This document is scheduled to be published in the Federal Register on 2026-09-30 and available online at https://www.federalregister.gov/d/2026-20023, and on https://govinfo.gov

ode: 3510-04

DEPARTMENT OF COMMERCE

National Technical Information Service

[Docket No.: 260924-0004]

Request for Information (RFI) Regarding the Digitization and Modernization of the National Technical Reports Library (NTRL) and the Leveraging of Scientific, Technical and Engineering Information (STEI) Stored in the NTRL

AGENCY: National Technical Information Service (NTIS), U.S. Department of Commerce.

ACTION: Notice; request for information.

SUMMARY: The National Technical Information Service (NTIS) is seeking general information, feedback, suggestions, and experiential and technical insights from stakeholders to inform NTIS's goal of fully digitizing, modernizing, and enhancing the utility of the National Technical Reports Library (NTRL). NTIS has historically collected, indexed, abstracted, and stored U.S. Government-sponsored technical reports and made them available to the public through the NTRL. To further unlock the NTRL's intrinsic value, NTIS aims to digitize records that are part of the NTRL but currently exist only in physical formats. Leveraging the information archived within the NTRL may include making the data more accessible to AI and advanced computing applications. NTIS seeks to understand the opportunities, challenges, and priorities that may attend NTRL's efforts to modernize and make any resultant high-quality technical federal data sets publicly available.

DATES: Comments must be received by 11:59 PM Eastern Time on **[INSERT DATE 45 DAYS AFTER DATE OF PUBLICATION IN THE *FEDERAL REGISTER*]**. Submissions received after that date may not be considered.

ADDRESSES: Comments must be submitted electronically via the Federal e-Rulemaking Portal (regulations.gov):

- To submit electronic public comments via the Federal e-Rulemaking Portal.

1. Go to www.regulations.gov and enter NIST-2026-0166 in the search field.
2. Click the "Comment Now!" icon and complete the required fields.
3. Enter or attach your comments.

Comments containing references, studies, research, and other empirical data that are not widely published should include copies of the referenced materials. All submissions, including comments, attachments and other supporting materials, will become part of the public record and subject to public disclosure. Relevant comments will generally be available on the Federal e-Rulemaking Portal at www.regulations.gov. NTIS will not accept comments accompanied by a request that part or all of the material be treated confidentially because of its business proprietary nature or for any other reason. Therefore, do not submit confidential business information or otherwise sensitive, protected, or personal information, such as account numbers, Social Security numbers, or names of other individuals.

NTIS will not accept comments for this notice by postal mail, fax, or email. To ensure that NTIS does not receive duplicate copies, please submit your comments only once.

Additional information on the use of regulations.gov, including instructions for accessing agency documents, submitting comments, and viewing the docket is available at: www.regulations.gov/faq.

FOR FURTHER INFORMATION CONTACT: For questions about this notice please contact: Bobby Khondker via email at bobby.khondker@nist.gov and ntrl@nist.gov, or by phone at (301) 873-2127.

SUPPLEMENTARY INFORMATION:

NTIS operates the NTRL as a clearinghouse for the collection and dissemination of federally-funded scientific, technical, and engineering information (STEI). The NTRL hosts the largest collection of U.S. government-sponsored technical reports, and historically has collected, authenticated, indexed, preserved, and made available to the public technical information generated through federal research, contracts, grants, laboratories, and other government-

sponsored activities—including material that may never be published in conventional academic journals.

Reports in the NTRL include detailed results of federally funded research and development (R&D), which can contain technical details that are often omitted from journal publications, such as: experimental configurations and methods; performance characteristics and failure modes; material properties and measurement results; prototype designs; field-test results; engineering tolerances and assumptions; cost, reliability, and manufacturability assessments; negative or inconclusive results; contractor-developed methods and tools; and recommendations for follow-on R&D. For technology developers, the detailed results found in the reports can help establish the existing state of the art, identify prior government investment, reproduce earlier experiments, locate promising but unfinished technology, and avoid repeating approaches that federal research has already shown to be impractical.

Other available and valuable data in the NRTL include emerging-technology assessments and horizon scanning (for example, the NTRL has the Army's *Emerging Science and Technology Trends: 2017–2047*, which examines artificial intelligence, quantum computing, additive manufacturing, materials science, renewable energy, biotechnology, and other emerging areas to inform strategic dialogue and future R&D investments); technology-feasibility and technology-limitation studies; materials science and advanced-manufacturing information; energy technology and energy-infrastructure research; transportation, aviation, and infrastructure engineering; computing, communications, sensing, and control systems; health, environmental, and occupational-safety research; standards, measurement, testing, and evaluation; and historical technical baselines.

Currently the NTRL houses nearly 3 million records dating back to the late 1920s and spanning more than 350 subject areas, from public health and industrial medicine to aeronautics. Approximately 2 million of these records were collected by NTIS prior to the widespread adoption of the Internet and related technologies and exist only in non-digital formats, such as

paper print, microform, tape, CDs and DVDs. Due to budgetary constraints, NTIS has been unable to convert many of these non-digital records into digital form. Bibliographic information of all records contained within the NTRL and copies of the reports that exist in digital format are available to the public free of charge at: https://ntrl.ntis.gov/NTRL/.

To advance its mission to make federally-funded STEI more accessible to the public and unlock the maximum potential value of the non-digitized records, NTIS is seeking input from the public on processes and procedures for modernizing the NTRL, with a focus on digitization.

NTIS is also seeking input on the various ways in which the digitized NTRL and its technical data might be leveraged by the private sector and scientific community to advance U.S. innovation, strengthen America's cyber resilience, and spur economic growth.

This insight may inform NTIS' actions on increasing the efficiency, scalability, automation, interoperability, and utility of the NTRL.

Respondents may choose to provide information on the topics below. NTIS has provided the following non-exhaustive list of topics and accompanying questions to guide respondents, and the submission of any relevant information germane to the subject but that is not included in the list of topics below is also encouraged. The inclusion of the topics in this Notice is not intended to indicate a particular relationship between them, nor are they intended to limit the topics that may be addressed by the public. Respondents need not address all questions in this Notice, though all responses should specify which questions are being answered. All relevant responses that comply with the requirements listed in the DATES and ADDRESSES sections of this Notice will be considered.

I. Digitization Best Practices

A. General Logistics

1. When digitizing physical records, what chain-of-custody frameworks or procedures might be utilized? How might the end-to-end movement of physical records be

managed in the travel from pickup, transport to digitization site, digitization, quality verification, and final destruction? What auditing or tracking procedures are helpful in preventing loss, identifying poor digitization, and ensuring regulation-compliant final destruction?

B. Physical Records (paper, microform, magnetic film, other legacy media)

2. With regards to physical records that exist in the form of paper documents, what technologies and processes might be utilized in the digitization of a large-scale collection containing documents of various ages? What factors inform the estimated cost of such an endeavor, and at what speed might digitization occur?

    a. What about in the case of microform (microfilm and microfiche) records? How might these records differ from paper records in terms of cost, throughput, and output quality?

    b. What about for recovering data from legacy ½-inch, 9-track open-reel magnetic tapes? What are the primary risks of data loss or degradation specific to this media type, and how should NTIS prioritize recovery before further deterioration occurs?

3. What preparation steps might be necessary before digitizing paper records of various ages? Are there certain preparation steps that are necessary for documents from a particular era but not others? How might these preparations steps, including deacidification, de-stapling, and flattening, inform a project's overall cost and timeline?

    a. In the case of microfilm, given variations in microfilm quality over decades, what specialized pre-processing or image-enhancement techniques are necessary to ensure high-accuracy OCR text extraction from older or degraded film stock?

4. In dots per square inch (DPI), what image resolution settings are recommended for paper documents, keeping in mind the balance between file size, storage costs, and machine readability? Are preferred image resolution settings different for documents intended for AI training versus human readability?
5. When utilizing Optical Character Recognition (OCR) engines, which document layout analysis tools or approaches have you found most useful, and why? Are preferred tools and approaches different for technical and scientific documents, such as those with mathematical notation, chemical formulas, engineering diagrams, tables, headers, footnotes, and multi-column layouts? What OCR confidence thresholds are reasonable when digitizing paper records that are up to 60 plus years old?
6. What tools and approaches are most useful when digitizing documents containing complex technical diagrams, schematics, chemical structures, or mathematical formulas, particularly with regard to capturing both textual and visual elements? What are examples of the practice of extracting tables and figures from such documents as separate metadata objects, and under what circumstances is this practice being utilized?
7. When digitizing a paper dataset for use in AI training, which digital file formats have been found to be best-suited, and why?
8. Once data is recovered from magnetic tapes, what normalization and conversion steps are typically required to make the content usable within a modern document management or AI accessibility?
9. What key performance benchmarks have proven most useful or are most often overlooked when evaluating vendor proposals for paper digitization?
    a. What about in the case of microform records? What performance benchmarks are most useful to evaluate vendor proposals for microform digitization at

scale?

b. What about for magnetic tape formats? What performance benchmarks or success metrics are most reliable?

10. Beyond paper, microfilm, microfiche, and magnetic tape, are there other legacy media formats NTIS should anticipate encountering in a collection of this age and scope, and how should they be addressed?

II. General Digitization Infrastructure

11. What cloud storage architectures, on-premises solutions, or hybrid approaches should NTIS consider for housing a collection of this size, and what are the tradeoffs in terms of cost, security, access speed, and long-term sustainability?
12. What quality assurance and quality control (QA/QC) frameworks — including automated and human-in-the-loop review — should be applied to ensure the accuracy and completeness of digitized records across all media types?
13. What metadata schemas and standards (e.g., Dublin Core, MARC, MODS, JATS XML) should be applied during digitization to maximize interoperability, discoverability, and long-term preservation value, as well as compatibility with existing AI data pipelines?
14. What strategies can be employed to harmonize legacy datasets created under older metadata standards with newly digitized collections that use more contemporary metadata schemas?
15. What hurdles to success commonly or surprisingly appear in large-scale digitization projects, and how can they be minimized or avoided?

III. Dataset Curation, Structuring & Preparation for Use

16. What data cleaning, normalization, and enrichment steps are necessary to transform a raw digitized corpus of scientific and technical documents into a useful tool for

future scientific and technical applications?

17. When encountering documents with low OCR confidence scores, missing metadata, or damaged and illegible pages, is there value in maintaining such documents as part of the NTRL? What thresholds or remediation strategies are recommended?
18. What strategies might NTIS consider for handling multimodal content (e.g., figures, charts, engineering diagrams, tables, and equations) for both archival purposes and providing the highest utility to the current research community?
19. What deduplication strategies are effective for identifying multiple copies or versions of a document, including copies and versions that may exist in multiple formats (e.g., paper, microfiche, magnetic tape, digital)?
20. Are there recommended automated filtering or screening mechanisms that can be deployed during digitization to ensure that no restricted, export-controlled (e.g., ITAR/EAR), sensitive, or personally identifiable information (PII) is inadvertently exposed, given that the collection consists primarily of public, federally-funded research?
21. What governance structures should NTIS establish to oversee ongoing dataset curation, quality control, update cycles, and stakeholder engagement after initial digitization is complete?
22. How might a fully digitized version of the NTRL be structured and formatted (e.g., JSONL, Parquet, Markdown, raw text with metadata sidecars) to maximize its usability across a range of CETs, including AI training pipelines and frameworks?
23. What approaches to versioning and provenance tracking have proven effective in large-scale digitization projects? When considering the importance of reproducibility in AI research and compliance with emerging AI transparency requirements, do different approaches suggest themselves?

24. What existing mechanisms for distributing and accessing large-scale data repositories (e.g., bulk download via cloud object storage, streaming APIs, vector databases) offer the greatest advantages in terms of efficiency, useability, and interoperability? How can interoperability with existing data repositories be maximized?

IV. Value of a Fully Digitized NTRL for furthering future U.S. Innovation

25. What characteristics of a fully digitized NTRL would be of the greatest value in furthering US innovation?
26. How would a fully digitized NTRL compare in value for the development of CETs, including AI models and tools, to other commonly used scientific text databases? What unique gaps might the NTRL fill?
27. Are there specific technical domains within the NTRL (e.g., aeronautics, public health, materials science, energy, defense) that if more accessible could produce outsized scientific or economic value? Please identify priority domains and explain your reasoning.
28. To what extent could an AI model trained or fine-tuned on the NTRL corpus help researchers, engineers, and scientists surface insights, connections, and findings buried in decades of forgotten or hard-to-search government research — including potentially identifying overlooked scientific breakthroughs or cross-disciplinary connections?

V. Access Models, Licensing & Public Benefit

29. What data distribution and pricing framework (access model)—encompassing technical delivery methods (e.g., bulk cloud downloads, APIs), licensing terms (e.g., open source, commercial), and fee structures (e.g., freemium, cost-recovery fees for high-volume commercial users)—would best balance maximizing public benefit,

supporting U.S. competitiveness, and enabling NTIS to recover costs (or eliminate government costs), while ensuring individual public users and non-commercial researchers retain free access?

30. How should NTIS structure access to ensure that small businesses, academic institutions, and non-profit researchers can meaningfully participate alongside large companies?
31. Are there precedents from other government data release efforts (e.g., NIH data sharing policies, NASA open data initiatives, the Pile dataset) that NTIS should study when designing its access and licensing framework?

VI. Partnerships & Implementation Models

32. Are there existing federal data infrastructure programs, university consortia, or national laboratory networks that NTIS should partner with to accelerate digitization and dataset curation?
33. What measures should NTIS take to ensure that tools, including AI tools built on the NTRL dataset, are accessible and useful to a broad range of users, including those without advanced technical expertise?
34. How should NTIS approach transparency and documentation of the dataset — for example, through dataset datasheets or model cards — to support responsible AI development practices consistent with current federal AI policy?
35. How can NTIS ensure that the NTRL dataset remains current and continues to grow in value over time as new federally-funded STEI is produced?
36. Are there international counterparts or foreign government initiatives to digitize and make available large scientific and technical corpora for AI development that NTIS should study or seek to distinguish the NTRL from in terms of strategic positioning?

Authority: 15 U.S.C. §§ 1151-1157, 3704b, 3704b-1, 3704b-2.

Alicia Chambers,
NIST Executive Secretariat.
[FR Doc. 2026-20023 Filed: 9/29/2026 8:45 am; Publication Date: 9/30/2026]