# De-Identification Tools

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | [**NLM Scrubber**](https://lhncbc.nlm.nih.gov/scrubber/) | [**CliniDeID**](https://clinacuity.com/clinideid/) | [**Apache OpenNLP**](https://opennlp.apache.org/) | [**MIST (MITRE Identity Scrubber Toolkit)**](https://mist-deid.sourceforge.net/) | [**ARX Data Anonymization Tool**](https://arx.deidentifier.org/) |
| **License** | No – freely available | open-source software, GPL License | [Apache License, Version 2.0](https://www.apache.org/licenses/LICENSE-2.0)  | open source, BSD license | [Apache License, Version 2.0](https://www.apache.org/licenses/LICENSE-2.0) |
| **Supported or funded by?** | National Library of Medicine | [Clinacuity](https://clinacuity.com/), MUSC, U.S. National Institute for General Medical Sciences (NIGMS) (R41GM11647, R42GM116479) | Apache Software Foundation, developed by volunteers | Developed by [MITRE](https://www.mitre.org/our-impact/intellectual-property/mist-mitre-identity-scrubber-toolkit-version-13), no current funding support, seeking government or non-profit partner | Institute of Medical Informatics, Statistics and Epidemiology at the Technical University of Munich (TUM); Berlin Institute of Health (BIH) at Charité |
| **Resources** | [NLM-Scrubber User Manual](https://data.lhncbc.nlm.nih.gov/public/scrubber/files/user_manual/windows/user_manual_v.19.0411W.pdf), some journal publications, announcements | [GitHub](https://github.com/Clinacuity/CliniDeID), 20+ journal publications | [Documentation – Apache OpenNLP](https://opennlp.apache.org/docs/), [FAQ](https://opennlp.apache.org/faq.html)  pageMailing lists: [Get Involved – Apache OpenNLP](https://opennlp.apache.org/get-involved.html) | Several publications (listed under “Where can I get more information?”)Mailing list: [through SourceForge](https://sourceforge.net/projects/mist-deid/lists/mist-deid-users) | [Anonymization tool | ARX](https://arx.deidentifier.org/anonymization-tool/) (User manual)20+ [publications](https://arx.deidentifier.org/publications/) |
| **Which file types?** | plain text, .txt | plain text, SQL | text data | plain text | SQL, CSV, Excel |
| **Training data** | Medical histories, Lab Reports | Discharge Summaries | Hint: check the ManualDepends on which OpenNLP application you are using. There are details about the training models in the documentation manual | Discharges, Lab Reports, Letters, Orders | Demographic datasets from multiple sources\* \*found in [A Flexible Approach to Distributed Data Anonymization](https://doi.org/10.1016/j.jbi.2013.12.002) |
| **Download and Installation** | [NLM-Scrubber Downloads](https://lhncbc.nlm.nih.gov/scrubber/download.html), to download a graphic user interface (GUI), no installation | Script on Github, beta version only, need Java | [Several versions](https://opennlp.apache.org/download.html), latest available is Apache OpenNLP 2.4.0 | [Documentation](https://mist-deid.sourceforge.net/current_docs/html/index.html), including technical requirements (e.g. need Java and Python) | [Downloads](https://arx.deidentifier.org/downloads/), source code in [GitHub](https://github.com/arx-deidentifier/arx); latest available is ARX 3.9.1 |