

US Department of Health and Human Services

Privacy Impact Assessment

Date Signed:

10/29/2024

OPDIV:

NIH

Name:

NIAID) Public Bioinformatics and Computational Biosciences Branch (BCBB) Developed (PBD)
Authorization Boundary

PIA Unique Identifier:

P-6532809-537337

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?

New

Does the system have Security Authorization (SA)?

Yes

Indicate the following reason(s) for updating this PIA.**Describe the purpose of the system.**

The National Institute of Allergies and Infectious Diseases (NIAID) Public Bioinformatics and Computational Biosciences Branch (BCBB) Developed (PBD) Authorization Boundary boundary consists of public-facing research applications hosted on cloud-based resources owned and managed by the Operations and Engineering Branch (OEB). Applications in the boundary are as follows:

METAGENOTE: An annotation system that allows users to store metadata associated with microbiome sequencing data and facilitates publishing of metadata to the National Center for Biotechnology Information (NCBI) sequence read archive (SRA) public resource.

NIAID Bioinformatics Portal (NBP): Enables researchers across the globe access to NIAID resources. The portal will provide links to NIAID applications, publications, code, news, events and training materials.

Nephele: Microbiome Analysis: Allows researchers to run sophisticated microbiome analysis pipelines on their data and to generate results output quickly and inexpensively using on-demand resources.

NIH 3D: an online repository for searching, browsing, downloading, creating, sharing, and annotating three dimensional (3D) models. These models most often relate to scientific content such as labware, anatomical subjects, personal protective equipment (PPE), educational materials, and biomolecules.

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

The PII referenced applies to the whole boundary. Each element listed may or may not be applicable to all systems in the boundary.

METAGENOTE: Metadata for microbiome analysis. Allows end users to input raw data files and metadata of sample collection and processing. System stores data and creates a submission in a format compatible for the NCBI repository.

NBP: Dashboard for applications, events, training resources, collaborating labs, research collaboration, outreach, and support.

Nephele: Microbiome/Genomic sequencing data uploaded by the user via FastQ file types.

NIH 3D: 3D modeling files, files to be converted to 3D modeling files (such as Digital Imaging and Communications in Medicine (DICOM)), and supplemental documentation and images.

Personally identifiable information (PII) collected, maintained, and/or shared by the PBD authorization boundary includes: Name, email address, phone number, medical notes, date of birth, photographic identifier, mailing address, medical records number, device identifiers, employment status, organization/affiliation, geolocation data, and internet protocol (IP) address.

End users accessing this system use specific login credentials which include unique username and password or through Google authentication (federated).

PBD operations and maintenance personnel login to the systems using the NIH Identity, Credential, and Access Management (IAM) Services which maintains its own unique privacy impact assessment (PIA) on record, including 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 collect unique user credentials and stores them in an encrypted format. The IAM Services are 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.

The National Institute of Allergies and Infectious Diseases (NIAID) Public Bioinformatics and Computational Biosciences Branch (BCBB) Developed (PBD) Authorization Boundary boundary consists of public-facing research applications hosted on cloud-based resources owned and managed by the Operations and Engineering Branch (OEB). Applications in the boundary are as follows:

METAGENOTE: An annotation system that allows users to store metadata associated with microbiome sequencing data and facilitates publishing of metadata to the National Center for Biotechnology Information (NCBI) sequence read archive (SRA) public resource.

NIAID Bioinformatics Portal (NBP): Enables researchers across the globe to access to NIAID resources. The portal will provide links to NIAID applications, publications, code, news, events and training materials.

Nephele: Microbiome Analysis: Allows researchers to run sophisticated microbiome analysis pipelines on their data and to generate results output quickly and inexpensively using on-demand resources.

NIH 3D: an online repository for searching, browsing, downloading, creating, sharing, and annotating three dimensional (3D) models. These models most often relate to scientific content such as labware, anatomical subjects, personal protective equipment (PPE), educational materials, and biomolecules.

METAGENOTE: Metadata for microbiome analysis. Allows end users to input raw data files and metadata of sample collection and processing. System stores data and creates a submission in a format compatible for the NCBI repository.

NBP: Dashboard for applications, events, training resources, collaborating labs, research collaboration, outreach, and support.

Nephele: Microbiome/Genomic sequencing data uploaded by the user via FastQ file types.

NIH 3D: 3D modeling files, files to be converted to 3D modeling files (such as Digital Imaging and Communications in Medicine (DICOM)), and supplemental documentation and images.

PII collected, maintained, and/or shared includes: Name, email address, phone number, medical notes, date of birth, photographic identifier, mailing address, medical records number, device identifiers, employment status, organization/affiliation, geolocation data, and internet protocol (IP) address.

End users accessing this system use specific login credentials which include unique username and password or through Google authentication (federated).

PBD operations and maintenance personnel login to the systems using the NIH Identity, Credential, and Access Management (IAM) Services which maintains its own unique privacy impact assessment (PIA) on record, including 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 collect unique user credentials and stores them in an encrypted format. The IAM Services are an essential service which facilitates and governs network access to various resources.

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

Yes

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

Date of Birth

Name

Photographic Identifiers

E-Mail Address

Mailing Address

Phone Numbers

Medical Records Number

Medical Notes

Device Identifiers

Organization/affiliation

IP address

Geolocation

Password

Username

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

Employees
Public Citizens
Business Partner/Contacts (Federal/state/local agencies)

How many individuals' PII is in the system?

5,000-9,999

For what primary purpose is the PII used?

User account creation

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

All PBD applications facilitate end user support request submission capabilities.

NBP: IP address is captured in application logs.

NIH 3D:

1. PII may be collected as being inherent to the uploaded DICOM files, but it is de-identified for the output and the source file is purged.
2. Registered users may opt to include their PII as part of their public username and biographical profile. This allows them to be associated to, or promote, the contributions they have made.
3. IP address is captured in application logs.
4. Registered users may upload supplemental files that support their 3D model contributions. The user is free to include images in the files, which may contain the geolocation of the image's origin.
5. The NIH 3D team manages registered user Application Programming Interface (API) key requests.

Nephele: Providing notification to the end user that the test results for their research projects are available.

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

42 U.S.C. 241, Research and Investigation; 42 U.S.C. 285f, National Institute on Allergy and Infectious Diseases

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 use to cover the system or identify if a SORN is being developed.

09-25-0200, Clinical, Basic and Population-based Research Studies of the National Institutes of Health

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

Non-Governmental Sources are exempt from an OMB Information Collection Number through Public Law 114-255 - 21st Century Cures Act, Section 2035: Exemption for the National Institutes of Health from the Paperwork Reduction Act requirements.

Is the PII shared with other organizations?

Yes

Identify with whom the PII is shared or disclosed and for what purpose.

Describe any agreements in place that authorizes the information sharing or disclosure.

NIH 3D: Registered users acknowledge the privacy policies at account creation and again when submitting data for processing. The policy includes verbiage stating that providing the PII is voluntary.

Describe the procedures for accounting for disclosures.

NBP: Internal NIH personnel end user support requests are submitted in the application. These generate a service ticket to an NIAID managed instance of Jira, which maintains its own privacy impact assessment (PIA), that includes the end user's information.

NIH 3D: Registered user API key requests are initiated from the application and processed through the NIH M365 portal tools Microsoft Forms and Lists, which maintains its own PIA. The end user provides their name, email address, NIH 3D username, and description of requested usage via MS Forms. Integrated workflows in the Lists application capture the request information for management purposes.

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

NIH 3D: End users are provided the option to share their 3D model contributions. If so, they must select a license to associate to the record. If no license is applicable, they must select "Public Domain" as the appropriate option. The user acknowledges the Terms and Conditions at account creation, which includes requirements for uploading content free of PII. DICOM files submitted by end users may contain PII for the image subject, but those fields are deidentified, not included in the converted output, and purged. Users are free to upload content relevant to the 3D modeling community. Users acknowledge the Terms and Conditions at account creation.

NBP: For support requests, the user (NIH users only) is prompted to provide the required information in order to submit the request.

Entry Submissions, references user responsibilities when uploading content, including notice that the user should not submit content with PII. The medical information, DoB, etc, is content potentially available on DICOM files uploaded by the end user to the NIH 3D application. All fields that may contain PII are de-identified by default as part of the image conversion to a 3D model file.

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.

There are no methods for users to opt-out of the collection or use of PII at account creation. Users may contact the application administrators to address the removal of their accounts as needed.

There is no method for the end user to opt out of the collection or use of their PII for reasons beyond the account related items. The end user chooses to upload / share their 3D modeling files. Once uploaded and shared, the content is considered public domain unless specifically licensed. Any PII that may be on image uploads are de-identified, then purged once the file is converted as part of the workflow.

NIH 3D: There are no methods for users to opt-out of the collection of PII for image file uploads. The end user acknowledges the entry submission requirements on the Terms and Conditions, which references the restriction of uploading content with PII. Any end user DICOM file types undergo de-identification by default.

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

Any major system/usage changes will still result in email notifications to the end users.

METAGENOTE and NBP: No notice is required as public end user PII is not retained.

Nephele and NIH 3D: Users are contacted via email for any major changes.

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.

Users may contact the site administrators via email or internal messaging mechanisms to resolve issues, including those involving PII. Administrators coordinate with the user to address their concerns.

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

End users provide the PII voluntarily. They are responsible for the accuracy of any self-asserted PII. PBD boundary applications are hosted on the OEB's Monarch platform (cloud-based). All components, including encrypted components containing PII, are reviewed periodically as part of system development lifecycle operations and maintenance practices to ensure integrity and availability of the PII. This includes daily vulnerability and compliance scans and platform controls that require system configurations to be reviewed and/or updated every thirty days.

Nephele and NIH 3D: At account creation, the user is sent an email with a link or a verification code that they must use or provide to complete the process. Users validate the accuracy or relevancy of retained email addresses when undergoing the password recovery processes.

NIH 3D: All medical imaging entries are reviewed for relevancy prior to being made public on the site.

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.

Role-based access controls are implemented to ensure only authorized PBD operations and maintenance personnel have access to information that has not been made public by the user.

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.

Principle of least privilege is applied to the administrator roles to ensure access to PII is applicable to the level required for the role.

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 awareness training. There are five categories of mandatory IT training (Information Security, Counterintelligence, Privacy Awareness, Records Management and Emergency Preparedness). 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).

Administrators and Privileged Users require additional training specific to their roles and responsibilities. This includes Application Security and Best Practices training available from the NIH Information Security and Information Management Training.

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.

01-003, Records of All Other Intramural Research Projects (DAA-0443-2012-0007-0003): Cut off annually at termination of project/program or when no longer needed for scientific reference, whichever is longer. Destroy 7 years after cutoff.

07-203, System access records. Systems not requiring special accountability for access (DAA-GRS-2013-0006-0001): Email addresses are retained for access control purposes. The records are destroyed when business use ceases.

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

Administrative Controls: All technical personnel who access IT systems which contain protected information have met background investigation criteria for Public Trust positions. All personnel have taken mandatory security and privacy training classes and annual refreshers. Administrative personnel accessing these systems use privileged and separate accounts for system administrative access.

NIH 3D: Facial images uploaded by end users are screened by content administrators to ensure they are appropriate prior to posting the output 3D model.

Technical Controls: Access to system components by PBD operations and maintenance staff is controlled by the NIH login, 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 resources 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. Administrators must be logged in to the NIH virtual private network (VPN). PII is encrypted at rest and in transit.

NIH 3D: DICOM file types uploaded by the end user are uploaded to a temporary encrypted directory. The file is immediately de-identified. Supplemental files that contain recognized medical imaging formats are also anonymized by removing their header information. Once de-identified, the data is converted to a 3D modeling file type. The source DICOM file is deleted. If this imaging happens to contain the face, it is only used to generate the 3D model, and never made public. It is stored in an encrypted directory in case workflows need to be rerun. The data is encrypted at rest and the application programming interface (API) to access it requires a key.

Physical controls: PBD applications are hosted on virtual servers on the Monarch platform. The

physical servers supporting Monarch reside in the Amazon Web Services (AWS) Cloud where policies and procedures are in place to restrict access to the machines.

Identify the publicly-available URL:

<https://metagenote.niaid.nih.gov/>

<https://bioinformatics.niaid.nih.gov/>

<https://nephele.niaid.nih.gov/>

<https://3d.nih.gov>

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.

Other technologies that do not collect PII:

Google Analytics

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?

No