

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



## NLM-Scrubber

**Description:** The NLM-Scrubber is a freely available tool used for clinical text de-identification currently available from the Lister Hill National Center for Biomedical Communication, a research arm of the National Library of Medicine. The NLM-Scrubber 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 replace them 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 de-identifies all Health Insurance Portability and Accountability Act of 1996 (HIPAA) defined personal identifiers. People have various options to configure the system in order to preserve some of those personal identifiers (e.g., ages above 89) and produce a limited data set.

### Popular uses for this product:

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

### Key Points:

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

### Considerations:

- The tool requires input folder information where the files of clinical notes are stored and an output folder for the de-identified data, but it can be further customized according to a person’s needs.
- NLM-Scrubber can de-identify 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:

- Have participants explore the annotation information. What are two categories of identifiers? Within each of those categories, what is an example of a de-identified label?
- Discuss the differences between anonymization and de-identification and the differences between sensitive information and personally identifying information. How does NLM-Scrubber protect this kind of information? Brainstorm additional ways to protect private health information.

### Real Life Examples:

- A researcher at a hospital uses NLM-Scrubber to de-identify clinic notes so information can be safely analyzed while respecting patient privacy.
- A public health researcher uses NLM-Scrubber to de-identify 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](#)[An Extensible De-Identification Framework for Privacy Protection of Unstructured Health Information: Creating Sustainable Privacy Infrastructures](#)**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.