing with disaster and research linked with tornado. Environmental and health ideas are absent. The predominant set of ideas deals with management.

**Figure 3. Idea Map of Tornado Ideas based on Emerald Sentences.**



Another display is the ***idea network*** involving higher frequency terms serving as nodes and the lines between them representing observed relationships among the terms. This graphic is much like a geographical map showing the cities and towns and the roads connecting them. Figure 4 shows a partial network graphic. Some of the relationships with tornado are considered. The principle nodes are shown together with the paths connecting them. The ideas shown as connections between tornado and the other bolder labels are the simple pairs, e.g., tornado with disaster. In addition, there are heavier links between health with disaster, people and management. These ideas are composed of three terms, e.g., tornado 🡪 health 🡪 disaster. The three must occur within the same sentence.

**Figure 4. Tornado Idea Network Example with Principle Nodes Illustrated.**



**Table 6. Idea Records from Emerald Database.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tornado** | **Related** | **Year** | **Ident** | **Sentence** |
| tornado | agent | 2011 | 2006454 | 5 |
| **tornado** | **disaster** | **2011** | **2006454** | **5** |
| tornado | government | 2011 | 2006454 | 5 |
| tornado | policy | 2011 | 2006454 | 5 |
| tornado | recommendation | 2011 | 2006454 | 5 |
| tornado | central | 2011 | 2006454 | 2 |
| tornado | citizen | 2011 | 2006454 | 2 |
| tornado | danger | 2011 | 2006454 | 2 |
| tornado | government | 2011 | 2006454 | 2 |
| tornado | provide | 2011 | 2006454 | 2 |
| tornado | question | 2011 | 2006454 | 2 |
| **tornado** | **research** | **2011** | **2006454** | **2** |

Table 6 shows an example of this. The table also shows the format of the idea record. Tornado is the primary term. The related terms were linked by the authors to the central one. The document was entered into PubMed in 2006. The identification number assigned by PubMed is given next, followed by the sentence number containing the ideas.

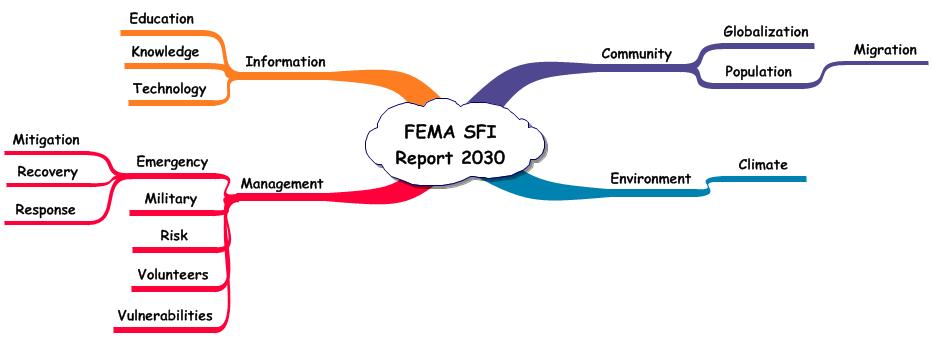
The sentence in question is shown in Exhibit 1. This particular sentence also shows a number of possible combinations of three informative terms as well as larger combinations. The contributing author offered one version of the combination of informative terms. There are numerous others. Depending on the analyst, this sentence offers different ideas that can serve in building a new description.

**Exhibit 1. Sentence 7 illustrating Authors’ Use of Informative Terms in Forming Ideas.**

***To respond properly to a serious type of disaster like a cyclone or a tornado or recurrent devastating flood, the disaster mental health team should be aware of the socio-economic status, local culture, tradition, language and local livelihood patterns.*** [***Choudhury WA***](http://www.ncbi.nlm.nih.gov/pubmed?term=%22Choudhury%20WA%22%5BAuthor%5D)***,*** [***Quraishi FA***](http://www.ncbi.nlm.nih.gov/pubmed?term=%22Quraishi%20FA%22%5BAuthor%5D)***,*** [***Haque Z***](http://www.ncbi.nlm.nih.gov/pubmed?term=%22Haque%20Z%22%5BAuthor%5D)***. Mental health and psychosocial aspects of disaster preparedness in Bangladesh.*** [***Int Rev Psychiatry.***](http://www.ncbi.nlm.nih.gov/pubmed?term=17162693) ***2006 Dec;18(6):529-35. PMID: 17162693***

**Duplicating Experts’ Results:** Capturing the authors’ ideas is based on the premise that these specialists can effectively guide the discovery of new knowledge. The advantages include the development of formalized learning with transparent methods involving application of cognitive functions. The process satisfies the needs of individuals using self-discovery methods. The recent report by FEMA’s SFI group considered the major components of an emergency management program by 2030. Figure 5 shows the elements considered by the consensus of experts participating in the SFI deliberations.

**Figure 5. Major Elements in the FEMA SFI’s Description of Emergency Management Programs by 2030.**



As seen, the major dimensions are community, environment, information, and management. The diversity and changing scene of communities and the population were stressed as critical in accomplishing successful emergency management programs. While there are other environmental threats, the major one of interest to FEMA was climate change and its relationship with the frequency and intensity of natural disasters. The utilization of information was recognized as important and involved the use of newer forms of communication as well as innovations in technology. Management programs emphasized the tasks required to rehabilitate a community after a catastrophic event. In addition, the use of military and volunteers was considered. Risk management and vulnerabilities in supply storage and transportation completed the major elements considered.

These issues were developed as a result of conferences, workshops, and analyses involving large numbers of experts in disaster management. The process has historic precedence but is a labor-intensive and time consuming one. A different approach involved the separation of the knowledge utilization tasks into three segments. The first was the capture and organization of information relevant to the topic. This was performed essentially by software. The second dealt with construction of new arrangements of the ideas and can be performed by technicians. The third involved the application of the higher cognitive functions where the experts were expected to develop measures, criteria, and decision-structures leading to solutions. The intent of this analysis was to determine the experts’ selection of ideas from those available in the scientific literature.

**Table 8. Terms From Sentences Containing Disaster Event 🡪 FEMA Terms and the Source of the Idea – PubMed or Emerald.\*\*\***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **FEMA Term** | **Disaster** | **Earthquake** | **Flood** | **Hurricane** | **Tornado** |
| **climate** | **16** |  | **61** | **5** | **2** |
| **community** | **168** | **15** | **46** | **6** | **3** |
| **education** | **43** | **9** | **13** | **8** | **2** |
| **emergency** | **374** | **34** | **38** | **9** | **17** |
| **environment** | **34** | **4** | **17** | **1** |  |
| **global** | **29** | **3** | **19** | **3** | **1** |
| **information** | **15** | **2** | **2** |  | **1** |
| **knowledge** | **72** | **4** | **14** | **2** |  |
| **management** | **381** | **23** | **66** | **2** | **1** |
| **migration** | **4** |  |  |  |  |
| **military** | **57** | **5** | **1** |  | **1** |
| **mitigation** | **42** | **7** | **15** | **6** |  |
| **population** | **88** | **21** | **49** | **13** | **1** |
| **recovery** | **135** | **14** | **22** | **9** | **1** |
| **response** | **507** | **53** | **53** | **16** | **4** |
| **risk** | **125** | **24** | **26** |  |  |
| **technology** | **31** |  | **2** | **1** |  |
| **volunteer** | **26** | **3** |  | **1** | **1** |
| **vulnerability** | **35** | **7** | **4** | **2** | **1** |

\*\*\****Contributions from PubMed = black, from Emerald = green***

.

**Summary**

A comparison of contributions from two bibliographic databases, PubMed and Emerald, was performed. The essential data consisted of author-presented ideas. In disaster-related events, seventeen percent of the terms used in forming ideas were contributed only by PubMed. Earthquake was the only topic where Emerald contributed vocabulary not included in PubMed. PubMed uniquely contributed 19% of the vocabulary and Emerald, 18%. Hurricane and tornado ideas were increasingly dependent on vocabulary from PubMed. Fifty-five percent and 92% of the terms were only from PubMed in those topics.

In order to determine the validity of this analysis, a report from FEMA SFI was used. Table 8 showed the higher frequency terms used in ideas. The frequency of links found in the scientific literature with each of the disaster-related terms was shown. The ideas were identified by contributing source with most from PubMed and others from Emerald.

One criterion for determining ‘important’ ideas was ***consistency of use*** by authors across the different disaster-related events. The terms satisfying this consistency criterion were: ***community, education, emergency, global, management, population, recovery***, and ***response***.

The number of ideas involving each of these terms with a disaster-related event was shown in Table 8. The number of ideas varied depending on the disaster term. As an example, community related ideas ranged from 3 involving tornado to 168 linked with the general term -- disaster. The term, emergency, ranged from 9 hurricane related ideas to 374 disaster related ones. Management was linked with tornado once and with disaster 381 times.

PubMed was the more frequent source of ideas. However, Emerald contributed key ideas as well. Those were highlighted in green. Disaster and education was reported in sentences from Emerald. The ideas involving risk with disaster, earthquake and flood were from Emerald

A second criterion for selecting ‘important’ ideas could be the ***frequency of occurrence*** of a term linked with a particular disaster-related event. The frequency of ideas involving disaster is an example. Response 🡪 disaster was used 507 times. Emergency and management, respectively, were linked with disaster over 370 times. All of the terms used in the FEMA report were linked with disaster in the PubMed and Emerald literature, thus satisfying the consistency criterion as well. The frequency of ideas ranged from 4 (disaster 🡪 migration) to 507 (disaster 🡪 response). The ideas linked with flood provide another example of the frequency criterion. All but two of the ideas are represented. The missing ones are migration and volunteer, respectively, with flood. The ideas present ranged from 1 (flood 🡪military) to 66 (flood 🡪 management).

An example of a sentence contributing these data was shown in Exhibit 2. The informative terms were highlighted. Those in red were those captured by the software. The two in green were not. ***Immigrant*** and ***Risk*** were in the informative term vocabulary and were missed in processing this sentence. Of the ten informative terms available for capture in this sentence, the software identified correctly 80%. An important advantage to this computerized approach is the determination of accuracy. Overall, the median percent of matching terms (expert vs. software) is 85% (range 60 to 99% depending on writing style).

**Exhibit 2. Sentence from Document 9130864 Containing Earthquake Related Ideas.**

***Selected findings and key conclusions are presented for each of six areas of investigation: service utilisation during the first 6 months post-disaster; patterns of earthquake experience and short-term (6-month) psychosocial outcome; earthquake exposure and medium term (2-year) psychosocial outcome; vulnerability factors and medium-term psychosocial outcome; specific community groups at increased risk (e.g., the elderly and immigrants from non-English-speaking backgrounds); the effects of stress debriefing for helpers.***

[***Carr VJ***](http://www.ncbi.nlm.nih.gov/pubmed?term=%22Carr%20VJ%22%5BAuthor%5D)***,*** [***Lewin TJ***](http://www.ncbi.nlm.nih.gov/pubmed?term=%22Lewin%20TJ%22%5BAuthor%5D)***,*** [***Webster RA***](http://www.ncbi.nlm.nih.gov/pubmed?term=%22Webster%20RA%22%5BAuthor%5D)***,*** [***Kenardy JA***](http://www.ncbi.nlm.nih.gov/pubmed?term=%22Kenardy%20JA%22%5BAuthor%5D)***. A synthesis of the findings from the Quake Impact Study: a two-year investigation of the psychosocial sequelae of the 1989 Newcastle earthquake.*** [***Soc Psychiatry Psychiatr Epidemiol.***](http://www.ncbi.nlm.nih.gov/pubmed?term=9130864) ***1997 Apr;32(3):123-36.***

**Figure 6. Agreement between Research Literature Ideas and FEMA Terms using the Criterion – Disaster 🡪 FEMA Term from PubMed or Emerald.**

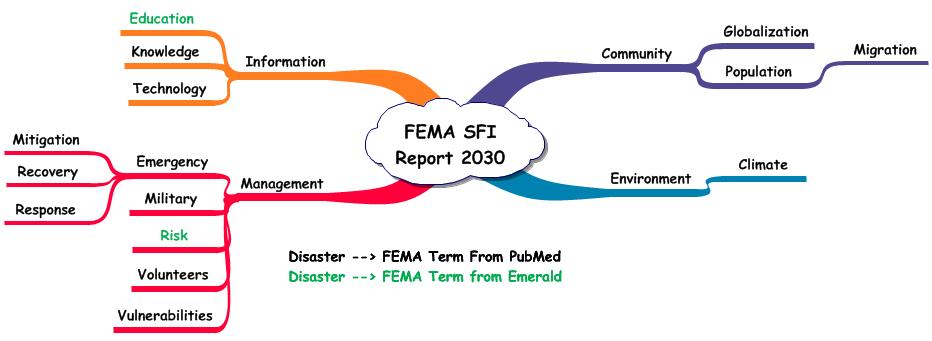


Figure 6 shows the FEMA terms now highlighted to show the agreement with the research literature. This agreement is based on finding ideas involving disaster 🡪 FEMA term from sentences in PubMed or Emerald. That criterion enabled identification of 100% of the terms selected by FEMA experts.

The other ideas shown in Table 8 provide a measure of validity. The general term, disaster, could represent natural or man-made crises. This non-specificity is a concern in interpreting the findings. The idea sets representing earthquake, flood, hurricane, and tornado are more specific as they are focused on the particular topic. Table 8 showed that the terms selected by the FEMA experts were pertinent in these specific sets.

As a result, if one bibliographic database was to be used as the source, PubMed provides sufficient information to effectively consider the relationships in a topic. The links between the type of disaster and health issues are numerous and cover many specific disciplines. In addition to breath of topics, PubMed offers ease in use making it a convenient first choice.