

06.1 HHS Privacy Impact Assessment (Form) / NIH NLM dbGaP (Database of Genotype and Phenotype) (Item)

Primavera
ProSight

Form Report, printed by: Peltekian, Jacqueline, **Oct 23, 2012**

PIA SUMMARY

1	<p>The following required questions with an asterisk (*) represent the information necessary to complete the PIA Summary for transmission to the Office of Management and Budget (OMB) and public posting in accordance with OMB Memorandum (M) 03-22.</p> <p>Note: If a question or its response is not applicable, please answer "N/A" to that question where possible. If the system hosts a website, the Website Hosting Practices section is required to be completed regardless of the presence of personally identifiable information (PII). If no PII is contained in the system, please answer questions in the PIA Summary Tab and then promote the PIA to the Senior Official for Privacy who will authorize the PIA. If this system contains PII, all remaining questions on the PIA Form Tabs must be completed prior to signature and promotion.</p>
----------	--

2	Summary of PIA Required Questions				
*Is this a new PIA?					
No					
If this is an existing PIA, please provide a reason for revision:					
PIA Validation					
*1. Date of this Submission:					
Aug 14, 2012					
*2. OPDIV Name:					
NIH					
*4. Privacy Act System of Records Notice (SORN) Number (If response to Q.21 is Yes, a SORN number is required for Q.4):					
No					
*5. OMB Information Collection Approval Number:					
No					
*6. Other Identifying Number(s):					
No					
*7. System Name (Align with system item name):					
dbGaP - Database of Genotype and Phenotype					
*9. System Point of Contact (POC). The System POC is the person to whom questions about the system and the responses to this PIA may be addressed:					
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th colspan="2" style="text-align: left;">Point of Contact Information</th> </tr> <tr> <td style="width: 50%;">POC Name</td> <td>Dar-Ning Kung</td> </tr> </table>		Point of Contact Information		POC Name	Dar-Ning Kung
Point of Contact Information					
POC Name	Dar-Ning Kung				
*10. Provide an overview of the system:					
<p>dbGaP, the database of Genotype and Phenotype, is a database designed to archive and distribute data from genome wide association (GWA) studies. GWA studies explore the association between specific genes (genotype information) and observable traits, such as blood pressure and weight, or the presence or absence of a disease or condition (phenotype information). Connecting phenotype and genotype data provides information about the genes that may be involved in a disease process or condition, which can be critical for better understanding the disease and for developing new diagnostic methods and treatments.</p> <p>The database does not contain names, social security numbers, fingerprints, photographs or anything enabling facial recognition. The data is strictly de-identified patient data and does not fall under the category of IIF.</p>					
*13. Indicate if the system is new or an existing one being modified:					
Existing					
*17. Does/Will the system collect, maintain (store), disseminate and/or pass through PII within any database(s), record(s), file(s) or website(s) hosted by this					

system?
<i>TIP: If the answer to Question 17 is "No" (indicating the system does not contain PII), only the remaining PIA Summary tab questions need to be completed and submitted. If the system does contain PII, the full PIA must be completed and submitted. (Although note that "Employee systems," – i.e., systems that collect PII "permitting the physical or online contacting of a specific individual ... employed [by] the Federal Government – only need to complete the PIA Summary tab.)</i>
No
<i>17a. Is this a GSS PIA included for C&A purposes only, with no ownership of underlying application data? If the response to Q.17a is Yes, the response to Q.17 should be No and only the PIA Summary must be completed.</i>
<i>*19. Are records on the system retrieved by 1 or more PII data elements?</i>
No
<i>*21. Is the system subject to the Privacy Act? (If the response to Q.19 is Yes, the response to Q.21 must be Yes and a SORN number is required for Q.4)</i>
No
<i>*23. If the system shares or discloses PII, please specify with whom and for what purpose(s):</i>
No
<i>*30. Please describe in detail: (1) The information the agency will collect, maintain, or disseminate (clearly state if the information contained in the system ONLY represents federal contact data); (2) Why and for what purpose the agency will use the information; (3) Explicitly indicate whether the information contains PII; and (4) Whether submission of personal information is voluntary or mandatory:</i>
Data collected includes the researchers name and institutional information, a research abstract (reason for requesting the data) and co-investigator information. This is collected for further contact with the PI and to provide controlled access to the data requested and to provide public access to the research uses of the data. The information collected from co-investigators is the same as that from Principal Investigators: name, business address, and email address. The submission of personal information is voluntary. The dbGaP database contains phenotype and genotype data from researchers and from centers who are conducting genome-wide association studies. NLM/NCBI summarizes, reformats, and redistributes these data acting as a central repository for these types of studies. The information collected is from studies sponsored by an NIH Institute and is sent from the principal investigator or the center conducting the study. All data received is certified as de-identified data. After NIH review of a request from an investigator and his/her sponsoring organization, the genotype and phenotype data is made available for that investigator to access. Data are categorized by an accession number assigned by NLM/NCBI (not the investigator) to the dataset. Information is retrieved by the name of the study. The capability exists to search the public data for the name of the study, the protocols used, and the dataset summaries but the retrieval is by accession number. No information in dbGaP is collected directly from patients. Data has not been collected from other NIH databases. If data were to be provided from other NIH databases, e.g., an intramural study, it would be provided under the same conditions as external data, i.e., all data would be de-identified. There are no names or personal identifiers linked to the phenotype/genotype records. All data are de-identified prior to the time it is delivered to NLM/NCBI.
<i>*31. Please describe in detail any processes in place to: (1) Notify and obtain consent from the individuals whose PII is in the system when major changes occur to the system (e.g., disclosure and/or data uses have changed since the notice at the time of the original collection); (2) Notify and obtain consent from individuals regarding what PII is being collected from them; and (3) How the information will be used or shared. (Note: Please describe in what format individuals will be given notice of consent [e.g., written notice, electronic notice, etc.]):</i>
No
<i>*32. Does the system host a website? (Note: If the system hosts a website, the Website Hosting Practices section is required to be completed regardless of the presence of PII)</i>
Yes
<i>*37. Does the website have any information or pages directed at children under the age of thirteen?</i>
No
<i>*50. Are there policies or guidelines in place with regard to the retention and destruction of PII? (Refer to the C&A package and/or the Records Retention and Destruction section in SORN)</i>
No
<i>*54. Briefly describe in detail how the PII will be secured on the system using administrative, technical, and physical controls:</i>
No

PIA REQUIRED INFORMATION

1 HHS Privacy Impact Assessment (PIA)

The PIA determines if Personally Identifiable Information (PII) is contained within a system, what kind of PII, what is done with that information, and how that information is protected. Systems with PII are subject to an extensive list of requirements based on privacy laws, regulations, and guidance. The HHS Privacy Act Officer may be contacted for issues related to Freedom of Information Act (FOIA) and the Privacy Act. Respective Operating Division (OPDIV) Privacy Contacts may be contacted for issues related to the Privacy Act. The Office of the Chief Information Officer (OCIO) can be used as a resource for questions related to the administrative, technical, and physical controls of the system. Please note that answers to questions with an asterisk (*) will be submitted to the Office of Management and Budget (OMB) and made publicly available in accordance with OMB Memorandum (M) 03-22.

Note: If a question or its response is not applicable, please answer "N/A" to that question where possible.

2 General Information

*Is this a new PIA?

No

If this is an existing PIA, please provide a reason for revision:

PIA Validation

*1. Date of this Submission:

Aug 14, 2012

*2. OPDIV Name:

NIH

3. Unique Project Identifier (UPI) Number for current fiscal year (Data is auto-populated from the System Inventory form, UPI table):

*4. Privacy Act System of Records Notice (SORN) Number (If response to Q.21 is Yes, a SORN number is required for Q.4):

No

*5. OMB Information Collection Approval Number:

No

5a. OMB Collection Approval Number Expiration Date:

*6. Other Identifying Number(s):

No

*7. System Name: (Align with system item name)

dbGaP - Database of Genotype and Phenotype

8. System Location: (OPDIV or contractor office building, room, city, and state)

System Location:	
OPDIV or contractor office building	
Room	
City	
State	

*9. System Point of Contact (POC). The System POC is the person to whom questions about the system and the responses to this PIA may be addressed:

Point of Contact Information	
POC Name	Dar-Ning Kung

The following information will not be made publicly available:

POC Title	
POC Organization	
POC Phone	
POC Email	

*10. Provide an overview of the system: (Note: The System Inventory form can provide additional information for child dependencies if the system is a GSS)

dbGaP, the database of Genotype and Phenotype, is a database designed to archive and distribute data from genome wide association (GWA) studies. GWA studies explore the association between specific genes (genotype information) and observable traits, such as blood pressure and weight, or the presence or absence of a disease or condition (phenotype information). Connecting phenotype and genotype data provides information about the genes that may be involved in a disease process or condition, which can be critical for better understanding the disease and for developing new diagnostic methods and treatments.

The database does not contain names, social security numbers, fingerprints, photographs or anything enabling facial recognition. The data is strictly de-identified patient data and does not fall under the category of IIF.

SYSTEM CHARACTERIZATION AND DATA CATEGORIZATION

1 System Characterization and Data Configuration

11. Does HHS own the system?

11a. If no, identify the system owner:

12. Does HHS operate the system? (If the system is operated at a contractor site, the answer should be No)

12a. If no, identify the system operator:

*13. Indicate if the system is new or an existing one being modified:

Existing

14. Identify the life-cycle phase of this system:

15. Have any of the following major changes occurred to the system since the PIA was last submitted?

Please indicate "Yes" or "No" for each category below:	Yes/No
Conversions	
Anonymous to Non-Anonymous	
Significant System Management Changes	
Significant Merging	
New Public Access	
Commercial Sources	
New Interagency Uses	
Internal Flow or Collection	
Alteration in Character of Data	

16. Is the system a General Support System (GSS), Major Application (MA), Minor Application (child) or Minor Application (stand-alone)?

Minor Application (child)

*17. Does/Will the system collect, maintain (store), disseminate and/or pass through PII within any database(s), record(s), file(s) or website(s) hosted by this system?

No

TIP: If the answer to Question 17 is "No" (indicating the system does not contain PII), only the remaining PIA Summary tab questions need to be completed and submitted. If the system does contain PII, the full PIA must be completed and submitted. (Although note that "Employee systems," – i.e., systems that collect PII "permitting the physical or online contacting of a specific individual ... employed [by] the Federal Government – only need to complete the PIA Summary tab.)

Please indicate "Yes" or "No" for each PII category. If the applicable PII category is not listed, please use the Other field to identify the appropriate category of PII.

Categories:	Yes/No
Name (for purposes other than contacting federal employees)	
Date of Birth	

Social Security Number (SSN)	
Photographic Identifiers	
Driver's License	
Biometric Identifiers	
Mother's Maiden Name	
Vehicle Identifiers	
Personal Mailing Address	
Personal Phone Numbers	
Medical Records Numbers	
Medical Notes	
Financial Account Information	
Certificates	
Legal Documents	
Device Identifiers	
Web Uniform Resource Locator(s) (URL)	
Personal Email Address	
Education Records	
Military Status	
Employment Status	
Foreign Activities	
Other	

17a. Is this a GSS PIA included for C&A purposes only, with no ownership of underlying application data? If the response to Q.17a is Yes, the response to Q.17 should be No and only the PIA Summary must be completed.

18. Please indicate the categories of individuals about whom PII is collected, maintained, disseminated and/or passed through. Note: If the applicable PII category is not listed, please use the Other field to identify the appropriate category of PII. Please answer "Yes" or "No" to each of these choices (NA in other is not applicable).

Categories:	Yes/No
Employees	
Public Citizen	
Patients	
Business partners/contacts (Federal, state, local agencies)	
Vendors/Suppliers/Contractors	
Other	

*19. Are records on the system retrieved by 1 or more PII data elements?

No

Please indicate "Yes" or "No" for each PII category. If the applicable PII category is not listed, please use the Other field to identify the appropriate category of PII.

Categories:	Yes/No
--------------------	---------------

Name (for purposes other than contacting federal employees)	
Date of Birth	
SSN	
Photographic Identifiers	
Driver's License	
Biometric Identifiers	
Mother's Maiden Name	
Vehicle Identifiers	
Personal Mailing Address	
Personal Phone Numbers	
Medical Records Numbers	
Medical Notes	
Financial Account Information	
Certificates	
Legal Documents	
Device Identifiers	
Web URLs	
Personal Email Address	
Education Records	
Military Status	
Employment Status	
Foreign Activities	
Other	

20. Are 10 or more records containing PII maintained, stored or transmitted/passed through this system?

*21. Is the system subject to the Privacy Act? (If the response to Q.19 is Yes, the response to Q.21 must be Yes and a SORN number is required for Q.4)

No

21a. If yes but a SORN has not been created, please provide an explanation.

INFORMATION SHARING PRACTICES

1 Information Sharing Practices

22. Does the system share or disclose PII with other divisions within this agency, external agencies, or other people or organizations outside the agency?

Please indicate "Yes" or "No" for each category below:	Yes/No
Name (for purposes other than contacting federal employees)	
Date of Birth	
SSN	
Photographic Identifiers	
Driver's License	
Biometric Identifiers	
Mother's Maiden Name	
Vehicle Identifiers	
Personal Mailing Address	
Personal Phone Numbers	
Medical Records Numbers	
Medical Notes	
Financial Account Information	
Certificates	
Legal Documents	
Device Identifiers	
Web URLs	
Personal Email Address	
Education Records	
Military Status	
Employment Status	
Foreign Activities	
Other	

*23. If the system shares or discloses PII please specify with whom and for what purpose(s):

No

24. If the PII in the system is matched against PII in one or more other computer systems, are computer data matching agreement(s) in place?

25. Is there a process in place to notify organizations or systems that are dependent upon the PII contained in this system when major changes occur (i.e., revisions to PII, or when the system is replaced)?

26. Are individuals notified how their PII is going to be used?

26a. If yes, please describe the process for allowing individuals to have a choice. If no, please provide an explanation.

27. Is there a complaint process in place for individuals who believe their PII has been inappropriately obtained, used, or disclosed, or that the PII is inaccurate?

27a. If yes, please describe briefly the notification process. If no, please provide an explanation.

28. Are there processes in place for periodic reviews of PII contained in the system to ensure the data's integrity, availability, accuracy and relevancy?

28a. If yes, please describe briefly the review process. If no, please provide an explanation.

29. Are there rules of conduct in place for access to PII on the system?

Please indicate "Yes," "No," or "N/A" for each category. If yes, briefly state the purpose for each user to have access:

Users with access to PII	Yes/No/N/A	Purpose
User		
Administrators		
Developers		
Contractors		
Other		

*30. Please describe in detail: (1) The information the agency will collect, maintain, or disseminate (clearly state if the information contained in the system ONLY represents federal contact data); (2) Why and for what purpose the agency will use the information; (3) Explicitly indicate whether the information contains PII; and (4) Whether submission of personal information is voluntary or mandatory:

Data collected includes the researchers name and institutional information, a research abstract (reason for requesting the data) and co-investigator information. This is collected for further contact with the PI and to provide controlled access to the data requested and to provide public access to the research uses of the data.

The information collected from co-investigators is the same as that from Principal Investigators: name, business address, and email address. The submission of personal information is voluntary.

The dbGaP database contains phenotype and genotype data from researchers and from centers who are conducting genome-wide association studies. NLM/NCBI summarizes, reformats, and redistributes these data acting as a central repository for these types of studies.

The information collected is from studies sponsored by an NIH Institute and is sent from the principal investigator or the center conducting the study. All data received is certified as de-identified data. After NIH review of a request from an investigator and his/her sponsoring organization, the genotype and phenotype data is made available for that investigator to access.

Data are categorized by an accession number assigned by NLM/NCBI (not the investigator) to the dataset. Information is retrieved by the name of the study. The capability exists to search the public data for the name of the study, the protocols used, and the dataset summaries but the retrieval is by accession number.

No information in dbGaP is collected directly from patients. Data has not been collected from other NIH databases. If data were to be provided from other NIH databases, e.g., an intramural study, it would be provided under the same conditions as external data, i.e., all data would be de-identified.

There are no names or personal identifiers linked to the phenotype/genotype records. All data are de-identified prior to the time it is delivered to NLM/NCBI.

*31. Please describe in detail any processes in place to: (1) Notify and obtain consent from the individuals whose PII is in the system when major changes occur to the system (e.g., disclosure and/or data uses have changed since the notice at the time of the original collection); (2) Notify and obtain consent from individuals regarding what PII is being collected from them; and (3) How the information will be used or shared. (Note: Please describe in what format individuals will be given notice of consent [e.g., written notice, electronic notice, etc.]

No

WEBSITE HOSTING PRACTICES

1 Website Hosting Practices

*32. Does the system host a website? (Note: If the system hosts a website, the Website Hosting Practices section is required to be completed regardless of the presence of PII)

Yes

Please indicate "Yes" or "No" for each type of site below. If the system hosts both Internet and Intranet sites, indicate "Yes" for "Both" only.	Yes/ No	If the system hosts an Internet site, please enter the site URL. Do not enter any URL(s) for Intranet sites.
Internet	Yes	http://www.ncbi.nlm.nih.gov/entrez/db=gap
Intranet		
Both		

33. Does the system host a website that is accessible by the public and does not meet the exceptions listed in OMB M-03-22?

Note: OMB M-03-22 Attachment A, Section III, Subsection C requires agencies to post a privacy policy for websites that are accessible to the public, but provides three exceptions: (1) Websites containing information other than "government information" as defined in OMB Circular A-130; (2) Agency intranet websites that are accessible only by authorized government users (employees, contractors, consultants, fellows, grantees); and (3) National security systems defined at 40 U.S.C. 11103 as exempt from the definition of information technology (see section 202(i) of the E-Government Act.).

Yes

34. If the website does not meet one or more of the exceptions described in Q. 33 (i.e., response to Q. 33 is "Yes"), a website privacy policy statement (consistent with OMB M-03-22 and Title II and III of the E-Government Act) is required. Has a website privacy policy been posted?

Yes

35. If a website privacy policy is required (i.e., response to Q. 34 is "Yes"), is the privacy policy in machine-readable format, such as Platform for Privacy Preferences (P3P)?

Yes

35a. If no, please indicate when the website will be P3P compliant:

36. Does the website employ tracking technologies?

Yes

Please indicate "Yes", "No", or "N/A" for each type of cookie below:	Yes/No/N/A
Web Bugs	No
Web Beacons	No
Session Cookies	Yes
Persistent Cookies	No
Other	

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

No

37a. If yes, is there a unique privacy policy for the site, and does the unique privacy policy address the process for obtaining parental consent if any information is collected?

38. Does the website collect PII from individuals?

No

Please indicate "Yes" or "No" for each category below:	Yes/No
Name (for purposes other than contacting federal employees)	
Date of Birth	
SSN	
Photographic Identifiers	
Driver's License	
Biometric Identifiers	
Mother's Maiden Name	
Vehicle Identifiers	
Personal Mailing Address	
Personal Phone Numbers	
Medical Records Numbers	
Medical Notes	
Financial Account Information	
Certificates	
Legal Documents	
Device Identifiers	
Web URLs	
Personal Email Address	
Education Records	
Military Status	
Employment Status	
Foreign Activities	
Other	

39. Are rules of conduct in place for access to PII on the website?

Not Applicable

40. Does the website contain links to sites external to HHS that owns and/or operates the system?

Yes

40a. If yes, note whether the system provides a disclaimer notice for users that follow external links to websites not owned or operated by HHS.

Yes

ADMINISTRATIVE CONTROLS

1

Administrative Controls

Note: This PIA uses the terms "Administrative," "Technical" and "Physical" to refer to security control questions—terms that are used in several Federal laws when referencing security requirements.

41. Has the system been certified and accredited (C&A)?

41a. If yes, please indicate when the C&A was completed:

41b. If a system requires a C&A and no C&A was completed, is a C&A in progress?

42. Is there a system security plan for this system?

43. Is there a contingency (or backup) plan for the system?

44. Are files backed up regularly?

45. Are backup files stored offsite?

46. Are there user manuals for the system?

47. Have personnel (system owners, managers, operators, contractors and/or program managers) using the system been trained and made aware of their responsibilities for protecting the information being collected and maintained?

48. If contractors operate or use the system, do the contracts include clauses ensuring adherence to privacy provisions and practices?

49. Are methods in place to ensure least privilege (i.e., "need to know" and accountability)?

49a. If yes, please specify method(s):

*50. Are there policies or guidelines in place with regard to the retention and destruction of PII? (Refer to the C&A package and/or the Records Retention and Destruction section in SORN):

No

50a. If yes, please provide some detail about these policies/practices:

TECHNICAL CONTROLS

1 Technical Controls

51. Are technical controls in place to minimize the possibility of unauthorized access, use, or dissemination of the data in the system?

Please indicate "Yes" or "No" for each category below:	Yes/No
User Identification	
Passwords	
Firewall	
Virtual Private Network (VPN)	
Encryption	
Intrusion Detection System (IDS)	
Common Access Cards (CAC)	
Smart Cards	
Biometrics	
Public Key Infrastructure (PKI)	

52. Is there a process in place to monitor and respond to privacy and/or security incidents?

52a. If yes, please briefly describe the process:

PHYSICAL ACCESS

1 Physical Access

53. Are physical access controls in place?

Please indicate "Yes" or "No" for each category below:	Yes/No
Guards	
Identification Badges	
Key Cards	
Cipher Locks	
Biometrics	
Closed Circuit TV (CCTV)	

*54. Briefly describe in detail how the PII will be secured on the system using administrative, technical, and physical controls:

No

APPROVAL/DEMOTION

1 System Information

System Name:	dbGaP - Database of Genotype and Phenotype
---------------------	--

2 PIA Reviewer Approval/Promotion or Demotion
--

Promotion/Demotion:	Promote
----------------------------	---------

Comments:	
------------------	--

Approval/Demotion Point of Contact:	Dar-Ning Kung
--	---------------

Date:	Aug 14, 2012
--------------	--------------

3 Senior Official for Privacy Approval/Promotion or Demotion

Promotion/Demotion:	Promote
----------------------------	---------

Comments:	
------------------	--

4 OPDIV Senior Official for Privacy or Designee Approval

Please print the PIA and obtain the endorsement of the reviewing official below. Once the signature has been collected, retain a hard copy for the OPDIV's records. Submitting the PIA will indicate the reviewing official has endorsed it

This PIA has been reviewed and endorsed by the OPDIV Senior Official for Privacy or Designee (Name and Date):

Name: _____ Date: _____

Name:	Karen Plá
Date:	Sep 28, 2012

5 Department Approval to Publish to the Web
--

Approved for web publishing	Yes
------------------------------------	-----

Date Published:	Sep 1, 2009
------------------------	-------------

Publicly posted PIA URL or no PIA URL explanation:	http://www.hhs.gov/pia/nih.html
---	---

PIA % COMPLETE

1	PIA Completion	
	PIA Percentage Complete:	100.00
	PIA Missing Fields:	