Department of Health and Human Services

OFFICE OF
INSPECTOR GENERAL

# NIH GENERALLY IMPLEMENTED SYSTEM CONTROLS OVER THE SEQUENCE READ ARCHIVE BUT SOME IMPROVEMENTS NEEDED

*Inquiries about this report may be addressed to the Office of Public Affairs at [Public.Affairs@oig.hhs.gov](mailto:Public.Affairs@oig.hhs.gov).*

Amy J. Frontz
Deputy Inspector General
for Audit Services

February 2024
A-18-22-03300

# *Office of Inspector General*

https://oig.hhs.gov

The mission of the Office of Inspector General (OIG) is to provide objective oversight to promote the economy, efficiency, effectiveness, and integrity of the Department of Health and Human Services (HHS) programs, as well as the health and welfare of the people they serve. Established by Public Law
No. 95-452, as amended, OIG carries out its mission through audits, investigations, and evaluations conducted by the following operating components:

**Office of Audit Services.** OAS provides auditing services for HHS, either by conducting audits with its own audit resources or by overseeing audit work done by others. The audits examine the performance of HHS programs, funding recipients, and contractors in carrying out their respective responsibilities and provide independent assessments of HHS programs and operations to reduce waste, abuse, and mismanagement.

**Office of Evaluation and Inspections.** OEI's national evaluations provide HHS, Congress, and the public with timely, useful, and reliable information on significant issues. To promote impact, OEI reports also provide practical recommendations for improving program operations.

**Office of Investigations.** OI's criminal, civil, and administrative investigations of fraud and misconduct related to HHS programs and operations often lead to criminal convictions, administrative sanctions, and civil monetary penalties. OI's nationwide network of investigators collaborates with the Department of Justice and other Federal, State, and local law enforcement authorities. OI works with public health entities to minimize adverse patient impacts following enforcement operations. OI also provides security and protection for the Secretary and other senior HHS officials.

**Office of Counsel to the Inspector General.** OCIG provides legal advice to OIG on HHS programs and OIG's internal operations. The law office also imposes exclusions and civil monetary penalties, monitors Corporate Integrity Agreements, and represents HHS's interests in False Claims Act cases. In addition, OCIG publishes advisory opinions, compliance program guidance documents, fraud alerts, and other resources regarding compliance considerations, the anti-kickback statute, and other OIG enforcement authorities.

# *Notices*

## THIS REPORT IS AVAILABLE TO THE PUBLIC

at https://oig.hhs.gov

Section 8M of the Inspector General Act, 5 U.S.C. App., requires that OIG post its publicly available reports on the OIG website.

## OFFICE OF AUDIT SERVICES FINDINGS AND OPINIONS

The designation of financial or management practices as questionable, a recommendation for the disallowance of costs incurred or claimed, and any other conclusions and recommendations in this report represent the findings and opinions of OAS.  Authorized officials of the HHS operating divisions will make final determination on these matters.

Report in Brief
Date: February 2024
Report No. A-18-22-03300

U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES
OFFICE OF INSPECTOR GENERAL

## Why We Did This Audit

The Department of Health and Human Services (HHS), Office of Inspector General (OIG) has identified securing HHS data and systems to positively impact the cybersecurity posture of HHS and the sectors HHS influences as a key component within HHS's top management challenges.

The National Institutes of Health (NIH) Sequence Read Archive (SRA), which is hosted by National Library of Medicine (NLM), is the largest publicly available repository of high throughput sequencing data used for genomic research. The SRA holds diverse genomic data, including early COVID-19 sequencing, and is part of the International Nucleotide Sequence Database Collaboration.

The objective was to determine whether NIH has adequate controls in place to ensure data integrity of the NCBI Sequence Read Archive. OIG engaged the independent certified public accounting firm Brown & Company CPAs and Management Consultants, PLLC (Brown & Company) to conduct this audit.

## How We Did This Audit

To accomplish our objective, Brown & Company interviewed NIH officials, reviewed NIH's SRA information security policies and procedures, tested system controls; and examined 50 samples of the SRA data normalization and SRA Lite files to determine if the files were normalized as intended.

# NIH Generally Implemented System Controls Over the Sequence Read Archive But Some Improvements Needed

## What We Found

Brown & Company found that NIH adequately implemented most of the system and information integrity controls that ensure the integrity of the SRA data. However, control weaknesses were identified that should be addressed to improve the security of the SRA and its data.

While NIH stated the overall security categorization for the SRA was low impact, NIH did not document the rationale for the security categorization as is required by the National Institute of Standards and Technology (NIST) Special Publication (SP) 800-60 Volume 1, Revision 1.

NIH also did not conduct an SRA system-level risk assessment to identify threats and vulnerabilities as required by NIH's policy. However, NIH was required by NIST SP 800-53, Revision 4, to perform a system-level risk assessment for the SRA before it was authorized to operate and put into production.

In addition, the SRA data normalization policy lacked the assignment of roles and responsibilities to ensure the integrity of the SRA and its data.

## What We Recommend and NIH Comments

Brown & Company recommends that the NIH implement the recommendations below to improve controls over its SRA.

1. Complete the security categorization in accordance with FIPS Pub 199 to include documenting results and supporting rationale in the security plan.
2. Conduct a system-level risk assessment for the SRA in accordance with NIST SP 800-53 requirements and NIH polices.
3. Ensure that the data normalization policy and procedures comply with Federal requirements to include defining roles and responsibilities.

In written comments on our draft report, NIH concurred with all the recommendations and described actions it plans to take to implement the recommendations.

The full report can be found at https://oig.hhs.gov/oas/reports/region18/A182203300.asp.

# NIH GENERALLY IMPLEMENTED SYSTEM CONTROLS OVER THE SEQUENCE READ ARCHIVE BUT SOME IMPROVEMENTS NEEDED
# (A-18-22-03300)

## Final Report

**January 23, 2024**

**Prepared by**

**BROWN & COMPANY**

**CERTIFIED PUBLIC ACCOUNTANTS AND MANAGEMENT CONSULTANTS, PLLC**

*Point of Contact: Gail Jenifer, Member*
**6401 Golden Triangle Drive, Suite 310**
**Greenbelt, Maryland 20770**
*(240) 770-4903*
*gjenifer@brownco-cpas.com*
*TIN No.: 54-1783275, DUNS No.: 18-372-0515, Cage Code No.: 04TF0*

# NIH GENERALLY IMPLEMENTED SYSTEM CONTROLS OVER THE SEQUENCE READ ARCHIVE BUT SOME IMPROVEMENTS NEEDED (A-18-22-03300)

## Table of Contents

BROWN & COMPANY

CERTIFIED PUBLIC ACCOUNTANTS AND MANAGEMENT CONSULTANTS, PLLC

# INDEPENDENT ACCOUNTANT'S REPORT

National Institutes of Health
Sequence Read Archive
Washington, DC

Enclosed is the audit report on the National Institutes of Health (NIH) Sequence Read Archive (SRA). The U.S. Department of Health and Human Services (HHS) contracted with the independent certified public accounting firm Brown & Company CPAs and Management Consultants, PLLC (Brown & Company), to conduct a performance audit evaluating system and information integrity controls for the SRA. The objective of this performance audit was to determine whether NIH has implemented adequate system and information controls to ensure the integrity of SRA data.

The audit scope included testing NIH's compliance with Federal information technology (IT) laws, regulations, and standards. Brown & Company performed a security controls audit of select National Institute of Standards and Technology (NIST) Special Publication (SP) 800-53, Revision 4[1], information integrity controls and general controls. To assess control effectiveness, attribute testing was completed for fifty (50) SRA submissions/files selected from the 12-month period ending December 31, 2022. Audit fieldwork was performed remotely by Brown & Company, from November 14, 2022, through October 31, 2023.

This performance audit was performed in accordance with *Government Auditing Standards* issued by the Comptroller General of the United States. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

Based on the results of our performance audit, Brown & Company concluded that NIH needs to improve select federally required controls to ensure the integrity of SRA data. Brown & Company found NIH did not (1) conduct and document a security categorization in accordance with NIST *Federal Information Processing Standards Publication* (FIPS Pub) 199 for the SRA; (2) conduct a system-level risk assessment for the SRA; and (3) adequately implement data integrity policy and procedures for its data normalization process. Brown & Company made three recommendations for NIH to improve its system and information integrity controls to ensure the integrity of SRA data.

This report is for the purpose of concluding on the audit objective described above. Accordingly, this report is not suitable for any other purpose.

We appreciate the assistance we received from the staff of NIH and the opportunity to serve you. We will be pleased to discuss any questions you may have.

Brown & Company

Greenbelt, Maryland
January 23, 2024

[1] NIST SF 800-53, Revision 4, *Security and Privacy Controls for Federal Information Systems and Organizations;* System and Information Integrity Controls (SI): SI-1, SI-2, SI-3, SI-4, SI-5 as well as Risk Assessment (RA) controls RA-2 and RA-3.

# INTRODUCTION

## WHY WE DID THIS AUDIT

The Department of Health and Human Services (HHS), Office of Inspector General (OIG) has identified securing HHS data and systems to positively impact the cybersecurity posture of HHS and the sectors HHS influences as a key component within HHS's top management challenges. As HHS expands its technological capabilities, increases data sharing among HHS programs and the public, and improves data interoperability in the broader health care and public health systems, it must take crucial steps to modernize its approach to cybersecurity. The importance of improving cybersecurity posture across the Federal Government has been recognized by the President, such as in the May 2021 Executive Order *Improving the Nation's Cybersecurity*, which directed Federal agencies to change their approach fundamentally and systemically to cybersecurity.

The National Institutes of Health (NIH) Sequence Read Archive (SRA), which is hosted by the National Library of Medicine (NLM), is the largest publicly available repository of high throughput sequencing data. The SRA data is part of the International Nucleotide Sequence Database Collaboration which includes the National Center for Biotechnology Information (NCBI), the European Bioinformatics Institute, and the DNA Database of Japan. The SRA is a crucial resource for the scientific research community.

## OBJECTIVE

To determine whether NIH has adequate controls in place to ensure data integrity of the NCBI Sequence Read Archive.

## BACKGROUND

### National Institutes of Health

NIH is the primary Federal agency for conducting and supporting medical research to enhance health, lengthen life, and reduce illness and disability. For FY23, NIH was allocated $47.5[2] billion to support important medical research projects on cancer, Alzheimer's, diabetes, arthritis, heart ailments, and acquired immunodeficiency syndrome (AIDS). More than 84% of NIH's funding is awarded through competitive grants to more than 300,000 researchers at universities, medical schools, and other institutions in every state and around the world.[3]

NIH is composed of multiple institutes, centers, and offices that focus on specific areas of biomedical research and provide support for scientific advancements and public health initiatives. Within NIH, two important entities related to biomedical information and research are NLM and NCBI.

NLM is the world's largest biomedical library and plays a vital role in collecting, organizing, and providing access to a vast amount of biomedical information and literature. Its primary mission is to ensure that biomedical information is available to scientists, healthcare professionals, and the public. NLM offers an extensive range of resources, databases, and tools that support research, clinical practice, and public health. NLM oversees various biomedical information services, including PubMed, ClinicalTrials.gov, and the Unified Medical Language System.  NLM works to ensure seamless access to biomedical literature,

[2] FY 2023 NIH Operating Plan updated as of January 31, 2023.
[3] NIH website, https://www.nih.gov/about-nih/what-we-do/budget.

genomic data, and other valuable resources. Its efforts contribute significantly to advancing biomedical research, clinical practice, and public health initiatives worldwide.

NCBI is a division of NLM and serves as a hub for storing, organizing, and analyzing vast amounts of biological data generated by research studies worldwide. It makes available numerous databases and tools that support genomic research, molecular biology, and bioinformatics.

**Sequence Read Archive**

SRA is the world's largest publicly available repository of raw, unassembled genetic sequencing data. The purpose of the SRA repository is to store and make available sequencing data for the research community to search and conduct further genomic analyses. An example of the types of data submitted to the SRA includes the early genomic sequencing for Coronavirus Disease 2019 (COVID-19), which was submitted by a foreign researcher. At the time of this audit, the SRA held 14.5 million submissions/files and 16.5 petabytes of data; by 2025, the dataset is expected to grow by 50 petabytes.

In 2019, NIH engaged the SRA Data Working Group of the NIH Council of Councils to address the long-standing challenge of ensuring SRA's sustainability as an archive of exponentially growing experimental data. We examined the SRA Data Working Group recommendations for maintaining the increased growth of the NIH repository and noted:

> The SRA Data Working Group believes the SRA is an important part of NIH's mission for providing critical research data to the scientific research community.

When data is submitted to the SRA, the original files are large and vary in format (e.g., FASTQ, BAM, CRAM). Due to the size and varying formats, the original files are expensive to store, not useful for cross-analysis with other submissions, and impossible and cost-prohibitive for users to download. As a result, NIH implemented a data normalization process to address these issues.[4] The normalization process aims to:

- Transform the submitted data into one format, allowing data from different projects to be cross analyzed
- Reduce the size of the data, compacting the data and reducing storage costs
- Reduce the cost and difficulty of the user downloading data

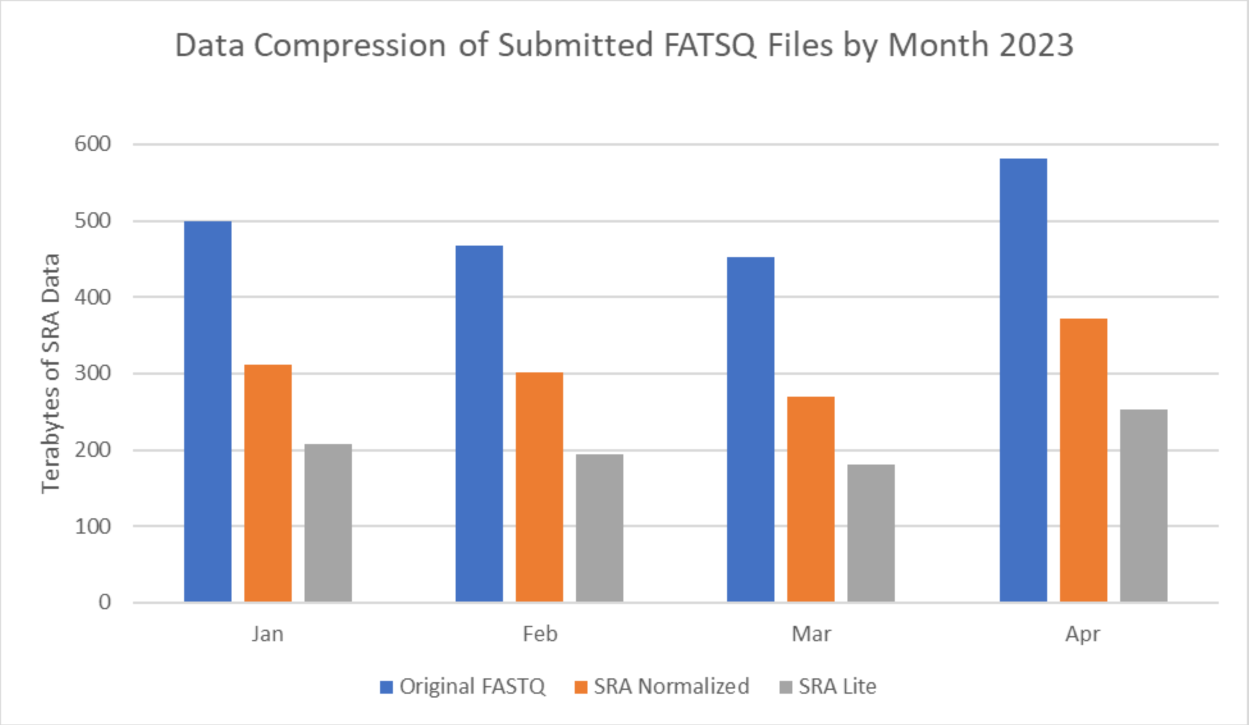
NIH provides two types of normalized files: SRA Normalized, the historical normalized format; and SRA Lite, a new, more compressed normalized format that was recently introduced. Both normalized formats can be stored, searched, downloaded, and most importantly, cross-analyzed in research conducted around the world.

Processing original data files into SRA Normalized and SRA Lite files not only provides compression, but also creates a file with a database-like structure that includes indexing and support information not present in the original submitted file.

It is important to note that the goal of SRA is to reduce the average size of submitted files across the entire archive. Processing of original data files into SRA Normalized and SRA Lite files achieves this goal (see Figure below), and on average, processing of submitted original files provides a significant reduction in

[4] Prior to data normalization, NIH saves the submitted source data in its original format in Amazon Web Services (AWS) Glacier, a secure and durable cloud storage service for low-cost data archiving and long-term backup. This step ensures that the source data is not altered by NIH processes.

size, even accounting for submitters utilizing generic file compressors. While not every submitted file will be reduced in size by SRA processing, this is a necessary design tradeoff to support fast data access and performance.



## HOW WE CONDUCTED THIS AUDIT

This performance audit concentrated on the controls that should be in place and operating effectively to protect the integrity of the SRA and its data, including flaw remediation, malicious code protection, and information input validation.

Specifically, we performed an audit of NIH's design and implementation of National Institute of Standards and Technology (NIST) Special Publication (SP) 800-53, Revision 4, system and information integrity (SI) controls, which included a review of other related security documentation to determine compliance with Federal requirements. We also examined a judgmental sample of fifty (50) SRA submissions/files to determine if the data normalization procedure was followed in accordance with NIH's data integrity policies.

Audit fieldwork was performed remotely by Brown & Company, located in Greenbelt, Maryland, from November 14, 2022 through October 31, 2023. This performance audit was conducted in accordance with *Generally Accepted Government Auditing Standards* (GAGAS), as specified in the Government Accountability Office's *Government Auditing Standards*. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objective. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objective.

**Appendix A** also briefly describes the details of our audit scope and methodology. **Appendix B** contains other relevant, specific Federal requirements and guidance.

## FINDINGS

NIH adequately implemented most of the system and information integrity controls to ensure the integrity of the SRA data. However, we identified control weaknesses that should be addressed to improve the security of the SRA and its data.

While NIH stated that the overall security categorization for the SRA was low impact, we found that NIH did not document the rationale for the security categorization as is required by NIST SP 800-60 Volume 1, Revision 1, *Guide for Mapping Types of Information and Information Systems to Security Categories*. If the impact level was not properly determined, NIH may not have implemented the minimum security controls needed to ensure the protection and integrity of the SRA and its data. The difference in the required controls for a low-impact versus moderate-impact baseline for the confidentiality, integrity, and availability of an information and system is significant, as a low-impact baseline consists of 115 security controls, but there are 159 controls for a moderate-impact. As it relates to SI controls, the difference between the number of baseline controls for a low-impact system and moderate-impact system is five controls.[5]

NIH also did not conduct an SRA system-level risk assessment to identify threats and vulnerabilities as required by NIH's policy. The information NIH would have ascertained as part of the risk assessment process could impact the implementation of SRA system integrity controls. NIH delayed conducting the SRA risk assessment because an agency-level security assessment to include "system risk" was being planned; however, NIH is required by NIST SP 800-53, rev 4, to perform a system-level risk assessment for SRA before it was authorized to operate and put into production. The absence of a system-level risk assessment means that NIH may not have identified threats to the SRA, and therefore NIH may not have implemented the appropriate cybersecurity controls to mitigate the threats.

We determined the SRA data normalization policy lacked the assignment of roles and responsibilities to ensure data integrity. In addition, the normalization procedures did not include a requirement for review processes (oversight) to ensure the modification (normalization) of the SRA submissions/files were completed correctly. NIH data normalization policies and procedures were not fully compliant with Federal requirements to assign roles and responsibilities to ensure the integrity of the SRA and its data.

## SRA SYSTEM SECURITY CATEGORIZATION

### Federal Requirements

NIST FIPS Pub 199, *Standards for Security Categorization of Federal Information and Information Systems,* provides those standards for categorizing information and information systems as low-impact, moderate-impact, or high-impact for confidentiality, integrity, and availability based on security objectives. "The security categories are based on the potential impact on an organization should certain events occur which jeopardize the information and information systems needed by the organization to accomplish its assigned mission, protect its assets, fulfill its legal responsibilities, maintain its day-to-day functions, and protect individuals" (section 2). The resulting security categorization helps the organization determine the security and privacy control baselines to protect the system, as detailed in NIST SP 800-53.

NIST SP 800-53*, Revision 4, Security and Privacy Controls for Federal Information Systems and Organizations*, §1.1 provides "guidelines for selecting and specifying security controls for organizations and information systems supporting the executive agencies of the federal government to meet the

[5] The number of security controls is based on NIST SP 800-53, Revision 4, and represents the total number of controls and control enhancements identified for each system impact level.

requirements of FIPS Publication 200, *Minimum Security Requirements for Federal Information and Information Systems*."

NIST SP 800-60 Volume 1, R*evision 1, Guide for Mapping Types of Information and Information Systems to Security Categories,* §1.1 addresses the FISMA direction to develop guidelines recommending the types of information and information systems to be included in each category of potential security impact. "This guideline is intended to help agencies consistently map security impact levels to types of: (i) information (e.g., privacy, medical, proprietary, financial, contractor sensitive, trade secret, investigation); and (ii) information systems (e.g., mission critical, mission support, administrative). This guideline applies to all Federal information systems other than national security systems. National security systems store, process, or communicate national security information."

### NIH Needs to Perform a System Security Categorization In Accordance With Federal Requirements

On November 7, 2022, NIH categorized the SRA as a FIPS Pub 199 low-impact system, which conveys that the loss of the largest publicly available repository of high throughput sequencing data would have a "low" impact on the NLM mission for providing critical research data to the scientific research community.

The NIH did not document the supporting rationale for its security categorization determination. By not documenting key information, NIH did not comply with the NIST SP 800-60 requirements. We are unable to assess or confirm that NIH completed the procedures and properly determined the appropriate impact. Since the results of the categorization process are used to select the security controls for the system, the SI security controls implemented to secure the SRA may have been based on a categorization that is lower than it should be, resulting in a set of controls that may not adequately protect the SRA and its data.

## SRA SYSTEM-LEVEL RISK ASSESSMENT

### Federal Requirements

NIST SP 800-37, Revision 2*, Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy*, requires that individual security controls applicable to a system be identified and assessed in support of the system's authorization.

NIST SP 800-53, Revision 4, *Security and Privacy Controls for Federal Information Systems and Organizations*, RA-3 Risk Assessment control states the requirements for conducting, documenting, reviewing, disseminating, and updating a risk assessment.

In addition, the *NIH NLM Information System Security Handbook,* Version 4.1, November 17, 2022, states a risk assessment shall be conducted on all information systems as required by NIST, Federal Information Security Modernization Act (FISMA), and NIH guidelines. The *NIH NLM Risk Management Policy and Procedures*, November 29, 2022, provides directions for assessing, monitoring, and communicating an agency's risk assessment.

### NIH Needs to Conduct a System-Level Risk Assessment

NIH has not conducted a system-level risk assessment for the SRA to identify the threats and vulnerabilities of the SRA as required by NIH policy. NIH has developed an SRA system security plan which documents the implementation of the baseline security controls selected; however, the controls may not be tailored to mitigate the risks specific to the SRA since an assessment was not completed.

NIH stated that it delayed conducting the SRA system-level risk assessment because management planned to issue an overall NIH security assessment to include system risk. As noted above, NIST requires that a system level risk assessment be completed prior to authorizing the system to operate. By not conducting a system-level risk assessment of the SRA, NIH has minimal assurance that risks specific to the SRA have been identified. Also, by not knowing these risks and related threats to the SRA, NIH may not have implemented the appropriate cybersecurity controls to mitigate the risks and threats to an acceptable level, as determined by Federal requirements and organizational goals and missions.

## SRA DATA NORMALIZATION POLICY AND PROCEDURES

### Federal Requirements

NIST SP 800-53, Revision 4, *Security and Privacy Controls for Federal Information Systems and Organizations*, SI-1 Policy and Procedures control states the requirements for developing, documenting, disseminating, reviewing, and updating system and information integrity policies and procedures. The policy should address purpose, scope, roles, responsibilities, management commitment, coordination among organizational entities, and compliance. The procedures should support the policy implementation.

*SRA Data Retention Policy* describes SRA's practice of storing and managing sequence archived data. The policy defines and documents format-specific data retention guidelines to help guide procedures, organizational management, software development, and user expectations.

*SRA End-to-End Procedures* detail the processes used for managing sequence data from submission to storage. The goal is to reduce the size of the data, which will reduce storage costs and reduce the cost and difficulty of downloading data.

### NIH Needs to Further Develop SRA Data Normalization Policy and Procedures

The SRA data normalization policy and procedures to ensure data integrity, which is included in the *SRA Data Retention Policy*, do not meet Federal requirements. Specifically, we determined through review that the policy lacked the requirements of assigning roles and responsibilities. In addition, NIH did not have documented procedures for the oversight of the normalization process to ensure the normalization of the SRA files were completed correctly.

We tested a sample of fifty (50) SRA submissions/files that were from December 31, 2022, to determine if the data normalization processes were followed in accordance with data normalization policies. Using the NIH SRA Tool, we examined 50 samples of the SRA normalized and SRA Lite files to compare the size of the files to the original data files (raw data).[6] The purpose of normalizing data is to reduce the sequencing data file size, which can arise from differences in sequencing depth, library preparation, and sequencing platforms. Therefore, after normalization, the goal is for the data to be smaller in size compared to the original uploaded file. Because SRA Lite files are a further reduction in file size of the already normalized data, they also should be smaller than both the original file and the normalized file in the SRA. While the goal of the normalization process is to reduce the average size of submitted files across the entire archive, this is not always achievable due to attributes of the original file. Our test results showed 49 out of 50 SRA normalized files were larger than the original files (see Figure 1 on next page), and 29 out of 50 SRA Lite files were larger than the original files (see Figure 2 on next page).

[6] The tested data came from a single submitter, a single sequencing technology, and a single small virus.

**Figure 1: Comparison of 50 samples of the SRA normalizated data files to original files. Results; 49 out of 50 SRA data normalization files were larger than 100% of the original files.**

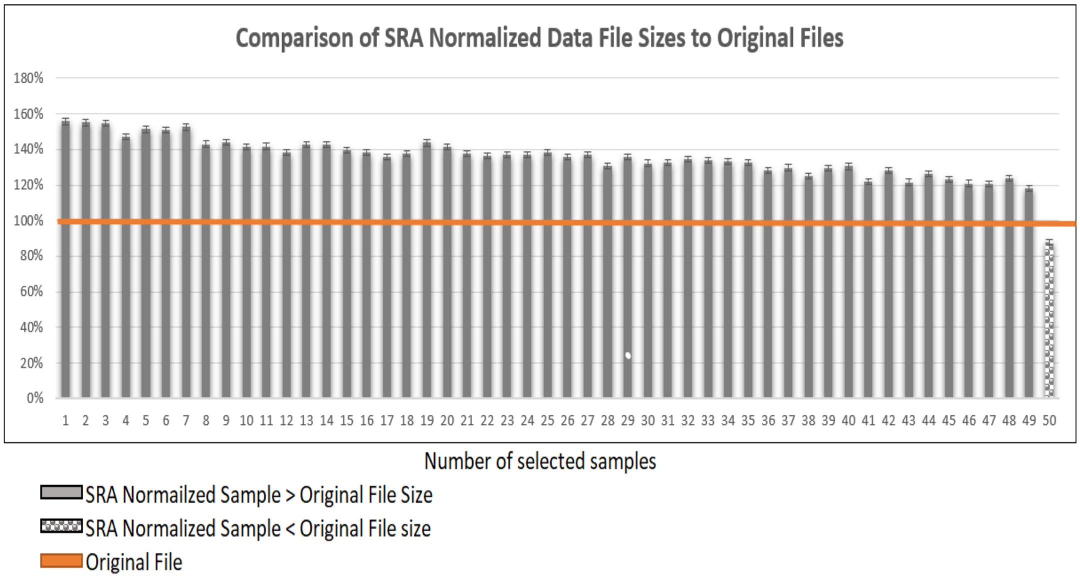
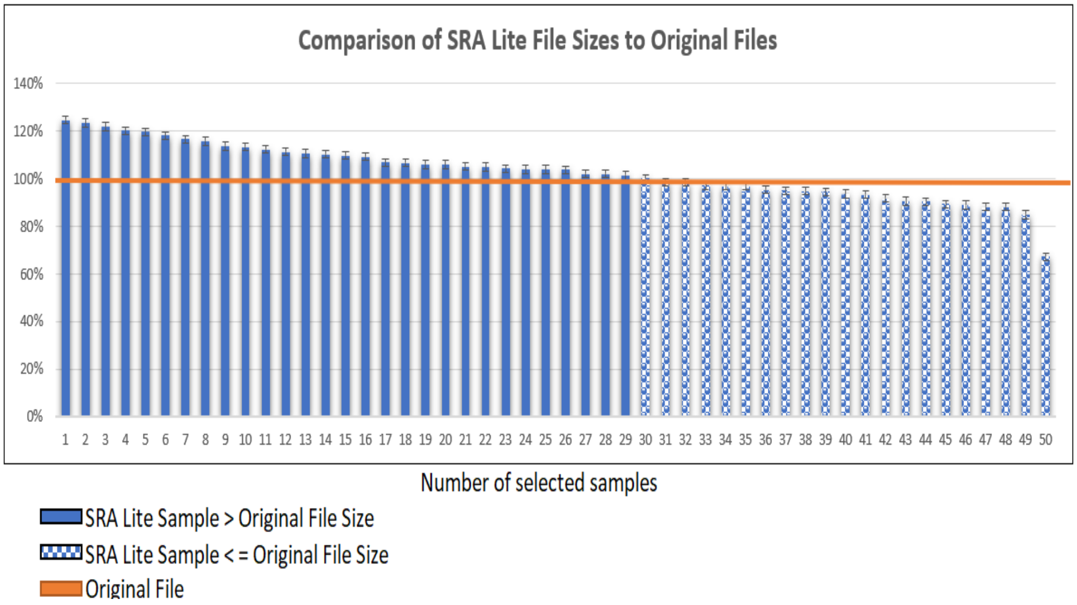


**Figure 2: Comparison of 50 samples of the SRA Lite files to original files. Results; 29 out of 50 SRA Lite files were larger than 100% of the original files.**



Once we shared our results with NIH, they conducted validation spot checks of the files and determined the increases in file sizes were rare occurrences that are explainable. However, these increases in file size were not previously validated because the normalization policy and procedures did not contain roles and responsibilities, to include assigning oversight responsibilities to ensure the normalization of the SRA files were completed correctly. The lack of defined roles and responsibilities within the normalization policy and procedures could lead to undetected errors and mistakes and missed opportunities for improving the data normalization process. Also, it can impact accountability and ownership of tasks, making it difficult to determine who is responsible for addressing issues or resolving conflicts, ultimately leading to delays in problem-solving and decision-making.

## RECOMMENDATIONS

Brown & Company recommends that the NIH implement the recommendations below to improve system and information integrity controls over its SRA.

- Complete the security categorization in accordance with FIPS Pub 199 to include documenting results and supporting rationale in the security plan.
- Conduct a system-level risk assessment for the SRA in accordance with NIST SP 800-53 requirements and NIH policies.
- Ensure that the data normalization policy and procedures comply with Federal requirements to include defining roles and responsibilities.

## NIH COMMENTS AND BROWN & COMPANY RESPONSE

In written comments on our draft report, NIH concurred with all our recommendations and described actions it plans to take to implement our recommendations. NIH also provided technical comments, which we addressed as appropriate. NIH's general comments are included in their entirety in Appendix C.

## APPENDIX A: SCOPE AND METHODOLOGY

### SCOPE

Brown & Company's audit scope included NIH's design and implementation of NIST SP 800-53, Revision 4, SI controls: SI-1, SI-2, SI-3, SI-4, SI-5, and other federally required controls related to securing the SRA and its data, such as system categorization and system-level risk assessment. In addition, Brown & Company tested a judgmental sample of 50 SRA submissions/files, to determine if the data normalization processes were followed in accordance with NIH's data integrity policies.

Audit fieldwork was performed remotely from November 14, 2022 to October 30, 2023.

### METHODOLOGY

To accomplish our audit objective, Brown & Company:

- Interviewed key personnel and reviewed legal and regulatory requirements stipulated by NIH;
- Obtained and reviewed policy and procedure documentation related to the SRA information security program and data normalization process;
- Obtained and reviewed federally required cybersecurity documentation, such as the system security plan and FIPS Pub 199 security categorization;
- Tested system processes to determine the adequacy of selected system and information integrity controls; and
- Examined 50 samples of the SRA data normalization and SRA Lite files to determine the size of the files compared to the original data files.

We tested internal controls that we considered significant to meet the audit objective. Accordingly, we obtained an understanding of the internal controls over the SRA system through interviews and observations, as well as inspection of various documents, including information technology and other related organizational policies and procedures.

Brown & Company conducted this audit in accordance with performance auditing standards, in accordance with *Generally Accepted Government Auditing Standards (GAGAS)*, also known as the Yellow Book, which is issued by the Government Accountability Office (GAO). Those standards require that the auditor plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for their findings and conclusions based on the audit objectives.

## APPENDIX B: FEDERAL REQUIREMENTS AND GUIDANCE

The federal requirements and guidance used in conducting this audit included:

1. **44 U.S. Code § 3554 - Federal agency, states:**

   *a) In General. —The head of each agency shall—*

   *(B)complying with the requirements of this subchapter, subchapter III of chapter 13 of title 41, and related policies, procedures, standards, and guidelines, including—*

   *(ii)operational directives developed by the Secretary under section 3553(b).*

2. **Office of Management and Budget (OMB) Circular NO. A-130, states:**

   ***General Requirements***

   *Agencies shall develop, implement, document, maintain, and oversee agency-wide information security and privacy programs including people, processes, and technologies to:*

   *1) Provide for agency information security and privacy policies, planning, budgeting, management, implementation, and oversight.*

3. **NIST Standards for Security Categorization of Federal Information and Information Systems, Federal Information Processing Standards Publication (FIPS Pub) 199:**

   FIPS Pub 199 requires agencies to categorize information and information systems as low-impact, moderate-impact, or high impact for the security objectives of confidentiality, integrity, and availability. *The security categories are based on the potential impact on an organization should certain events occur which jeopardize the information and information systems needed by the organization to accomplish its assigned mission, protect its assets, fulfill its legal responsibilities, maintain its day-to-day functions, and protect individuals*.

   *The characterization of information or an information system is based on an assessment of the potential impact that a loss of confidentiality, integrity, or availability of such information or information systems would have on organizational operations, organizational assets, or individuals.*

4. **NIST SP-800-37 Risk Management Framework for Information Systems and Organizations, A System Life Cycle Approach for Security and Privacy, Revision 2, states:**

   *The purpose of the Categorize step is to inform organizational risk management processes and tasks by determining the adverse impact to organizational operations and assets, individuals, other organizations, and the Nation with respect to the loss of confidentiality, integrity, and availability of organizational systems and the information processed, stored, and transmitted by those systems.*

*Security categorization results reflect the organization's risk management strategy.*

*The purpose of the Assess step is to determine if the controls selected for implementation are implemented correctly, operating as intended, and producing the desired outcome with respect to meet the security and privacy requirements for the system and the organization.*

*The purpose of the Monitor step is to maintain an ongoing situational awareness about the security and privacy posture of the information system and the organization in support of risk management decisions.*

5. **NIST SP 800-53, Revision 4, Security and Privacy Controls for Federal Information Systems and Organizations, states:**

***RA-2 SECURITY CATEGORIZATION***

*Control Statement*

*The organization:*

*a. Categorizes information and the information system in accordance with applicable federal laws, Executive Orders, directives, policies, regulations, standards, and guidance;*
*b. Documents the security categorization results (including supporting rationale) in the security plan for the information system; and*
*c. Ensures that the authorizing official or authorizing official designated representative reviews and approves the security categorization decision.*

***RA-3 RISK ASSESSMENT***

*Control Statement*

*The organization:*

*a. Conducts an assessment of risk, including the likelihood and magnitude of harm, from the unauthorized access, use, disclosure, disruption, modification, or destruction of the information system and the information it processes, stores, or transmits.*
*b. Documents risk assessment results in [Selection: security plan; risk assessment report; [Assignment: organization-defined document]];*
*c. Reviews risk assessment results [Assignment: organization-defined frequency];*
*d. Disseminates risk assessment results to [Assignment: organization-defined personnel or roles]; and*
*e. Updates the risk assessment [Assignment: organization-defined frequency] or whenever there are significant changes to the information system or environment of operation (including the identification of new threats and vulnerabilities), or other conditions that may impact the security state of the system.*

### *SI-1 System and Information Integrity Policy and Procedures*

*Control Statement*

*The organization:*

a. *Develops, documents, and disseminates to [Assignment: organization-defined personnel or roles]:*

   1. *A system and information integrity policy that addresses purpose, scope, roles, responsibilities, management commitment, coordination among organizational entities, and compliance; and*
   2. *Procedures to facilitate the implementation of the system and information integrity policy and associated system and information integrity controls; and*

b. *Reviews and updates the current:*

   1. *System and information integrity policy [Assignment: organization-defined frequency]; and*
   2. *System and information integrity procedures [Assignment: organization-defined frequency].*

### *SI-2 Flaw Remediation*

*Control Statement*

*The organization:*

a. *Identifies, reports, and corrects information system flaws;*
b. *Tests software and firmware updates related to flaw remediation for effectiveness and potential side effects before installation;*
c. *Installs security-relevant software and firmware updates within [Assignment: organization defined time period] of the release of the updates; and*
d. *Incorporates flaw remediation into the organizational configuration management process. A system and information integrity policy that addresses purpose, scope, roles, responsibilities, management commitment, coordination among organizational entities, and compliance; and*

### *SI-3 Malicious Code Protection*

*Control Statement*

*The organization:*

a. *Employs malicious code protection mechanisms at information system entry and exit points to detect and eradicate malicious code;*
b. *Updates malicious code protection mechanisms whenever new releases are available in accordance with organizational configuration management policy and procedures;*
c. *Configures malicious code protection mechanisms to:*

   1. *Perform periodic scans of the information system [Assignment: organization-defined frequency] and real-time scans of files from external sources at [Selection (one or more); endpoint; network entry/exit points] as the files are*

*downloaded, opened, or executed in accordance with organizational security policy; and*

2. *[Selection (one or more): block malicious code; quarantine malicious code; send alert to administrator; [Assignment: organization-defined action]] in response to malicious code detection; and*

d. *Addresses the receipt of false positives during malicious code detection and eradication and the resulting potential impact on the availability of the information system.*

### *SI-4 Information System Monitoring*

*Control Statement*

*The organization:*

a. *Monitors the information system to detect:*

1. *Attacks and indicators of potential attacks in accordance with [Assignment: organization defined monitoring objectives]; and*
2. *Unauthorized local, network, and remote connections;*

b. *Identifies unauthorized use of the information system through [Assignment: organization defined techniques and methods];*

c. *Deploys monitoring devices:*

1. *Strategically within the information system to collect organization-determined essential information; and*
2. *At ad hoc locations within the system to track specific types of transactions of interest to the organization;*

d. *Protects information obtained from intrusion-monitoring tools from unauthorized access, modification, and deletion;*

e. *Heightens the level of information system monitoring activity whenever there is an indication of increased risk to organizational operations and assets, individuals, other organizations, or the Nation based on law enforcement information, intelligence information, or other credible sources of information;*

f. *Obtains legal opinion with regard to information system monitoring activities in accordance with applicable federal laws, Executive Orders, directives, policies, or regulations; and*

g. *Provides [Assignment: organization-defined information system monitoring information] to [Assignment: organization-defined personnel or roles] [Selection (one or more): as needed; [Assignment: organization-defined frequency]].*

***SI-5 Security Alerts, Advisories, and Directives***

*Control Statement*

*The organization:*

*a. Receive system security alerts, advisories, and directives from [Assignment: organization-defined external organizations] on an ongoing basis;*
*b. Generate internal security alerts, advisories, and directives as deemed necessary;*
*c. Disseminate security alerts, advisories, and directives to: [Assignment (one or more): [Assignment: organization-defined personnel or roles], [Assignment: organization-defined elements within the organization], [Assignment: organization-defined external organizations]; and*
*d. Implement security directives in accordance with established time frames or notify the issuing organization of the degree of noncompliance.*

6. **NIST SP 800-60 Volume 1, Revision 1, Guide for Mapping Types of Information and Information Systems to Security Categories:**

   Addresses the FISMA direction to develop guidelines recommending the types of information and information systems to be included in each category of potential security impact. This guideline is intended to help agencies consistently map security impact levels to types of: (i) information (e.g., privacy, medical, proprietary, financial, contractor sensitive, trade secret, investigation); and (ii) information systems (e.g., mission critical, mission support, administrative). This guideline applies to all Federal information systems other than national security systems. National security systems store, process, or communicate national security information.

7. **NIH NLM Information System Security Handbook, Version 4.1, November 17, 2022**

   States that a risk assessment shall be conducted on all information systems as required by NIST, Federal Information Security Modernization Act (FISMA), and NIH guidelines.

8. **NIH NLM Risk Management Policy and Procedures, November 29, 2022**

   Provides directions for assessing, monitoring, and communicating an agency’s risk assessment.

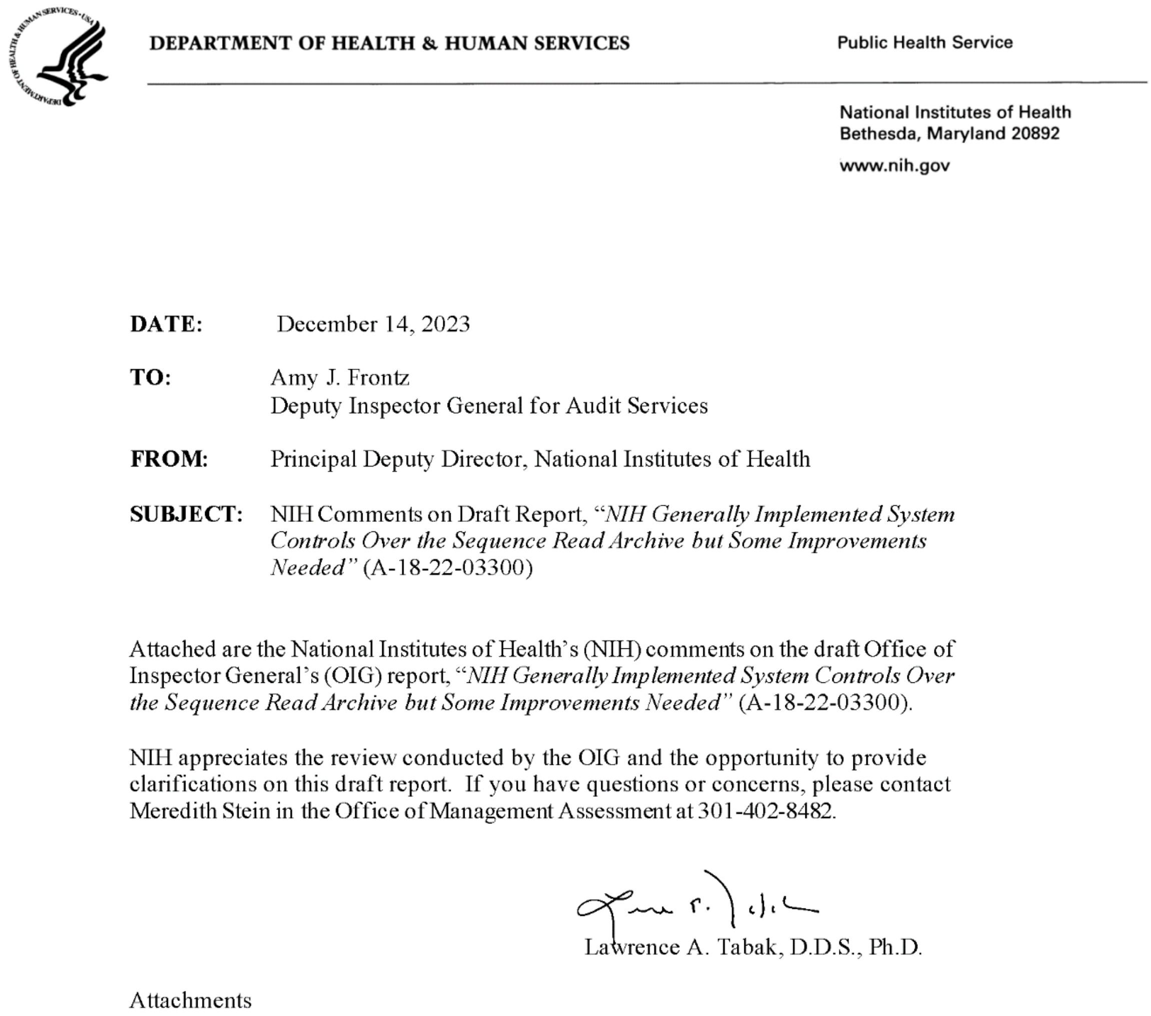
9. **SRA Data Retention Policy**

   Describes SRA's practice of storing and managing sequence archived data. The policy defines and documents format-specific data retention guidelines to help guide procedures, organizational management, software development, and user expectations.

10. **SRA End-to-End Procedures**

    Details the processes used for managing sequence data from submission to storage. The goal is to reduce the size of the data, which will reduce storage costs and reduce the cost and difficulty of downloading data.

# APPENDIX C: NIH'S MANAGEMENT COMMENTS



**DEPARTMENT OF HEALTH & HUMAN SERVICES**

Public Health Service

National Institutes of Health
Bethesda, Maryland 20892
www.nih.gov

**DATE:** December 14, 2023

**TO:** Amy J. Frontz
Deputy Inspector General for Audit Services

**FROM:** Principal Deputy Director, National Institutes of Health

**SUBJECT:** NIH Comments on Draft Report, "*NIH Generally Implemented System Controls Over the Sequence Read Archive but Some Improvements Needed*" (A-18-22-03300)

Attached are the National Institutes of Health's (NIH) comments on the draft Office of Inspector General's (OIG) report, "*NIH Generally Implemented System Controls Over the Sequence Read Archive but Some Improvements Needed*" (A-18-22-03300).

NIH appreciates the review conducted by the OIG and the opportunity to provide clarifications on this draft report. If you have questions or concerns, please contact Meredith Stein in the Office of Management Assessment at 301-402-8482.

Lawrence A. Tabak, D.D.S., Ph.D.

Attachments

## GENERAL COMMENTS OF THE NATIONAL INSTITUTES OF HEALTH (NIH) ON THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (HHS) OFFICE OF INSPECTOR GENERAL (OIG) DRAFT REPORT ENTITLED: "NIH GENERALLY IMPLEMENTED SYSTEM CONTROLS OVER THE SEQUENCE READ ARCHIVE BUT SOME IMPROVEMENTS NEEDED" (A-18-22-03300)

The National Institutes of Health (NIH) appreciates the opportunity to review and comment on this report and wants to thank the Office of Inspector General (OIG) for their partnership and efforts during this audit.

NIH concurs with the recommendations made by OIG. Since November 2022, NIH has worked collaboratively with OIG and initiated corrective actions. As a result, NIH is pleased to report that it has already taken actions to address the three recommendations. NIH anticipates that this effort will be complete within the next 4 months.

NIH appreciates the opportunity to comment on this report and looks forward to continued collaboration with the OIG audit team on this important effort.

**OIG Recommendation 1:**
Complete the security categorization in accordance with FIPS Pub 199 to include documenting results and supporting rationale in the security plan.

**NIH Response:**
NIH concurs with OIG's finding and corresponding recommendation regarding the Sequence Read Archive (SRA) security categorization.

A new authorization boundary for information systems managed by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine, including SRA, is being documented. As part of that process, security categorizations for NCBI-managed repositories are being reviewed to ensure that they are complete, in accordance with FIPS PUB 199, and contain appropriate justification. The expected completion date of this process is March 31, 2024.

**OIG Recommendation 2:**
Conduct a system-level risk assessment for the SRA system in accordance with NIST SP 800- 53 requirements and NIH policies.

**NIH Response:**
NIH concurs with OIG's finding and corresponding recommendation that NLM should conduct a risk assessment.

NLM has a planned risk assessment of SRA for the second quarter of FY 2024, which will be conducted in accordance with NIST SP 800- 53 requirements and NIH policies. The expected completion date of this process is March 31, 2024.

**OIG Recommendation 3:**
Ensure that the data normalization policy and procedures comply with Federal requirements to include defining roles and responsibilities.

Page **1** of **2**

**GENERAL COMMENTS OF THE NATIONAL INSTITUTES OF HEALTH (NIH) ON THE DEPARTMENT OF HEALTH AND HUMAN SERVICES (HHS) OFFICE OF INSPECTOR GENERAL (OIG) DRAFT REPORT ENTITLED: "NIH GENERALLY IMPLEMENTED SYSTEM CONTROLS OVER THE SEQUENCE READ ARCHIVE BUT SOME IMPROVEMENTS NEEDED" (A-18-22-03300)**

**NIH Response:**
NIH concurs with OIG's recommendation to improve the documentation of data normalization policy and procedures included in the *SRA Data Retention Policy* and the *SRA End-to-End Procedures*, including defining roles and responsibilities.

These documents will be revised to meet Federal requirements by March 31, 2024.

Page **2** of **2**