



Privacy Impact Assessment  
for the  
**Genomic Data Network and Analysis  
(GDNA)**

**DHS/S&T/PIA-039**

**February 25, 2020**

**Contact Point**

**Julie Brewer**

**Director**

**Office of National Laboratories  
Science & Technology Directorate  
(202) 254-8227**

**Reviewing Official**

**Jonathan R. Cantor**

**Acting Chief Privacy Officer  
Department of Homeland Security  
(202) 343-1717**



## Abstract

The National Biodefense Analysis and Countermeasures Center (NBACC) is a federally funded research and development center (FFRDC) laboratory sponsored by the U.S. Department of Homeland Security's (DHS) Science & Technology Directorate (S&T). NBACC's National Bioforensic Analysis Center (NBFAC) created the Genomic Data Network and Analysis (GDNA) system to analyze bioforensic data for unclassified law enforcement casework and forensic capability development projects. NBACC uses the GDNA system to analyze the deoxyribonucleic acid (DNA) sequences of bacteria, viruses, plants, and non-human animals. S&T is conducting this Privacy Impact Assessment (PIA) because GDNA also will process human DNA for casework, as part of ongoing human forensics capability development projects, and for quality assurance purposes. This PIA identifies and assesses the potential privacy risks posed by the collection, use, handling, storage, retention, and disposal of human DNA by NBFAC and the GDNA system and the mitigations necessary to address the identified privacy risks.

## Overview

### Background

#### *National Biodefense Analysis and Countermeasures Center (NBACC)*

NBACC is a one-of-a-kind facility with the capability to address critical biodefense-related shortcomings in the nation's scientific knowledge of biological agents (e.g., bacteria, viruses, toxins) that may pose a threat of harm, as well as the nation's ability to respond to such threats. Battelle National Biodefense Institute, LLC (BNBI) operates and manages NBACC as an FFRDC for S&T.<sup>1</sup> NBACC bears the same responsibility as S&T for ensuring that the programs and systems NBACC manages and operates, on S&T's behalf, provide adequate privacy protections.

NBACC is the premier national security biocontainment laboratory that responds to and provides data required to support prevention, protection, mitigation, and recovery from current and future bio-crime and bioterrorism threats. NBACC's work focuses on: (1) scientific analyses to support bio-crime and bioterror attribution investigations (i.e., analyses to support the identification or source of a biological threat for known, emerging, enhanced, genetically engineered, and synthetically derived biological agents); (2) timely scientific biological threat characterization analyses (e.g., analyzing the nature of the threat a biological agent poses and the needed or existing capabilities to respond to the threat); and (3) a continuously available national

---

<sup>1</sup> FFRDCs are public-private partnerships that conduct research for the U.S. Government. These are typically administered by universities and corporations in accordance with 48 CFR 35.017. An FFRDC meets some special long-term research or development need that cannot be met as effectively by existing in-house or contractor resources. FFRDCs enable agencies to use private sector resources to accomplish tasks that are integral to the mission and operation of the sponsoring agency.



capability of broad and adaptable competencies that provide both pre-positioned information and responsive scientific experiments for newly identified threats.<sup>2</sup> These threat analysis capabilities support intelligence assessments, preparedness planning, and response to both events and identified potential threats.

### *National Bioforensic Analysis Center (NBFAC)*

In accordance with the Homeland Security Act of 2002 and Homeland Security Presidential Directive-10 (HSPD-10), NBACC operates NBFAC as the lead federal technical forensic analysis facility responsible for analyzing materials recovered following a biological attack.<sup>3</sup> NBFAC uses state-of-the-art, agent-specific assays and next-generation sequencing<sup>4</sup> and bioinformatic<sup>5</sup> capabilities to identify and characterize biological threat agents in a highly contained and regulated laboratory environment. NBFAC provides this forensic analysis in support of whichever federal agency the President designates as the lead agency to conduct the biological attack investigation.

The Federal Bureau of Investigation (FBI) has lead responsibility for investigating acts of bio-crime and bioterror. DHS entered into a Memorandum of Agreement (MOA) with the FBI through which DHS and the FBI collaborate on NBACC's operation and management. S&T, as the NBACC FFRDC sponsor, provides governance and program direction. Under the MOA, the FBI provides a portion of the funding for NBFAC and oversees and sets NBFAC's requirements.

Aside from its role in the event of a large-scale biological attack, NBFAC analyzes evidence samples from bio-crime or bioterrorism investigations and develops advanced bioforensic capabilities. As a result, NBFAC performs two major capacities as both an operational (casework analysis) and a research (capability development) laboratory. In the first capacity, NBFAC serves as the operational analysis laboratory that identifies information about biological materials in evidentiary samples delivered by the FBI. In the second capacity, NBFAC develops new methodologies and capabilities to expand and improve operational analysis. NBFAC research and development (R&D) efforts are focused on a number of areas, including Genomics<sup>6</sup> and Bioinformatics, based on biothreat planning analysis.

### *Deoxyribonucleic Acid (DNA)*

DNA is the molecule that stores the blueprint information essential to an organism's survival.

---

<sup>2</sup> See The NBACC Strategic Plan, available at <http://bnbi.org/about-us-2/strategic-plan/>.

<sup>3</sup> See Public Law 107-296 Sections 101(b), 302-303; 6 U.S.C. § 111(b)(1) and §§ 182-183, and Homeland Security Presidential Directive/HSPD-10: *Biodefense for the 21<sup>st</sup> Century*, available at <https://www.gpo.gov/fdsys/pkg/CPRT-110HPRT39618/pdf/CPRT-110HPRT39618.pdf>.

<sup>4</sup> Next generation sequencing (NGS, NextGenSeq) is a relatively new method for sequencing genomes at high speed and at low cost. It is also known as second generation sequencing (SGS) or massively parallel sequencing (MPS).

<sup>5</sup> Bioinformatics is the science of collecting and analyzing complex biological data such as genetic codes.

<sup>6</sup> Genomics is the branch of molecular biology concerned with the structure, function, evolution, and mapping of genomes.



DNA stores long linear sequences of four different subcomponents (nucleotides) that form a code of information translated into messages used by all living organisms for life functions.<sup>7</sup> The DNA in a casework sample could originate from any organism, including bacteria, viruses, plants, animals, and humans.

Human DNA collection, use, and maintenance presents particular privacy concerns because each individual's DNA sequence is unique and DNA samples can never be made truly anonymized.<sup>8</sup> Although DNA is inherently identifiable, additional information is needed in order to attribute the DNA to a specific individual. NBFAC refers to anonymization of DNA as using a non-attributable identifier (i.e., one that cannot be linked to an individual in any way, without a key that would allow attribution to an individual).

### The GDNA System

NBFAC owns, operates, and uses the GDNA system as a key tool to meet the mission need for DNA-based identification. GDNA is an information technology system that performs advanced analysis on data collected for both operational and research activities. GDNA is a stand-alone, air-gapped system that resides only within the physical NBACC boundary. GDNA is not connected to any external networks and can only be accessed by a limited number of authorized users.

While not networked externally, GDNA, itself, is a network of laboratory instruments and high-performance computing (HPC) hardware that enables DNA sequencing, storage of DNA sequence data, and DNA sequence data analysis. The HPC components of GDNA include a large shared file system, a system for backing up data, nodes that are used for algorithmic processing, nodes with large amounts of memory that are used for various bioinformatic tasks, and end-user workstations that provide users with an interface to the system.

GDNA uses its DNA sequencers, which are automated devices that can "read" raw DNA material (e.g., casework biological samples), to derive and transcribe the nucleotide sequences present on DNA molecules into machine-readable sequences that have been extracted from the biological material. A "DNA sequence" is a digital representation of a single strand of a DNA molecule. In some cases, GDNA receives DNA information provided to NBFAC already sequenced. Once sequenced, GDNA performs the computational analysis necessary to convert the raw DNA sequence information into high-quality genomic sequences that allow for the identification of particular features and characteristics.

GDNA obtains, processes, and stores DNA for casework, capability development, and quality assurance from various sources. NBFAC and GDNA use data collected from volunteers

---

<sup>7</sup> A DNA sequence is a digital representation of a single strand of a DNA molecule, using the symbols A (for adenine), C (for cytosine), G (for guanine), and T (for thymine).

<sup>8</sup> See [www.genomics.gov](http://www.genomics.gov), *Privacy in Genomics*, National Human Genome Research Institute, National Institutes of Health.



and through law enforcement activities that includes personally identifiable information (PII), such as biological samples, DNA sequences, and, in the case of the volunteers, other identifying information including name, ethnicity, hair color, and eye color. Other DNA sources include publicly available datasets, information from collaborators, such as universities or other laboratories, or commercial sources, as well as information from laboratory staff and visitors. These data sources are described in more detail below.

### *The GDNA System’s DNA Sources*

In many instances, the DNA samples and sequences are coded with a unique identifier not associated with any PII to minimize the ability to link a sample or sequence to an individual. The extent to which NBFAC’s DNA samples and sequences are traceable to an individual differs depending on their source and purpose. In some instances, NBFAC works with coded samples or DNA sequences provided by a partner. The partner may or may not maintain identifying information that can be retrieved using the code. In either event, NBFAC does not possess the code and cannot connect the DNA to an individual. In other instances, NBFAC collects DNA from known individuals and codes the DNA for internal use. NBFAC limits, through physical and technological controls, access to the key that would enable the DNA to be traced to the individual from whom it was collected.

Table 1, below, provides an overview of the sources from which NBFAC obtains DNA, the individuals from whom the DNA is collected, how the DNA is used (i.e., for casework, capability development, and/or quality assurance), and additional information about NBFAC’s retrieval of the PII by unique identifier and its ability to link the DNA information to an individual or infer an individual’s identity from the DNA.

*Table 1. Summary of information about DNA Sequences stored or processed by the GDNA System*

<b>DNA Source for NBFAC</b>	<b>Individuals whose DNA NBFAC collects, processes, and/or maintains</b>	<b>DNA Use</b>	<b>PII: Retrieval, Linkage, and Inference</b>
Law enforcement	Individuals whose DNA may be present in a law enforcement-related casework sample or in a reference sample	Casework	NBFAC retrieves this PII by FBI-provided case file number and/or NBFAC-assigned ID number  NBFAC does not have the information needed to link the PII to an individual or to infer an individual’s identity



Publicly available information	Individuals unknown to NBFAC	Casework Capability Development Quality Assurance	PII is anonymized and/or aggregated NBFAC does not retrieve this PII by unique identifier NBFAC does not have the information needed to link the PII to an individual or to infer an individual's identity
Collaborators or commercial entities	Individuals unknown to NBFAC; may or may not be known by the source	Capability Development Quality Assurance	PII is anonymized NBFAC does not have the information needed to link the PII to an individual or to infer an individual's identity
Volunteers	Individuals known to NBFAC	Capability Development Quality Assurance	PII will be retrieved by a unique identifier PII is linkable to an individual, allowing an individual's identity to be inferred
NBFAC laboratory staff and visitors	Individuals known to NBFAC	Quality Assurance	PII will be retrieved by a unique identifier PII is linkable to an individual, allowing an individual's identity to be inferred

## Casework Forensic Analysis

NBFAC sequences and analyzes a wide variety of biological samples in the performance of FBI casework following established scientific processes. NBFAC determines whether DNA is present in the sample, and, if so, identifies what type of DNA it is (e.g., plant or animal) and particular characteristics the DNA may possess (e.g., the DNA is synthetic or genetically altered) by comparing the sequences to known reference databases. The majority of casework samples consist predominantly of microbial, plant, or animal (non-human) DNA.

To facilitate casework analysis, NBFAC downloads DNA reference sequences (plant, animal, microbial, human) from publicly available National Institutes of Health (NIH) databases and makes these sequences available on GDNA for sequence matching purposes. These NIH databases are collectively referred to as GenBank.<sup>9</sup> Human DNA sequence data in these public databases often does not represent a specific individual, but rather an “average” of the sequence

<sup>9</sup> GenBank databases are maintained by the National Center for Biotechnology Information, which is part of the United States National Library of Medicine, a branch of NIH. A submitter of data to GenBank is allowed to submit data derived from sequencing human DNA. GenBank documentation states the following: “If you are submitting human sequences to GenBank, do not include any data that could reveal the personal identity of the source. GenBank assumes that the submitter has received any necessary informed consent authorizations required prior to submitting sequences.” See <https://www.ncbi.nlm.nih.gov/genbank/>.



across multiple people. To the extent the GenBank human DNA sequence does represent a specific individual, GDNA does not and cannot collect or store any identifying information about individuals from GenBank DNA sequences. NIH does not store identifying information associated with the human sequences in GenBank. By comparing DNA sequences derived from NBFAC evidentiary samples to the GenBank sequences, NBFAC analysts ascertain the sequence types present in the evidence samples (e.g., the sample contains material from a specific type of plant or animal, or the sample includes human DNA).

When NBFAC determines that human DNA is present in an evidence sample, NBFAC may perform additional testing to differentiate between individuals or generate information (such as physical characteristics) useful for investigative leads. In these instances, the FBI may provide NBFAC with one or more reference human DNA samples for sequencing and comparison to the evidence sample. These reference human DNA samples come from individuals known to the FBI (or other federal agency submitting the casework) but will be anonymized for provision to NBFAC. Thus, the individual's identity will not be known or knowable at NBFAC.

NBFAC uses the reference sequences to make a comparison between an evidence sample human DNA sequence and reference sample sequence to determine if the comparison produces a match. If so, statistical calculations may be performed to determine the confidence level of the match (e.g., is it a perfect match or less than perfect match). The calculations produce a probability that the matching sequences could occur by chance and are based on how common or rare the DNA sequence is in a relevant human population group (e.g., Caucasians). To perform the calculations, NBFAC must use additional information about the match frequency rates within particular human population groups. NBFAC obtains the additional information in the form of sequence frequency datasets developed for public use by both federal agencies and academic labs.<sup>10</sup> NBFAC may download human DNA sequence frequency datasets from a public source for storage and use on GDNA for statistical calculations. These frequency datasets are by definition aggregated across a human population group sample and thus contain no identifying information about individuals.

Information about the sequences present in the casework sample and any statistical calculations are compiled into a textual report that is delivered to law enforcement for investigative lead generation and attribution.

---

<sup>10</sup> For example, see the sequence frequency datasets for four U.S. population groups developed by the National Institute of Standards and Technology (NIST), available at <https://catalog.data.gov/dataset/sequence-based-u-s-population-data-for-27-autosomal-str-loci>.



## **Human Forensic Analysis Capability Development**

### *Current Development Efforts*

NBFAC is developing human forensic analysis capabilities. To develop and validate these capabilities, NBFAC works primarily with anonymized human DNA sequences, obtained from various sources. For some applications, NBFAC additionally requires DNA sequences from known individuals that could serve as positive controls for the purpose of validating results.

The anonymized human DNA sequences used for capability development come from publicly available data, data obtained through collaboration with federal agencies and academic labs, and samples purchased from commercial entities, as described below:

- Publicly available data includes, for example, sequences that can be downloaded from GenBank or have been made available in a scientific publication.
- Anonymized human sequences provided by collaborators, such as the FBI, other federal agencies, or academic labs, directly to NBFAC for use on GDNA, or developed by NBFAC from anonymized human samples provided by those entities. For all such anonymized sequences, NBFAC cannot link a sample to a specific individual, as NBFAC does not have access to any data that would indicate from which individuals the samples were collected. The collaborating agency or lab providing the samples is responsible for ensuring that the sequences were collected in accordance with an approved human use protocol and with appropriate privacy protections in place.
- NBFAC purchases DNA materials from commercial entities developed specifically for validation use in a scientific setting. For all such anonymized sequences, NBFAC cannot link a sample to a specific individual. NBFAC does not have access to any data that would indicate from which individuals the samples were collected.

### *Positive Controls*

Positive controls in human DNA sequencing are samples for which the correct “answer” is already known. NBFAC uses DNA sequencing from known and/or unknown individuals as positive controls, depending on what NBFAC is testing. For example, positive controls from unknown individuals (such as a cell line purchased from a commercial source, or a positive control DNA sample provided with a commercial testing kit) may be used in casework to confirm that an established laboratory or analysis process has worked correctly. In capability development, NBFAC uses DNA sequences from both unknown and known individuals as positive controls to test whether a new method will produce accurate results.

Samples NBFAC collects from volunteers have been the primary source of the known samples used as positive controls for capability development. For these samples, NBFAC collects



PII such as name, ethnicity, and hair and eye color from volunteers. The known information associated with the samples, such as hair and eye color, is used to test whether the sequencing and subsequent sequence analysis methods can determine the correct hair and eye color from the DNA sample. For the capability development work NBFAC has performed to date, the DNA sequence information generated from these known samples represents less than 0.0001% of the human genome, the minimum amount necessary to conduct appropriate validation work.<sup>11</sup>

### Quality Assurance

#### *Training and Proficiency Testing*

GDNA will analyze human DNA sequences as part of internal training and proficiency testing. Depending on the aim of the training or proficiency testing (i.e., laboratory methods or analysis methods), the human DNA sequences processed on GDNA for these purposes may come from publicly available data, collaborators, commercial entities, or volunteers.

#### *Staff Elimination database*

Elimination databases are routinely developed and maintained by human DNA testing laboratories, as well as by commercial laboratories that make consumables and reagents for human DNA testing, and are recommended for forensic practice by authoritative bodies such as the Scientific Working Group on DNA Analysis Methods (SWGDM).<sup>12</sup> Elimination databases serve to ensure that human DNA sequences identified and reported in casework are not the result of contamination by a person who has been in contact with or in proximity to the casework evidence.

To ensure the quality of NBFAC's casework involving human DNA, NBFAC plans to develop an internal elimination database on GDNA. Elimination databases contain DNA sequences of individuals who may be in a position to inadvertently contaminate a sample. For NBFAC's casework, these individuals are the NBACC personnel who access the laboratory suites where human DNA casework may be performed or who handle casework items prior to laboratory processing, and visitors who enter the laboratory suites where human DNA casework may be performed.

The staff elimination database will be developed by collecting a DNA sample (cheek swab) from NBACC personnel and visitors who may be in position to contaminate a casework sample. Only a DNA sample, the individual's name, sex, and self-described race/ethnicity/ancestry will be collected. At the time of collection, the individual will be provided with an informed consent form specifying details regarding:

---

<sup>11</sup> The human genome is estimated to comprise over 3 billion base-paired nucleotides organized into 22 paired chromosomes, plus the X chromosome (one in males, two in females) and the Y chromosome (one in males).

<sup>12</sup> For more information, see the SWGDAM webpage at <https://www.swgdam.org/> and the SWGDAM Contamination Prevention and Detection Guidelines for Forensic DNA Laboratories at [https://media.wix.com/ugd/4344b0\\_c4d4dbba84f1400a98eaa2e48f2bf291.pdf](https://media.wix.com/ugd/4344b0_c4d4dbba84f1400a98eaa2e48f2bf291.pdf).



- the purpose of the collection;
- disposition of the sample and any data developed from the sample;
- confidentiality and privacy protections; and
- foreseeable risks.

Individuals will have an opportunity to decline to provide a sample. However, individuals who decline to provide a sample may not be permitted entry to the laboratory spaces in which casework that may contain human DNA is processed.

Two cheek swab samples will be collected from each individual. At the time of collection, the cheek swab samples will be coded with a unique identifier. NBFAC will extract DNA from cellular materials in one of the swabs, and the swab will be discarded. The DNA extracts and the second (unextracted) swabs will be stored in a locked room within the NBFAC Genomics laboratories. Only individuals enrolled in NBACC's Personal Reliability Program and who have received Human Subjects Protections training will be permitted unescorted access to the room. Sequence data developed from each sample will be stored anonymously on GDNA in a stand-alone elimination database using the same unique identifier. Access to the elimination database will be limited to a small number of NBFAC staff involved in human DNA casework analysis, and the elimination database will be accessed only when contamination is suspected. A lookup document (i.e., key) connecting the unique identifiers to specific individuals, as well as the signed informed consent forms (which necessarily include an individual's name) will be maintained on GDNA, but will be stored on a separate GDNA computer, password-protected, and accessible by only two NBFAC staff members.

### **Future Forensic Capability Development, Collaboration Activities, or Changes to Quality Assurance Practices**

NBFAC continually develops new methodologies and capabilities to expand and improve operational analysis for GDNA, including capabilities based on human forensics project planning and FBI requirements. NBFAC research and development efforts are focused in a number of areas, including Genomics and Bioinformatics using biothreat planning analysis. NBFAC will use GDNA to perform human DNA analysis to develop and validate these new forensic capabilities.

NBFAC will submit Privacy Threshold Analyses (PTA) to determine formally whether a new capability development project, collaboration activity, or change in quality assurance practice involving PII or potentially privacy-sensitive technology requires an update to this PIA, other further privacy risk analysis and management, or other required privacy compliance documentation such as a System of Records Notice (SORN).



## Section 1.0 Authorities and Other Requirements

### 1.1 What specific legal authorities and/or agreements permit and define the collection of information by the project in question?

The Homeland Security Act of 2002, National Security Presidential Memorandum 14, *Support for National Biodefense*, and the National Biodefense Strategy serve as the primary authorities for NBACC and NBFAC bioforensic work.<sup>13</sup> DHS designated NBFAC to be the lead federal facility to conduct and facilitate the technical forensic analysis and interpretation of materials recovered following a biological attack. NBFAC conducts bioforensic analysis of evidence for the FBI and develops advanced bioforensic capabilities under these authorities.

DHS and the FBI entered into a Memorandum of Agreement (MOA) that sets terms and conditions that govern NBFAC's work on the FBI's behalf. The MOA also addresses the FBI's responsibility for setting the law enforcement and evidentiary standards for the collection and analysis of evidence in connection with FBI casework samples.

Additional authorities that require or permit DHS and the FBI to collect and analyze data relevant to their Chemical, Biological, Radiological, and Nuclear (CBRN) forensic and attribution responsibilities include:

- Homeland Security Act of 2002, Section 302;<sup>14</sup>
- National Defense Authorization Act for Fiscal Year 2017, Division A, Title X, § 1086;<sup>15</sup>
- 18 U.S.C. Chapter 11B, *Chemical Weapons*;<sup>16</sup>
- 28 U.S.C. § 533, *Investigative and Other Officials; Appointment*;<sup>17</sup>
- 31 U.S.C. § 1535, *The Economy Act, Agency Agreements*;<sup>18</sup>
- 28 CFR 0.85, *Organization of the Department of Justice – General Functions of the*

---

<sup>13</sup> Available at <https://www.whitehouse.gov/presidential-actions/presidential-memorandum-support-national-biodefense/>.

<sup>14</sup> Available at [https://www.dhs.gov/sites/default/files/publications/hr\\_5005\\_enr.pdf](https://www.dhs.gov/sites/default/files/publications/hr_5005_enr.pdf).

<sup>15</sup> Available at <https://uscode.house.gov/statviewer.htm?volume=130&page=2423>.

<sup>16</sup> Available at <https://www.gpo.gov/fdsys/granule/USCODE-2011-title18/USCODE-2011-title18-partI-chap11B/content-detail.html>.

<sup>17</sup> Available at <https://www.gpo.gov/fdsys/granule/USCODE-2011-title28/USCODE-2011-title28-partII-chap33-sec533>.

<sup>18</sup> Available at <https://www.gpo.gov/fdsys/granule/USCODE-2009-title31/USCODE-2009-title31-subtitleII-chap15-subchapIII-sec1535/content-detail.html>.



*Federal Bureau of Investigation;*<sup>19</sup>

- National Security Presidential Directive 17 (NSPD-17) / Homeland Security Presidential Directive 4 (HSPD-4), *National Strategy to Combat Weapons of Mass Destruction;*<sup>20</sup>
- National Security Presidential Memorandum 14, *Support for National Biodefense;*
- National Biodefense Strategy;
- Homeland Security Presidential Directive-5 (HSPD-5), *Management of Domestic Incidents;*<sup>21</sup> and
- Presidential Decision Directive 39, *U.S. Policy on Counterterrorism.*<sup>22</sup>

## **1.2 What Privacy Act System of Records Notice(s) (SORN(s)) apply to the information?**

DHS/Science & Technology Directorate-001 DHS Research, Development, Test, and Evaluation Records System of Records provides notice of PII collected and maintained for capability development.<sup>23</sup>

DHS is developing a new SORN to provide notice for PII collected and maintained to use staff elimination database records, a future database. NBFAC will not collect information for this database until a SORN is published.

DHS/ALL-004 General Information Technology Access Account Records System (GITAARS) provides notice of PII collected and maintained to ensure appropriate access control for DHS information resources and technology systems.<sup>24</sup>

The DNA information in the FBI-provided case samples, processed DNA sequences, and the resulting reports for the NBFAC casework analysis does not require a SORN. Although DNA is inherently identifiable, sequence information and other records in GDNA have no associated identifying information that makes DNA linkable to any specific individual. Further, DHS will not use DNA stored in GDNA to match against any other information it holds in any other DHS system that would make this information linkable to a specific individual.

---

<sup>19</sup> Available at <https://www.gpo.gov/fdsys/granule/CFR-2010-title28-vol1/CFR-2010-title28-vol1-sec0-85/content-detail.html>.

<sup>20</sup> Available at <https://www.gpo.gov/fdsys/pkg/CPRT-110HPRT39618/pdf/CPRT-110HPRT39618.pdf>.

<sup>21</sup> Available at <https://www.dhs.gov/sites/default/files/publications/Homeland%20Security%20Presidential%20Directive%205.pdf>.

<sup>22</sup> Available at <https://www.hsdl.org/?view&did=462942>.

<sup>23</sup> DHS/S&T-001 Research, Development, Test, and Evaluation Records, 78 FR 3019 (January 15, 2013).

<sup>24</sup> DHS/ALL-004 General Information Technology Access Account Records System (GITAARS), 77 FR 70792 (November 27, 2012).



### **1.3 Has a system security plan been completed for the information system(s) supporting the project?**

NBFAC completed a System Security Plan (SSP) and the Authority to Operate (ATO) for GDNA was granted on June 1, 2016. The GDNA system was accepted into the DHS Ongoing Authorization (OA) Program on August 20, 2018, and is currently in OA status. NBFAC and DHS are currently in the process of updating the SSP to change the GDNA availability impact-level from “Low” to “Moderate” in accordance with the role GDNA plays in supporting federal law enforcement. OA requires GDNA to be reviewed by the S&T Chief Information Officer on a regular basis and maintain its security and privacy posture to maintain its ATO.

### **1.4 Does a records retention schedule approved by the National Archives and Records Administration (NARA) exist?**

NBACC retains GDNA system records in accordance with General Records Schedule 3.1, General Technology Management Records.

Records from volunteers that contain PII are being proposed for coverage by a National Archives and Records Administration (NARA)-approved retention schedule. Until GDNA has a NARA-approved records retention schedule, NBACC is not legally permitted to dispose of these records.

NBACC has four requests for Records Disposition Authority pending approval for the casework and capability development scientific records, including the human DNA. The following paragraphs outline the requested schedules for the scientific records. Table 1 sets forth the applicable records retention schedules that have been requested through the Request for Records Disposition Authority.

*Table 2. Relevant records schedules*

<b>Records Schedule No.</b>	<b>Retention</b>	<b>Reason for Retention of Documentation</b>
DAA-0563-2019-0005-0025	20 years	Records generated for use in Law Enforcement Cases, with potential for appeals
DAA-0563-2019-0005-0024	20 years	Research and Development Files or Projects, not used in Law Enforcement Cases, allows time to evaluate historical significance
DAA-0563-2019-0005-	Permanent	Records used in significant law enforcement cases or projects involving novel or complex issues, public interest, media



0032		attention or congressional scrutiny.
DAA-0563-2019-0005-0036	5 years after superseded or obsolete	Records that document compliance with ISO 17025 requirements to carry out tests and/or calibrations, including sampling. Accreditation demonstrates technical competence and the ability to produce precise and accurate test and/or calibration data.

GDNA maintains records downloaded from publicly available sources in accordance with GRS 5.2, Item 20: Intermediary Records. Retention is temporary, and the data files can be destroyed when no longer needed for a business use.

**1.5 If the information is covered by the Paperwork Reduction Act (PRA), provide the OMB Control number and the agency number for the collection. If there are multiple forms, include a list in an appendix.**

NBACC is working with S&T to develop the appropriate PRA documentation.

**Section 2.0 Characterization of the Information**

**2.1 Identify the information the project collects, uses, disseminates, or maintains.**

As Table 1 indicates, GDNA processes biological DNA sequences from five sources: (1) biological information from FBI, or other agency, evidentiary and reference casework samples; (2) downloaded DNA sequences from publicly available sources; (3) non-human and fully anonymized human DNA samples or sequences received from collaborators such as federal agencies or academic labs, or purchased from commercial entities; (4) samples submitted by volunteers; and (5) samples from NBFAC-associated staff or visitors.

*Table 3. Summary of information NBFAC uses based on type of activity*

<b>NBFAC Activity</b>	<b>Information NBFAC collects, uses, disseminates, or maintains</b>
<b>Casework Analysis</b>	<ul style="list-style-type: none"> <li>• Casework sample materials that may contain human DNA</li> <li>• DNA sequences</li> </ul>



	<ul style="list-style-type: none"><li>• Population frequencies</li><li>• Textual analytical reports</li></ul>
<b>Human Forensic Capability Development</b>	<ul style="list-style-type: none"><li>• DNA sequences</li><li>• Human biological samples</li><li>• Name</li><li>• Self-identified race/ethnicity/ancestry</li><li>• Sex</li><li>• Hair color</li><li>• Eye color</li></ul>
<b>Staff Elimination Database</b>	<ul style="list-style-type: none"><li>• Human biological samples</li><li>• Name</li><li>• Self-identified race/ethnicity/ancestry</li><li>• Sex</li></ul>
<b>Training and Proficiency Testing</b>	<ul style="list-style-type: none"><li>• DNA sequences</li></ul>

The GDNA system stores anonymized DNA sequences from volunteers for the purposes of (1) developing and validating human forensic analysis protocols and (2) internal training and proficiency testing. These sequences are (1) provided in an anonymized form to NBFAC by a collaborator or commercial entity; (2) are derived from anonymized human samples provided to NBFAC by a collaborator or commercial entity; or (3) are derived from samples collected from volunteers.

The information collected from volunteers for capability development projects includes some identifying information. For instance, the information that was collected for a current Institutional Review Board (IRB)-approved capability development project was (1) volunteer name; (2) self-reported information about race/ethnicity/ancestry and phenotypic characteristics (such as hair and eye color); and (3) a buccal (cheek) swab. NBFAC extracted DNA from cellular materials in the swab and stored the materials in a locked room within the NBFAC Genomics laboratories. The DNA sequence data collected from these samples as part of the IRB-approved project included a very small amount of information—less than 0.0001% of the human genome—and the specific parts of the genome sequenced in these analyses have no known uses beyond differentiating between individuals and prediction of ancestry, eye color, and hair color.



For the elimination database, information will be collected from individuals accessing specific NBFAC casework laboratories or who may handle NBFAC casework materials containing human DNA. The information that will be collected is name, sex, self-described race/ethnicity/ancestry, and two buccal swabs. NBFAC will extract DNA from cellular materials in one of the swabs and will discard the used swab. The DNA extract from the first swab and the second (unextracted) swab will be stored in a locked room within the NBFAC Genomics laboratories. The DNA sequence data collected from these samples will be based on the capabilities that have been developed by NBFAC for casework sample sequencing. That is, the specific parts of the genome that may be sequenced in NBFAC's casework will also be sequenced for the elimination database samples.

## 2.2 What are the sources of the information and how is the information collected for the project?

### *Casework*

- DNA sequences associated with FBI casework may be received directly from the FBI or may be generated through analysis of an FBI-provided sample by a GDNA sequencing instrument.
- Publicly available data used as reference information for FBI casework is downloaded from sources such as the National Center for Biotechnology Information (NCBI), which is part of the United States National Library of Medicine (NLM), a branch of the NIH,<sup>25</sup> and may be obtained from human DNA studies, such as those published in the journal *Forensic Science International: Genetics*<sup>26</sup> or made publicly available by federal entities such as the National Institute of Standards and Technology (NIST)<sup>27</sup> and the FBI.<sup>28</sup>

### *Capability Development*

- DNA samples include volunteers. DNA has been collected by buccal swab and DNA sequences have been generated. In the future, volunteers may be sought to provide samples that can be anonymized and used for internal training and proficiency testing purposes.
- Publicly available data that are used as reference information such as GenBank data

---

<sup>25</sup> More information is available at <https://www.ncbi.nlm.nih.gov/guide/dna-rna/>.

<sup>26</sup> More information is available at <https://www.fsigenetics.com/>.

<sup>27</sup> For an example of the information made publicly available by NIST that may be used by NBFAC, see <https://catalog.data.gov/dataset/sequence-based-u-s-population-data-for-27-autosomal-str-loci>.

<sup>28</sup> For an example of the information made publicly available by the FBI that may be used by NBFAC, see <https://ucr.fbi.gov/lab/biometric-analysis/codis/expanded-fbi-str-2015-final-6-16-15.pdf>.



downloaded from NCBI or from published studies.

- Anonymized human sequences either received from collaborators such as the FBI (or other federal agencies) or academic labs or purchased from commercial entities.
- Anonymized human sequences from commercial entities are typically widely available standards or cell lines that labs may use for testing or as controls.

### *Quality Assurance*

- Information from staff or visitors who have access to the laboratory for the staff elimination database.
- Information from staff who handle casework material prior to laboratory processing for the staff elimination database.
- DNA sequences from publicly available or commercial sources described under Capability Development for training and proficiency testing.

In the future, it is also possible that capability development work may extend to the collection and analysis of human DNA from other sample types from volunteers. Despite the fact that the same DNA is found in all body tissues, certain tissues are less prone to contamination from microbes or other sources, and thus may have advantages over buccal swabs. Additionally, capabilities may need to be extended to address specific sample types that are commonly encountered in human forensics casework, such as hair, fingernails, or skin cells. This work is still in the planning stages, however, and would only commence following regulatory and privacy review and approval within NBACC, S&T, and DHS.

### **2.3 Does the project use information from commercial sources or publicly available data? If so, explain why and how this information is used.**

Yes. GDNA uses publicly available reference DNA sequence databases, commonly referred to as “GenBank.” NBFAC downloads the GenBank data from the NIH website and transfers the data to GDNA using encrypted media. NBFAC uses the GenBank reference data to characterize unknown DNA sequences when performing bioforensic casework.

GDNA may also store and use population-level frequency data for human DNA sequences that come from published studies or other public sources, such as the short tandem repeat (STR) marker frequency data made available by the FBI Laboratory.<sup>29</sup> In human DNA casework, if a comparison between an evidence sample sequence and reference sample sequence produces a match, statistical calculations may be performed to determine the importance of the match. The

---

<sup>29</sup> See <https://ucr.fbi.gov/lab/biometric-analysis/codis/expanded-fbi-str-2015-final-6-16-15.pdf>.



calculations produce a probability that the matching sequences could occur by chance, and are based on how common or rare the DNA sequence is in a relevant human population group (e.g., Caucasians). The calculations thus require the use of sequence frequency datasets for human population groups, which are developed for public use by both federal agencies and academic labs.

Human DNA sequences obtained from published studies or commercial entities that cannot be ascribed to an individual may also be used on GDNA for capability development. For instance, human genome reference data from GenBank may be used to help design, develop, and test laboratory methods for human DNA typing or sequence data analysis methods. Anonymized human samples, standards, and cell lines, some of which come from commercial sources or whose DNA sequences are publicly available, may also be used for capability development. For example, NBFAC uses human DNA standards developed by the NIST,<sup>30</sup> or control DNA samples included with the purchase of commercial human genotyping assays, as positive controls to assess laboratory method or analysis protocol performance.

The same types of anonymized samples or sequences from commercial and public sources that are used for capability development may be used to train NBFAC staff in validated laboratory processing or data analysis methods, or to implement a proficiency testing program for validated methods.

## **2.4 Discuss how accuracy of the data is ensured.**

NBACC ensures that data received is appropriately encoded for DNA sequences. NIH uses technical controls (i.e., MD5 checksum files) to ensure the data is downloaded without errors.<sup>31</sup> Human subject matter experts review the accuracy of matches when developing the text for casework and other external reports.

Several bioinformatics workflows are ISO 17025-accredited, which entails measuring analysis accuracy using various datasets as controls. Accuracy is periodically confirmed through the performance of internal and external Proficiency Tests (PTs), and re-accreditation of bioinformatics workflows is performed as required by ISO 17025 standards.<sup>32</sup>

---

<sup>30</sup> See <https://www.nist.gov/programs-projects/dna-profiling-standard-reference-materials> for information on the development and usage of NIST Standard Reference Materials for forensic human DNA profiling. Genotype information for the 2391b Standard Reference Material can be found at [https://www-s.nist.gov/srmors/view\\_datafiles.cfm?srm=2391d](https://www-s.nist.gov/srmors/view_datafiles.cfm?srm=2391d).

<sup>31</sup> MD5 (Message Digest algorithm 5) is an algorithmic hashing function that ensures that a file has not been changed as a result of a faulty file transfer, disk error, or other actions.

<sup>32</sup> Additional information about ISO 17025 (general requirements for the competence of testing and calibration laboratories) is available at <https://www.iso.org/ISO-IEC-17025-testing-and-calibration-laboratories.html>.



## 2.5 Privacy Impact Analysis: Related to Characterization of the Information

**Privacy Risk**: There is a risk that an individual's identity may be determined from the DNA sequences.

**Mitigation**: This risk is extremely low for individuals from whom samples have been collected by NBFAC due to the intense security measures in place that restrict access to the lookup document (key). NBFAC does not possess or have access to a key for anonymized human samples or anonymized human DNA sequences provided by external partners or obtained from public or commercial sources. Determining identity without a lookup document would require access to a forensic database and would also require the individual to have a profile in such a database. NBFAC does not have access to law enforcement databases.

**Privacy Risk**: There is a risk that anonymized data will be de-anonymized.

**Mitigation**: This is extremely unlikely because it would require the external partner that submitted the data to also inadvertently share information that would enable the data to be de-anonymized. They would not do so willfully, and their own security measures should ensure that such an event would not happen accidentally. If NBFAC were to receive information not properly anonymized, upon discovery, NBFAC would take steps to return or destroy the information without further use or processing of it.

## Section 3.0 Uses of the Information

### 3.1 Describe how and why the project uses the information.

In the performance of bioforensic casework, GDNA is used to compare unknown DNA sequences to reference databases of known DNA sequences to determine the most likely biological source of the sequences (animal, plant, bacterium, virus, etc.). GDNA then outputs a report that lists scores and sequence alignment details when matching sequences are found.<sup>33</sup> These metrics include length of match, percent similarity, locations of matching regions on the unknown sequence and known sequence, probability that the match occurred by random chance, and a description of the known sequence. A subject matter expert analyst reviews the match information for each unknown sequence and develops a human-readable report summarizing the analysis.

As previously stated, anonymized human DNA samples and sequences are used for capability development, and DNA sequences have been collected from volunteers for the purpose of developing and validating human forensic analysis protocols. At present, the human volunteer

---

<sup>33</sup> When a statistically significant match is found between two sequences, the length of the match in nucleotides, the percent similarity (number of matching nucleotides divided by length of match expressed as a percentage), and the location of the matching sub-regions within the original sequences are output for a subject matter expert to review.



sequences are only used to develop and validate capabilities. In the future, volunteers may be sought to provide samples that can be used for internal training and proficiency testing as part of a quality assurance program. For either purpose, however, human volunteer samples are not used in any way during the performance of casework (e.g., human DNA that may be present in casework samples is *not* being matched against human volunteer sequences).

DNA sequences from known individuals collected for contamination control purposes will be assigned a unique identifier so that the sequences can be stored separately from the individual's name on GDNA. The sequences will be used to construct an elimination database housed on GDNA that can be queried when human DNA contamination is suspected or identified in casework that involves human DNA. Access to the NBFAC elimination database will be limited to a small number of NBFAC staff members involved in human DNA casework analysis, and the elimination database will be queried only when contamination is suspected. A lookup document (i.e., key) connecting the unique identifiers to specific individuals, as well as the signed informed consent forms (which necessarily include an individual's name) will be maintained on GDNA, but will be stored on a separate GDNA computer, password protected and accessible by only NBFAC staff members.

### **3.2 Does the project use technology to conduct electronic searches, queries, or analyses in an electronic database to discover or locate a predictive pattern or an anomaly? If so, state how DHS plans to use such results.**

No. DHS does not use GDNA to discover predictive patterns or anomalies.

### **3.3 Are there other components with assigned roles and responsibilities within the system?**

No other DHS components have access to GDNA. The FBI receives casework reports that contain information produced via data analysis on GDNA, but the FBI does not have direct access to the GDNA system.

### **3.4 Privacy Impact Analysis: Related to the Uses of Information**

**Privacy Risk:** There is a risk that GDNA users may use human DNA sequences stored on GDNA for unauthorized purposes, such as a project out of scope.

**Mitigation:** The risk is mitigated. All individuals with access to GDNA must read, acknowledge, and sign the NBACC Rules of Behavior, which prohibits unauthorized use of the system or data. All commands run on GDNA are logged and discoverable. Furthermore, as stated in Section 2.5, the NBACC volunteer DNA sequences and elimination database sequences that will be stored will be tightly access-controlled. Additionally, as described in Section 2.1, the



sequences themselves represent a tiny fraction of the human genome and are limited in their application to specific forensic uses.

All GDNA users must also take annual Privacy Training, Security Awareness Training, and Insider Threat training. Additionally, all users with Privileged Access, must take S&T Privileged Users training before receiving a Privileged User account and on an annual basis. GDNA logs are reviewed on a weekly basis by an Information Assurance team member for anomalies such as unauthorized access attempts.

## **Section 4.0 Notice**

### **4.1 How does the project provide individuals notice prior to the collection of information? If notice is not provided, explain why not.**

The only information NBFAC collects directly from individuals at present is the volunteer data. NBFAC provides notice by asking potential volunteers if they would be willing to provide a DNA sample along with self-reported information such as race/ethnicity/ancestry and phenotype (hair and eye color) to help develop and test human forensic analysis protocols. All volunteers received an informed consent form. The current project and any future projects are subject to review and approval by an Institutional Review Board (IRB), which requires volunteers to provide written informed consent.

NBFAC also intends to collect DNA samples from individuals who access the laboratory suites in which casework involving human DNA is processed, as well as from individuals who handle casework material containing human DNA prior to laboratory processing. NBFAC intends to develop an informed consent form in coordination with an IRB. The informed consent form will be reviewed and signed at the time of data collection.

NBACC relies on the agencies collecting the other information processed by GDNA (e.g., GenBank and other publicly available data, and anonymized human sample data) to provide appropriate notice to individuals when collecting the information. Such notice may not be required for law enforcement casework evidentiary samples, depending on that agency's requirements.

### **4.2 What opportunities are available for individuals to consent to uses, decline to provide information, or opt out of the project?**

The collection of data from volunteers is completely voluntary. All participants are required to provide written informed consent and have full autonomy to decline to provide information or to opt out of the project entirely.

The collection of DNA samples for the elimination database will also be entirely voluntary. However, individuals who decline to provide a DNA sample may not be permitted to enter laboratory spaces in which casework involving human DNA is processed, or otherwise handle



casework material that may contain human DNA.

NBACC relies on the agencies collecting the other information processed by GDNA to obtain consent from individuals when collecting the information. Such consent is not required for law enforcement casework evidentiary samples.

### **4.3 Privacy Impact Analysis: Related to Notice**

**Privacy Risk:** There is a risk that an individual is unaware that NBACC is using his or her information for DNA sequence analysis or that the information was obtained without the individual's consent.

**Mitigation:** This PIA provides notice to the public of NBACC's use of FBI casework data, publicly available reference sequence data, anonymized human sequence data, data collected from volunteers, and data collected for an elimination database. FBI casework data is considered Law-Enforcement Sensitive (LES), the nature of which may not implicate notice and consent requirements. The publicly available and commercial data used by NBFAC does not associate (or otherwise connect) human DNA sequences with the individual(s) from which the DNA was obtained. For human DNA sequences provided to NBFAC or developed by NBFAC from human samples provided by external partners, NBACC relies on the external partners to ensure consent is obtained for the collection and sharing of the information. The collection of data by NBACC from volunteers is performed with written informed consent from the volunteer, which includes information about the intended uses of the data. Data collected from individuals to establish and maintain an elimination database will be performed with written informed consent, which will include information about the intended uses of the data.

## **Section 5.0 Data Retention by the project**

### **5.1 Explain how long and for what reason the information is retained.**

NBACC retains GDNA system records in accordance with GRS 3.1, General Technology Management Records.

As described in Section 1.4, GDNA maintains records downloaded from publicly available sources in accordance with GRS 5.2, Item 20: Intermediary Records. Retention is temporary, and the data files can be destroyed when no longer needed for a business use. Records from volunteers that contain PII are not currently covered by a NARA-approved records retention schedule, as the recommended retention has not yet been archived to-date. Until GDNA has a NARA-approved records retention schedule, NBACC is not legally permitted to dispose of these records. NBACC has several requests for Records Disposition Authority pending approval for the casework and capability development scientific records, including the human DNA. The following paragraphs outline the requested schedules for the scientific records.



1. **Casework:** NBACC will retain records associated with casework, including sample analysis plans, chain of custody, sample accessioning, procedures for sample analyses, personnel authorizations, results and final case reports, in accordance with requested Records Schedule No. DAA-0563-2019-0005-0025 (i.e., 20 years), beginning at the submission of the final case report, for records generated for use in Law Enforcement Cases.

Records related to significant cases may have a permanent value and will be maintained in accordance with requested DAA-0563-2019-0005-0032. Significant sample case files are those of enduring historical value, which involved complex and/or novel issues or involved intense public interest or controversy that usually is reflected in a high degree of media attention or congressional scrutiny.

2. **Bioforensic Capability Research and Development (R&D):** The scientific records resulting from the R&D efforts include laboratory technical and scientific records and notes, including plans, results, technical reports, notebooks, and laboratory notes such as daily observations, detailed procedures, tabulations, charts, analyses, research progress, results of experiments, and other raw data. These records will be maintained in accordance with DAA-0563-2019-0005-0024 (i.e., 20 years), beginning at the submission of the final project report.

Some scientific records may have permanent value, and like casework significant records, will be maintained in accordance with DAA-0563-2019-0005-0032.

Some records will be retained as method accreditation records. These include records that document technical competence and compliance with International Organization for Standardization (ISO) requirements to carry out tests and/or calibrations, including sampling. Accreditation demonstrates technical competence and the ability to produce precise and accurate test and/or calibration data, in accordance with Records Schedule No. DAA-0563-2019-0005-0036. NBACC destroys these records five years after the method is superseded or obsolete. However, longer retention is authorized.

## 5.2 Privacy Impact Analysis: Related to Retention

**Privacy Risk:** There is a risk data may reside on GDNA for longer than necessary because GDNA does not have an associated records retention schedule or automated data destruction processes.

**Mitigation:** This risk is currently unmitigated. As mentioned in Section 5.1, records are pending coverage by a NARA-approved records retention schedule. Once a records retention schedule is approved, the schedule will be followed.

For case work, NBFAC follows evidentiary practices. NBFAC completes a log to document



the removal. For completed casework maintained by NBFAC, the corresponding casework data is archived, access-controlled, and can no longer be modified.

## Section 6.0 Information Sharing

### **6.1 Is information shared outside of DHS as part of the normal agency operations? If so, identify the organization(s) and how the information is accessed and how it is to be used.**

GDNA casework reports are shared with the FBI. NBACC does not share GDNA casework analysis results with any other organization outside of DHS.

R&D results developed using publicly available data, from commercial samples and data, or from anonymized human samples or sequences received from external partners may be shared outside of DHS with approved external collaborators. The sharing would occur under the terms of the DHS Memorandum of Agreement (MOA) with the FBI. Any external sharing also must comply with DHS information sharing processes and include privacy review. In addition, the information at issue may be made publicly available only following IRB approval, NBACC approval, FBI approval, and favorable DHS sensitivity review. In all such instances, NBFAC will not possess a key that could link the analysis results to any specific individual, and thus any results shared outside of DHS will not be traceable to a specific person. NBACC relies on the agencies collecting the samples or sequences processed by GDNA to obtain consent from individuals for data sharing and public release (if applicable).

Though information generated in the current capability development project that uses volunteers will not be shared outside of DHS, future R&D results developed from samples provided by volunteers may be shared outside of DHS. As previously described (see Section 2.5), sequence data from NBACC volunteer samples is stored anonymously with a unique identifier such that they cannot be linked to the source individual without a “lookup” document (key). Any release of NBACC volunteer data outside of GDNA and DHS would only be made in accordance with several procedures and requirements: (1) pre-approval of a project by an authorized IRB; (2) according to an informed consent document; (3) represented using only anonymous identifiers; (4) governed by GDNA access/data removal policies and the DHS MOA with the FBI; (5) FBI and DHS sensitivity reviews in the case of data that would be made public; and (6) in accordance with DHS information sharing processes, privacy review, and other oversight.

Data collected from individuals for the purpose of establishing and maintaining an elimination database will not be shared outside of DHS.



## **6.2 Describe how the external sharing noted in 6.1 is compatible with the SORN noted in 1.2.**

Currently, NBFAC does not share GDNA data in an identifiable form with agencies outside DHS.

## **6.3 Does the project place limitations on re-dissemination?**

GDNA casework reports are only shared with authorized officials. Original addresses or NBFAC officials must authorize the re-dissemination of GDNA casework reports and these reports must be shared in their entirety.

For anonymized human data to be shared with external partners, limitations on re-dissemination are included in the agreement developed between DHS and the partner, in accordance with DHS information sharing processes.

## **6.4 Describe how the project maintains a record of any disclosures outside of the Department.**

NBFAC does not make any disclosures of GDNA casework data. FBI maintains control of the casework data submitted and the subsequent reporting on the results. NBFAC would send any request for disclosure to the FBI for resolution. Any non-casework data disclosed to external entities will be subject to standard NBACC and DHS review procedures for sharing of information and accounting for disclosures.

## **6.5 Privacy Impact Analysis: Related to Information Sharing**

**Privacy Risk:** Information may be disclosed to external entities, and possibly re-disseminated, for unauthorized purposes.

**Mitigation:** The DHS Memorandum of Agreement (MOA) with the FBI defines the relationship between the parties in administering the forensic programs with specific provisions addressing the processes to be followed before information could be shared. Sharing for unauthorized purposes is mitigated through the physical, technical, and administrative controls previously discussed.

**Privacy Risk:** There is a risk that there may be an unauthorized disclosure of the DNA sequences or identifiers maintained in GDNA.

**Mitigation:** NBACC mitigates this risk by administering physical, technical, and administrative controls. NBFAC stores all casework DNA sequences and all DNA sequences collected from known individuals (e.g., NBACC employees and their families) on GDNA servers. NBACC physically controls GDNA servers by locating them in a server room with limited access within a secure facility guarded by armed guards and monitored by 24-hour video surveillance.



NBACC implements technical controls by requiring approved users to authenticate their identity in two ways. Users must use a Personal Identification Verification (PIV) card to access GDNA. The system then requires the users to input a password created by the individual user. Internal administrative controls include NBACC policies and processes requiring that users obtain approval for access through the submission of an application before a user can access GDNA.

NBACC and the FBI protect bioforensic casework data through an established administrative procedure. Only the FBI can submit bioforensic casework to NBACC.

DNA sequences generated from samples from known individuals (e.g., volunteers) are stored on a single workstation within the GDNA system. This workstation is disconnected from the outside world and is only accessible by staff members within the NBFAC Genomics group. Sequence data from each sample is stored anonymously with a unique identifier such that it cannot be linked to the source individual without a “lookup” document. The lookup document is stored on a separate computer, is password-protected, and can be accessed by only NBFAC staff members.

Finally, all storage media removed from GDNA (e.g., hard drives, random access memory) are subject to extensive sanitization procedures before disposal, thus ensuring that no data can be recovered from the media that is disposed.

In summary, this system of severely limited access, no external connectivity, separation of PII from sequence data, and proper disposal of storage media provides strong safeguards on the ability to access the document and map known individual identities to the DNA data that is stored and processed on the GDNA system.

## Section 7.0 Redress

### 7.1 What are the procedures that allow individuals to access their information?

Forensic information generated for the FBI by NBFAC forensic programs under the MOA is received from the FBI. Any requests NBACC receives for disclosure of forensic information to the public under the Freedom of Information Act or the Privacy Act<sup>34</sup> regarding projects and programs covered by the MOA will be forwarded to the FBI. The FBI shall determine whether such information must be released and will inform DHS of its determination.

Procedures for volunteers are outlined in the DHS RDT&E SORN. Procedures for access to information for the staff and visitors will be addressed in the forthcoming SORN DHS is developing.

---

<sup>34</sup> 5 U.S.C. § 552 and § 552a.



## **7.2 What procedures are in place to allow the subject individual to correct inaccurate or erroneous information?**

Procedures for volunteers are outlined in the DHS RDT&E SORN. Procedures for staff and visitors to correct inaccurate or erroneous information for the staff and visitors will be addressed in the forthcoming SORN DHS is developing.

## **7.3 How does the project notify individuals about the procedures for correcting their information?**

For the most part, this is not applicable for what is being processed and how it is being processed, as the sequences are not directly identifiable to an individual.

Procedures for volunteers are outlined in the DHS RDT&E SORN. Procedures for correcting information for the staff and visitors will be addressed in the forthcoming SORN DHS is developing.

## **7.4 Privacy Impact Analysis: Related to Redress**

**Privacy Risk**: GDNA does not provide a method of redress for incorrect information within the system.

**Mitigation**: NBACC does not make any law enforcement decisions based on GDNA analysis results. Redress procedures for volunteers are outlined in the DHS RDT&E SORN. Procedures for redress to information for the staff and visitors will be addressed in the forthcoming SORN DHS is developing. GDNA processes the other DNA information for the purpose of capability development and does not present redress issues.

**Privacy Risk**: There is a risk that an individual is unaware of how to access his or her records for redress or withdrawal from this system.

**Mitigation**: This risk is being mitigated and will be addressed through the IRB review process and development of informed consent forms with information about records access for purposes of redress or withdrawal. Current volunteers will be informed of their ability to update or withdraw from the program through a comprehensive informed consent notice.

## **Section 8.0 Auditing and Accountability**

### **8.1 How does the project ensure that the information is used in accordance with stated practices in this PIA?**

As part of the GDNA system accreditation and Ongoing Authorization, a System Security Plan update is being prepared and will be reviewed by DHS assessors. GDNA also implements the System Privacy Plan. Third-party audits were performed as part of the ATO accreditation process



as well as periodic reviews of security controls as part of the Ongoing Authorization program.

As part of the GDNA account provisioning process, all personnel requesting accounts must first complete Security Awareness training and document that they have read and understand the Rules of Behavior for the system. Remote access to the GDNA system is not possible. Only approved storage devices are permitted to be connected to the GDNA system and a weekly audit review of the system is conducted by the NBACC Security Analyst which includes a review of authorized and unauthorized login attempts.

## **8.2 Describe what privacy training is provided to users either generally or specifically relevant to the project.**

All GDNA system users must complete the following DHS S&T mandated training courses:

1. Privacy at DHS: Protecting Personal Information Training (annually);
2. IT Security Awareness Training (annually);
3. Safeguarding National Security Information (NSI) Training (one-time); and
4. Insider Threat Awareness Training (annually).

In addition, all NBFAC personnel with access to human DNA samples or sequences will complete Human Subjects Protection training.

The NBACC Training Coordinator ensures that all required training is completed.

## **8.3 What procedures are in place to determine which users may access the information and how does the project determine who has access?**

GDNA system users must possess appropriate clearance levels and must have a “need to know.” A user’s need to know is determined by whether the user needs access to the information in the performance of his or her duties. In addition, once approved by his or her supervisor, a user must complete the provisioning process described above before access is granted. The provisioning process starts when a user requests access and the supervisor determines the user has a need to know, then the supervisor passes the request to the program manager who ensures the user has met all training requirements. The program manager then sends the request to the system owner who formally creates the account and provides access.



## **8.4 How does the project review and approve information sharing agreements, MOUs, new uses of the information, new access to the system by organizations within DHS and outside?**

The MOA with the FBI defines the relationship between the parties in administering the forensic programs with specific provisions addressing the processes to be followed before information can be shared. The MOA states that any part of the MOA can be changed by mutual agreement of the executive agents. If another agency wants to use NBFAC's analytical capabilities, it can do so only by coordinating its request through the FBI. Any agreement established with a non-FBI party would be subject to review and approval by the FBI. All external sharing by NBACC or NBFAC remains subject to DHS information sharing processes and oversight, including privacy review.

### **Responsible Officials**

Julie S. Brewer  
Director  
Office of National Laboratories  
DHS Science & Technology Directorate

George Korch  
Laboratory Director  
National Biodefense Analysis and Countermeasures Center (NBACC)  
DHS Science & Technology Directorate

Maria Petrakis  
Privacy Officer  
DHS Science & Technology Directorate

### **Approval Signature**

Original, signed copy on file at the DHS Privacy Office.

---

Jonathan R. Cantor  
Acting Chief Privacy Officer  
Department of Homeland Security