

US Department of Health and Human Services

Privacy Impact Assessment

Date Signed:

12/09/2025

OPDIV:

NIH

Name:

PubMed and PubMed Central

PIA Unique Identifier:

P-9014561-609268

The subject of this PIA is which of the following?

Minor Application (child)

Identify the Enterprise Performance Lifecycle Phase of the system.

Operations and Maintenance

Is this a FISMA-Reportable system?

No

Does the system include a Website or online application available to and for the use of the general public?

Yes

Identify the operator.

Agency

Is this a new or existing system?

Existing

Does the system have Security Authorization (SA)?

Yes

Indicate the following reason(s) for updating this PIA.

PIA Validation

Significant System Management Change

Describe in further detail any changes to the system that have occurred since the last PIA.

This system was realigned and moved under the Biomedical and Biological Information System (BBIS) General Support System (GSS).

Describe the purpose of the system.

PubMed is a database retrieval system containing over 27 million bibliographic citations both from biomedical literature and non-biomedical life science journals. Some citations include author's name.

PubMed Central® (PMC) is a free full-text archive of biomedical and life sciences journal literature at the NIH National Library of Medicine (NLM). NLM's legislative mandate is to collect and preserve the biomedical literature which includes NLM's extensive print and licensed electronic journal holdings and supports contemporary biomedical research and practice.

PubMed is not an acronym.

Describe the type of information the system will collect, maintain (store), or share.

PubMed can be accessed on a read-only basis by the general public through a web browser to retrieve citations from the PubMed database. Some citations include author's names. The user submits a query that is processed by the PubMed servers, which submit the query to indexes maintained in the PubMed Database. Users are also able to submit batch queries for querying PubMed through the PubMed Utilities servers. The database returns the results of the query to the PubMed portal servers, which format and display the results. The text of the abstract along with Medical Subject Headings (MeSH) indexing terms is retrieved from the Article servers. Related articles and links to data in other NCBI databases are retrieved through the Link servers.

Publishers submitting journal data to PubMed transfer their files to private directories in the PubMed File Transfer Protocol (FTP) site. The PubMed Loader retrieves the Publisher-supplied data from the FTP site and loads to the PubMed Data Management (PMDM) system for processing. A unique PubMed Identification (ID) is assigned to each citation. Once loaded to PMDM system, the data is available for editing by the PubMed data providers and NLM Data Review staff. The MeSH assignments are determined in a separate system, the Indexing Management System (IMS), and exported to PMDM to be added to the citations and included in PubMed.

PubMed Central® (PMC) is a free full-text archive of biomedical and life sciences journal literature at NLM.

Members of the public can voluntarily submit their email address and set up a password to get on the PMC Announce Mail List, Utilities Announce Mail List, and/or Tagging Guideline Mail List request email. Giving a name is optional but the user is required to create a password.

Those requiring access to this system log in using the NIH Identity, Credential, and Access Management (IAM) Services which maintains its own unique privacy impact assessment (PIA) on record, with all legal authorities documented. The purpose of IAM Services is to authenticate and authorize all users and computers in a Windows domain type network assigning and enforcing information security policies for all computers and installing or updating software. The IAM Services collects unique user credentials and stores them in an encrypted format. The IAM Service is an essential service which facilitates and governs network access to various resources.

Provide an overview of the system and describe the information it will collect, maintain (store), or share, either permanently or temporarily.

PubMed is a database retrieval system containing over 27 million bibliographic citations both from biomedical literature and non-biomedical life science journals. Some citations include author's name.

PMC is a free full-text archive of biomedical and life sciences journal literature at NLM. NLM's legislative mandate is to collect and preserve the biomedical literature which includes NLM's extensive print and licensed electronic journal holdings and supports contemporary biomedical research and practice.

PubMed can be accessed on a read-only basis by the general public through a web browser to retrieve citations from the PubMed databases. Some citations include the author's name. The end user accesses PubMed through the NLM load-balanced Web servers. He or she submits a query that is processed by the PubMed servers, which submit the query to indexes maintained in the PubMed Database. Users are also able to submit batch queries using a defined API for querying PubMed through the PubMed Utilities servers. The database returns the results of the query to the PubMed portal servers, which format and display the results. The text of the abstract along with MeSH indexing terms is retrieved from the Article servers. Related articles and links to data in other NCBI databases are retrieved through the Link servers.

Publishers submitting journal data to PubMed transfer their files to private directories in the PubMed File Transfer Protocol (FTP) site. The PubMed Loader retrieves the Publisher-supplied data from the FTP site and loads to the PMDM system for processing. A unique PubMed Identification (ID) is assigned to each citation. Once loaded to PMDM system, the data is available for editing by the PubMed data providers and NLM Data Review staff. The MeSH assignments are determined in a separate system, the IMS, and exported to PMDM to be added to the citations and included in PubMed.

PMC is a free full-text archive of biomedical and life sciences journal literature at the NLM. Members of the public can voluntarily submit their email address and set up a password to get on the PMC Announce Mail List, Utilities Announce Mail List, and/or Tagging Guideline Mail List request email. Giving a name is optional but the user is required to create a password. Once registered, users can also receive the following:

PMC Announce List informs list members of news and announcements related to PMC. Utilities Announce Mail List announcements of new or updated tools and utilities related to PubMed Central. Tagging Guidelines List is the forum that the PubMed Central Conversion Group uses to announce updates to the following resources: Tagging Guidelines, Manuscript Tagging Guidelines, and Style Checker tools.

Does the system collect, maintain, use or share PII?

Yes

Indicate the type of PII that the system will collect or maintain.

Name

E-Mail Address

Member password

Username

Indicate the categories of individuals about whom PII is collected, maintained or shared.

Employees

Public Citizens

How many individuals' PII is in the system?

5,000-9,999

For what primary purpose is the PII used?

The primary purpose of the PII is to allow individuals the ability to sign up for emails and announcements from PMC.

Describe the secondary uses for which the PII will be used.

There is no secondary use of PII.

Identify legal authorities governing information use and disclosure specific to the system and program.

42 U.S.C. section 286, 42 U.S.C. § 282(i) and (j)), 44 U.S.C. Sec. 2904, 42 U.S.C. 241.

Are records on the system retrieved by one or more PII data elements?

Yes

Identify the number and title of the Privacy Act System of Records Notice (SORN) that is being used to cover the system or identify if a SORN is being developed.

09-25-0005, Administration: Library Operations and NIH Library User I.D. File

Identify the sources of PII in the system.

Directly from an individual about whom the information pertains

Online

Government Sources

Identify the OMB information collection approval number and expiration date

None. PubMed and PubMed Central do not solicit information from the public. Submission of an individual's email address is voluntary and not required for use of PubMed and PubMed Central. Its only function is to allow individuals to be updated of changes within the system.

Is the PII shared with other organizations?

No

Describe the process in place to notify individuals that their personal information will be collected. If no prior notice is given, explain the reason.

PII is voluntarily submitted by individuals. A confirmation email is sent once an individual subscribes to verify that the individual wants to be signed to the email lists.

Is the submission of PII by individuals voluntary or mandatory?

Voluntary

Describe the method for individuals to opt-out of the collection or use of their PII. If there is no option to object to the information collection, provide a reason.

Individuals may un-subscribe at any time through the website. Individuals may also reach out to the PMC help desk if they need further assistance.

Process to notify and obtain consent from individuals whose PII is in the system when major changes occur to the system.

PII within the system is not changed or modified. PII is entered by the individual and may be edited at any time by the individual themselves. Individuals may also reach out to the PMC help desk if they have any questions regarding their information or have concerns that need to be resolved.

Describe the process in place to resolve an individual's concerns when they believe their PII has been inappropriately obtained, used, or disclosed, or that the PII is inaccurate.

PII within the system is not changed or modified. PII is entered by the individual and may be edited at any time by the individual themselves. Individuals may also reach out to the PMC help desk if they have any questions regarding their information or have concerns that need to be resolved.

Describe the process in place for periodic reviews of PII contained in the system to ensure the data's integrity, availability, accuracy and relevancy.

Periodic data integrity audits are conducted by PubMed and PubMed Central Administrative Staff.

Identify who will have access to the PII in the system and the reason why they require access.

Describe the procedures in place to determine which system users (administrators, developers, contractors, etc.) may access PII.

Determinations are made based on role-based access controls and least privilege. User rights are provisioned based on controls within the system, allowing users only access to the minimum amount of PII necessary to perform their job.

Access to PII is assigned to personnel based upon current job responsibilities. An IAM account login is required to gain access to the stored PII data.

Describe the methods in place to allow those with access to PII to only access the minimum amount of information necessary to perform their job.

Individuals who sign up for email announcements are verified through a confirmation email. Individuals only have access to their PII.

Identify training and awareness provided to personnel (system owners, managers, operators, contractors and/or program managers) using the system to make them aware of their responsibilities for protecting the information being collected and maintained.

According to NIH policy, all personnel who manage or operate NIH applications must successfully complete annual security and privacy awareness training. Training is completed on the <http://irtsectraining.nih.gov> site with valid NIH credentials.

Describe training system users receive (above and beyond general security and privacy awareness training).

NA

Do contracts include Federal Acquisition Regulation and other appropriate clauses ensuring adherence to privacy provisions and practices?

Yes

Describe the process and guidelines in place with regard to the retention and destruction of PII.

Records are retained and disposed of under the authority of the NIH Records Retention Schedule.

Item 10-101 - Administrative records maintained in any agency office. Administrative records maintained in any agency office. Records accumulated by individual offices that relate to routine day-to-day administration and management of the office rather than the mission-specific activities for which the office exists, excluding records scheduled elsewhere in the General Records Schedule (GRS) such as timekeeping and procurement. Disposition: Destroy when business use ceases. DAA-GRS2016-0016-0001.

Item 07-204 - System access records; Systems requiring special accountability for access; These records are created as part of the user identification and authorization process to gain access to systems. Records are used to monitor inappropriate systems access by users. Disposition: Destroy 6 years after password is altered or user account is terminated, but longer retention is authorized if required for business use. DAA-GRS-2013- 0006-0004.

Item 08-301 - Library Administrative Records.

Records documenting a library's planning and management. Records may document library policies, procedures, and statistics. Destroy when 3 years old or 3 years after superseded or obsolete, whichever is applicable. Longer retention is authorized for business use (DAA-GRS-2015-0003-0001).

Describe, briefly but with specificity, how the PII will be secured in the system using administrative, technical, and physical controls.

Administrative Controls: System administrators are approved by for access based on their technical/functional role in administering, developing, and supporting PubMed and PubMed Central's daily job functions, and PubMed and PubMed Central administrators perform periodic reviews to assure users adhere to system policies.

Technical Controls: Access to the system is controlled by NIH log-in which authenticates the user prior to granting access. Access level and permissions are controlled by the system and based on user, role, organizational unit, and status of the report. All servers have been configured to remove all unused applications and system files and all local account access except when necessary to manage the system and maintain integrity of data.

Physical Controls: The servers reside in the Center for Information Technology (CIT) Computer Room where policies and procedures are in place to restrict access to the machines. This includes guards at the front door and entrance to the machine room.

Identify the publicly-available URL:

PubMed: <https://www.ncbi.nlm.nih.gov/pubmed>

PubMed Central (PMC): <https://www.ncbi.nlm.nih.gov/pmc/>

Note: web address is a hyperlink.

Does the website have a posted privacy notice?

Yes

Is the privacy policy available in a machine-readable format?

Yes

Does the website use web measurement and customization technology?

Yes

Select the type of website measurement and customization technologies is in use and if it is used to collect PII.

Does the website have any information or pages directed at children under the age of thirteen?

No

Does the website contain links to non- federal government websites external to HHS?

Yes

Is a disclaimer notice provided to users that follow external links to websites not owned or operated by HHS?

Yes