

# PubMed, ClinicalTrials.gov: a critical analysis of new features after three years from the launch of the new release. Results from an interactive training course

Francesca Gualtieri, Silvia Molinari, Ivana Truccolo, Chiara Formigoni for the GIDIF-RBM Working Group

GIDIF-RBM, Monza, Italy

## Abstract

Three years after the launch of the new release of the PubMed database and the new platform for searching clinical trials "ClinicalTrials.gov", GIDIF Academy organized an interactive course aimed at biomedical documentalians and librarians. These portals, with free access, are known to the scientific community as the reference for bibliographic resources of information in scientific literature.

In the classroom, the high motivation of the participants was further powered by the careful analysis of the modernization innovations proposed by the NLM for the two platforms.

The speakers, interacting lively with the learners, highlighted the strengths and analysed the fragilities of the two systems, with the aim of finding possible solutions to obtain effective and safe queries; finally, the collection of opinions offered ideas and room for improvement in the performance of the platforms themselves.

**Key words:** *clinicaltrials.gov; interactive course; search filters; MeSH.*

## Introduction

The GIDIF-RBM (Italian Association of Documentalists and Librarians of the Pharmaceutical Industry and Biomedical Research Institutes) Group was set up informally in 1973 after a meeting of biomedical information professionals and became a non-profit association in 1985. Its aims, as stated in its bylaws, are as follows:

- to promote and protect the image of the information professional;
- to foster schemes for training and updating information professionals in the biomedical and allied fields;
- to contribute to the study of materials and methods helpful to the profession.

The Association under the "GIDIF Academy" project, three years after the announcement of the new release of the PubMed database and the very recent launch of the new platform for searching clinical trials "ClinicalTrials.gov", organized a course aimed at biomedical

documentalians and librarians. These portals, with free access, are known to the scientific community as milestone resources for bibliographic research of information in scientific literature.

The "Umberto Veronesi" Library of the National Cancer Institute of Milan hosted the in-person course on September 29, 2023.

The Faculty Members were experienced documentalists who have long held positions in scientific libraries and National Networks.

In the classroom we tried to discover the performance of the algorithms behind the search queries of the two databases, which are offering simple and user-friendly interfaces to the end-user rather than to researchers or information professionals. The speakers, interacting lively with the learners, highlighted the strengths and analysed the fragilities of the two systems with the aim of finding practical solutions to obtaining effective and safe queries.

*Address for correspondence:* Francesca Gualtieri, GIDIF-RBM, via F. Cavallotti, 126 Monza, Italy.  
E-mail: [direttivo@gidif-rbm.org](mailto:direttivo@gidif-rbm.org)

## Methods and Materials

The training course was organized into two sessions: the morning dedicated to reports that analysed the main changes made to the NLM platform accompanied by interactive interventions with considerable involvement of the learners; the afternoon focused on the analysis of the changes in the construction of search strategies and in the extraction of citations starting from specific research cases.

29 participants, divided in 5 groups, worked to test some search queries and discussed different results (Figure 1).

Time	Activity
09,45	Registrazione partecipanti
10,00	Presentazione dell'evento, F. Gualtieri, Presidente GIDIF-RBM
10,05	Contenuti, thesaurus, algoritmo, risultati. Cosa è cambiato? C. Formigoni
10,30	Come condurre una ricerca e gestire i risultati, S. Pizzarelli
11,30	Pausa
11,45	Focus on.....Clinical Queries, C. Formigoni
12,00	Novità: Proximity tools, V. Scotti
12,15	La nuova piattaforma ClinicalTrials.gov, F. Gualtieri
12,45	Pausa pranzo
13,45	Gruppi di lavoro sui casi di ricerca proposti
15,30	Presentazione dei risultati dei gruppi
16,00	Chiusura dei lavori

**Fig. 1.** Programme of the GIDIF-RBM training course (in Italian).

### Applying proximity indicators

It was explained how bibliographic research works by applying proximity indicators. This type of technique allows you to find two words that are close, in proximity, within a sentence or concept by specifying the distance between the words within a document. Words can occur in any syntactic order within a given range; she considered, as a search example: “colorectal cancer” [field: ~N] where “field” is the search field tag for the fields [title/abstract], N represents the maximum number of words that can appear among the search terms search and “tilde” indicates approximation.

The limitations highlighted using proximity operators are the lack of application of automatic mapping terms simultaneously with the use of proximity operators and the impossibility of specifying that an exact phrase appears within a certain distance from other terms.

### New search filters

It was illustrated how the NLM presents all available news items on the results page after a google-like search, showing the ability to sort the results by clicking on the “display options” drop-down menu, which keeps the sorting criteria unchanged. The list of authors is fully listed by the options “cite”, “save”, and “email”, exploding the “see abstract for full author list” field; this functionality was obtained thanks to the reports that expert users forwarded to the NLM through the Help Desk dialog box. In fact, one of the purposes of this meeting was also to point out needs and corrections to establish a dialogue with the NLM. The functionality of new search filters was also illustrated, including “other” (includes the possibility of excluding preprints [pt] and adding the check for Medline subset [sb]) and the one that allows you to filter in the “article type” category the “systematic reviews”; the old filters are still available in special queries section.

### Clinical query

A different scenario arises when users query PubMed through the “Clinical Query” option.

It was pointed out how the change linked to the mapping and indexing with MeSH performed automatically by artificial intelligence, is substantial when querying the database with this functionality.

The growth of MeSH items and the increase in publications have led the NLM to argue that automated indexing can provide users with timely access to metadata by speeding up the query response process to the growing volume of published biomedical literature.

The selection of indexers by experts has been and will continue to be involved in the refinement of automated indexing algorithms, significantly ensuring the quality required in the indexing process itself. In fact, at NLM, automated indexing has been under development for many years, and the most significant achievement was the development of the Medical Text Indexer (MTI).

MTI has been used to provide suggestions to “human” indexers since 2002 and has been integrated with subsequent care by experts from journals’ editorial board since 2011. With this automation, the period for associating MeSH with Medline records, has been reduced from around 30 days to just 24 hours.

### **ClinicalTrials.gov**

Finally, the new features of the ClinicalTrials.gov database, also offered by the NLM, were presented. The platform launch in August 2023, after a three-year modernisation period that started in 2020 and will end with the retirement of the old platform by June 2024. ClinicalTrials.gov (CTs.gov) is a large library designed to hold information from registered clinical trials of new molecules or existing drugs, for which possible new therapeutic indications are being investigated. The NLM offers limited reviews of the information provided by investigators and sponsors, who remain solely responsible for and owners of the accuracy of the safety data. The search explosion for clinical trial publications, access to the study protocol and the statistical analysis plan is accurate, as is the geolocation of clinical trials, which on the one hand loses the geographic map view of the previous version, and on the other hand becomes more sensitive when queried via the API (Application Programming Interface).

ClinicalTrials.gov is a valuable tool for searching registered clinical studies whose results have not yet been published. Some examples of clinical trial records were analysed, and some exercises were performed during the afternoon section to test results.

### **Discussion**

The interactive in-person mode, and the afternoon exercises to test the various solutions in querying and searching, allowed a high level of analysis of the content covered, as well as the pros and cons offered by the main new releases of the two NLM platforms. The most critical issues exposed in the classroom were also the result of expert use and professional analysis of the two databases. Amongst the trainees, considerable satisfaction with the training day emerged, and the comparison between, and with experts (peers), made them appreciate the added value of attendance.

On the substance, the participants greatly appreciated the reconstruction of the evolution of the ATM (Automatic Term Mapping) algorithm in PubMed and the possibility that remains, in this transitional phase, to deactivate the automatism and compare the results between the search with and without ATM.

What emerged, in the end, is the recognition of the enormous importance of the NIH/NLM platform that continues to make valuable products and services avail-

able, free of charge, to different audiences by taking care of their evolution. ClinicalTrials.gov accessible to researchers, health professionals, sponsors, patients, and citizens, is only the most recent example; the accuracy of this database makes it extremely competitive with other paid-for ones. Such products/services entail a considerable financial investment on the part of the US government, and it is obvious that it is also seeking sustainability using artificial intelligence.

What, in the opinion of the Italian biomedical librarians' community, should not happen is that PubMed, ClinicaTrial.gov, MedlinePlus, MeSH, as well as all the rest of the tools of the NIH/NLM platform, lose their added value in the effort to be "Google-like", i.e., easily accessible to all. Today, the loss of some options/filters and the AI algorithm, which sometimes produces results that are difficult to understand, generate some perplexity.

### **Concluding remarks**

It is precisely these perplexities that GIDIF-RBM wants to turn into an opportunity; to collect comments and suggestions to get the message across to the NIH/NLM staff that PubMed and other products must be easily accessible to different audiences (to each his own product!), but without losing accuracy, relevance, and quality. This is the challenge we would like to help achieve.

Biomedical librarians and documentalists are a significant help desk in the test of functionality immediately after launch. Associations such as GIDIF-RBM in the field of bibliographic research represent an authoritative reference and an opportunity for mediated and reasoned dialogue between the producer (NLM) and the ultimate user (researcher and public).

The Association will act as a guarantor in the collection of the technical wishes of the bibliographic search operators and, through the comments made in the room, will try to open a working table and active dialogue with the National Library of Medicine.

### **Acknowledgements**

The GIDIF-RBM Board would like to thank Antonio Florita and Alessio Adani of the "Umberto Veronesi" Library of the National Institute of Tumours (Milan) for their exquisite welcome and hospitality, the speakers and in particular Scilla Pizzarelli of the Istituto Su-

periore di Sanità (Rome) and Valeria Scotti of the Library of the San Matteo Hospital (Pavia) for their timely and clear expositions of ever-changing topics.

*Received on 10 November 2023.*  
*Accepted on 28 November 2023.*

## BIBLIOGRAPHIC LINKS

<https://gidif-rbm.org/>  
<https://pubmed.ncbi.nlm.nih.gov/>  
[https://www.nlm.nih.gov/psd/special\\_queries.html](https://www.nlm.nih.gov/psd/special_queries.html)  
<https://www.nlm.nih.gov/mesh/meshhome.html>  
<https://learn.nlm.nih.gov/documentation/training-packets/T0042010P/>  
<https://clinicaltrials.gov/>

