

Intended Users:	Healthcare Providers	Information Professionals	Researchers
-----------------	----------------------	---------------------------	-------------



NLM-Scrubber

Description: Developed at the Lister Hill National Center for Biomedical Communications, the NLM-Scrubber is a freely available tool used for clinical text deidentification. It can be downloaded directly from its website and run locally without any prerequisite installation or internet connection. This tool uses natural language processing to find personally identifying information (PII) and replaces it

with labels in square brackets. For example, “John Smith, diagnosed at age 92,” is replaced with “[PERSONALNAME], diagnosed at age [AGE90+].” People can provide a list of terms to be preserved from deidentification, and a list of terms to be redacted. By default, the NLM-Scrubber deidentifies all Health Insurance Portability and Accountability Act of 1996 (HIPAA)-defined personal identifiers. People have various options to configure the system to preserve some of those personal identifiers (e.g., ages above 89) and produce a limited data set.

Popular uses of this product:

Healthcare Providers	Information Professionals	Researchers
<ul style="list-style-type: none"> Deidentify clinical notes and records so information can be combined, analyzed, and shared. 	<ul style="list-style-type: none"> Guide patrons to current literature on best practices for anonymization and deidentification. Guide patrons to resources on privacy, reidentification risks, and data organization. 	<ul style="list-style-type: none"> Deidentify electronic medical records (EMRs) so information can be combined and analyzed without breaching patient privacy. Deidentify private health information before sharing under an appropriate data use agreement.

Key Points:

- NLM-Scrubber is a freely available, HIPAA-compliant, clinical deidentification tool.
- NLM-Scrubber download is available for Windows and Linux users.

Potential Predicaments:

- The tool requires input folder information where the files of clinical notes are stored and an output folder for the deidentified data, but it can be further customized according to a person’s needs.
- NLM-Scrubber can deidentify 1,000 typical size clinical notes in about a minute on a modern laptop computer.
- People are responsible for verifying that all information is properly deidentified before sharing.

Teaching Examples:

- The NLM-Scrubber User Manual provides examples and step-by-step guidance on running the scrubber with provided examples.
- Have participants explore the annotation information. What are two categories of identifiers? Within each of those categories, what is an example of a deidentified label?

Real Life Examples:

- A researcher at a hospital uses NLM-Scrubber to deidentify clinic notes so information can be safely analyzed while respecting patient privacy.
- A public health researcher uses NLM-Scrubber to deidentify health-related datasets so that research data can be safely preserved, published, and shared.

More Information:

[User Manual](#)

[NLM-Scrubber Announcements](#)

[List of Other Deidentification Tools](#)

Commercial Equivalents:

Numerous commercial and open-source deidentification tools are available. New machine learning and natural language processing algorithms are created every day to perform highly specialized methods of deidentification. Commercial options like [Google Cloud](#), [AirCloak](#), [John Snow Labs](#), and [Skyflow](#), often include other data analytics options with their packages.