Office of Scientific and Technical Information

Strategic plan

Fiscal Year 2020-2024

U.S. DEPARTMENT OF ENERGY | Office of Science | Office of Scientific and Technical Information

Office of Scientific and Technical Information (OSTI) Strategic Plan 2020-2024

Table of Contents

Foreword

The Department of Energy invests billions of dollars annually in energy-related research and development (R&D). The majority of this R&D investment is unclassified, and DOE has a statutory obligation to provide public access to its unclassified R&D results. Historically, the primary "result" of R&D has been recorded in scientific papers, such as journal articles, conferences, and technical reports. In today's data-intensive, computer-driven "open science" landscape, other critical components include datasets and scientific software. DOE's Office of Scientific and Technical Information (OSTI) is responsible for collecting, preserving, and providing seamless access to all these research "products." This strategic plan describes how OSTI will fulfill this responsibility from 2020 to 2024. OSTI's mission is to advance science and sustain technological creativity by making R&D findings available and useful to Department of Energy researchers and the public. DOE's longstanding directive for Scientific and Technical Information Management (DOE Order 241.1B or its successor version) sets forth OSTI's responsibilities and mission.

For the classified and sensitive components of its R&D mission, DOE is obligated to manage R&D results in a secure, need-to-know-based environment. While the rules and technologies in this environment are different from unclassified information, the underlying core functions of "collect, preserve, and disseminate" are the same, albeit through highly-controlled, secure systems. This plan describes how OSTI will continue to fulfill these classified and sensitive information responsibilities.

Finally, as a result of the tremendous body of scientific knowledge produced by DOE and the tools and technologies used to manage it, OSTI is positioned to provide specialized services to DOE and other federal agencies, as well as analytical capabilities to accelerate science, measure impacts, and identify trends and opportunities. This plan describes how OSTI will provide value to DOE, the Office of Science, other DOE program offices, and other federal agencies through these services.

Structurally, OSTI's goals, as laid out in this plan, are built around the three core functions essential to fulfilling our mission: 1) building a comprehensive collection of DOE R&D results; 2) permanently preserving the collection; and 3) broadly disseminating unclassified R&D results and providing secure access to classified and sensitive information. The fourth goal focuses on the technology required to deliver on these core functions as well as the specialized tools and services to maximize the value of DOE's information and OSTI capabilities.

Overlaid across the four strategic goals are the following guiding principles and values. 1) The public good. The public invests in energy-related R&D through DOE, and OSTI is committed to contributing to the public good by making the results of that investment useful and accessible. 2) Customer service. OSTI provides information and services to a wide range of individual and institutional customers; we strive to surpass customer expectations in every transaction. 3) Stewardship. OSTI is responsible for a tremendous body of scientific knowledge. We are committed to its integrity and security. We are also committed to the stewardship of the OSTI facility and budgetary resources. 4) Partnering. For OSTI to be successful, it is essential that we forge and maintain intra- and interagency, international, and public-private partnerships and that our partners view us as good-faith, responsive, and contributing counterparts. 5) Workforce excellence. OSTI is committed to the well-being of its workforce and to the development and promotion of mission-essential skills and values. We are also committed to diversity and inclusivity as sources of creativity, flexibility, resilience, civility, and strategic thinking.

Goal 1: Accountability for DOE R&D Results – Collection of Scientific and Technical Information

The comprehensive acquisition of DOE's R&D results, or scientific and technical information (STI), and the advancement of open science are focal points for OSTI over the next five years. OSTI will continue to foster strong relationships within the Scientific and Technical Information Program (STIP) community and will build and strengthen alliances with DOE offices and with other government agencies. In support of open science practices, OSTI will increasingly integrate its digital identifier service at the point of STI collection to support the interlinking of people, papers, software, and data as well as other research objects such as events, awards, instruments, projects, and tools. OSTI's submission systems will be reengineered to fully support these workflows. OSTI's existing E-Link ingest system will be updated and transformed to streamline STI submissions, continue to support the broad STIP community while improving the user experience of awardees, and accommodate new technologies and processes. OSTI will ensure that systems continue to support secure acquisition of DOE-funded classified STI and Unclassified Controlled Nuclear Information (UCNI) and that processes are adapted to meet new Departmental requirements for Controlled Unclassified Information (CUI).

Strategic Objective 1.1

Transform and modernize STI ingest processes and systems

OSTI's STI ingest system, E-Link, was first deployed two decades ago. E-Link has undergone significant periodic upgrades since that time to support STI management requirements and achieve efficiencies; however, new technologies and new workflow processes necessitate a modernization effort over the next five years. E-Link must be able to support new forms of R&D results; facilitate connections between text, data, and software tools, and other research objects; and transform the submission process for enhanced usability. Along with streamlining the submission process, the ability to obtain and enhance metadata via automated means such as text mining, machine learning, and artificial intelligence (AI) will be evaluated. Integration of E-Link with other DOE corporate or programmatic systems, to promote increased STI submissions and programmatic awareness, will be achieved.

Actions:

a) Research and benchmark how other federal agencies are acquiring STI.
b) Modernize the existing E-Link system with current technology and streamline the STI submission process.
c) Support assignment of persistent identifiers and interlinkage of related research objects through the initial STI submission process.
d) Reengineer existing systems to allow associations of grants, awards, instruments, projects, models, tools, and collections to specific STI.
e) Evaluate feasibility of using artificial intelligence, text mining, and machine learning to create or improve metadata at time of submission or during processing.
f) Integrate with DOE corporate or programmatic systems to facilitate STI submission.
g) Improve metadata quality via automated means and more sophisticated validation processes, while accommodating new data fields as warranted to facilitate connectivity of the research community with relevant and comprehensive results.

Strategic Objective 1.2

Target increased comprehensiveness for all forms of STI

OSTI will continue to support requirements set forth in the DOE Public Access Plan as well as support U.S. government open science policies and practices to increase public access to federally-funded scientific research. OSTI will target comprehensive collection of all forms of STI resulting from DOE-funded research, including scholarly publications, research datasets, scientific software, technical reports, conference proceedings, patents, etc. OSTI will work to demonstrate DOE's public access metrics at the federal level and will work with DOE program offices to improve the accuracy of acknowledging DOE funding and to incentivize STI submissions. Potential public-private partnerships to increase the submission of DOE-funded STI from the university community will be considered, along with expansion of partnerships with other federal agencies engaged in jointly-funded research with DOE. In concert with STIP, OSTI will engage records management communities to identify and submit STI from backlogged or legacy collections and will analyze and remedy gaps in STI acquisition practices.

Actions:

a) Coordinate continued inclusion of STI submissions as a measurable component in DOE labs' annual performance measurement plans, providing routine, targeted feedback to labs and site offices.
b) Increase awareness of scope of STI products and requirements through targeted communications.
c) Increase submissions of non-textual research products.
d) Engage Program Office support in incentivizing STI submission and in improving practice of accurately acknowledging DOE funding.
e) Support Awarding Offices' management of new funding models and requirements.
f) Evaluate possible public-private partnerships to acquire STI from university communities.
g) Investigate additional federal agency partnerships for obtaining STI resulting from multi-agency funded research.
h) Work with DOE Labs/Sites on submission of backlogged or legacy collections to improve historical comprehensiveness.
i) Engage records management community at Labs/Sites and determine requirements of site-specific records schedules.

Strategic Objective 1.3

Build a comprehensive collection of DOE classified/UCNI, CUI R&D results

In addition to focusing on comprehensive coverage of all forms of STI resulting from DOE funded R&D, OSTI's responsibilities encompass STI that requires protection in accordance with applicable laws, regulations, and national security requirements. Classified STI and Unclassified Controlled Nuclear Information (UCNI), as well as Controlled Unclassified Information (CUI), are key components in documenting vital research within the Department. Ongoing management of secure ingest systems for metadata and full text continues to be a priority. OSTI will work to expand both the utilization of submission processes and expand the recognition of the importance of a comprehensive collection of current and legacy STI which provides a critical resource for authorized Departmental researchers, now and in the future.

Actions:

a) Increase the submission of classified and UCNI STI in partnership with Classified STI Managers and designated points of contact at individual labs and sites.

b) Establish a current National Nuclear Security Administration (NNSA) Scientific and Technical Information Program (STIP) representative and other programmatic representatives.
c) Work to support performance measurement processes and provide Performance Evaluation and Measurement Plan (PEMP) feedback on labs' classified, UCNI, and CUI STI performance to NNSA.
d) Identify and coordinate with existing STI-related preservation efforts underway within NNSA.
e) Continue modernizing the existing Classified E-Link system with current technology to streamline the STI submission and validation process, while accommodating new data fields as warranted.
f) Evaluate feasibility of using artificial intelligence, text mining, and machine learning to create or improve metadata at time of classified submission or during processing.
g) Participate in ongoing Departmental CUI related working groups in accordance with Departmental efforts required by Executive Order 13556, Controlled Unclassified Information.
h) Work to ensure OSTI is positioned to appropriately ingest and manage STI identified as CUI, in accordance with Departmental directives and guidance.
i) Research and benchmark classified and CUI STI acquisition methods across other federal agencies.

Goal 2: Permanently Preserve DOE's Scientific Record

A prerequisite to providing permanent access to DOE's R&D results is ensuring preservation of this collection, dating from the Manhattan Project to the present. Much of the early collection remains in paper media, and OSTI will convert and continually migrate STI to the most modern digital formats and will maximize public access to digitized content and secure access to classified/UCNI and sensitive content. Since 2000, all STI collected by OSTI is "born digital," and along with digitized content, OSTI will ensure the protection, integrity, and backup redundancy of this comprehensive collection of DOE R&D results.

Strategic Objective 2.1

Accelerate preservation of legacy scientific collection and strategically manage and preserve "born digital" STI

OSTI maintains a historic physical collection of DOE's R&D results dating back to the Manhattan Project. Currently, the un-digitized component of this full-text STI content is inaccessible to the public and at risk of degradation as aging of the media continues. To address this risk and inaccessibility, OSTI continuously digitizes physical documents. OSTI will leverage emerging technology and AI techniques to lead the Department as a preservation and curation center of excellence.

By innovating modern digitization pipelines and capitalizing on expanding expertise from OSTI's development and records management workforce, OSTI will strategically reduce the time and complexity of converting a physical document to a digitally accessible product, which has been curated, tagged, and reviewed to ensure it can be released. Specifically, technologies such as AI-powered quality assurance processes near the beginning of the digital preservation workflow will allow OSTI to prioritize the preservation process of the at-risk documents, while also providing additional enrichment of the final available product.

Finally, all contemporary STI is "born digital," and OSTI will continuously migrate this content to the latest standards and formats to enable accurate indexing and permanent preservation. OSTI will deploy a mix of cloud, on premise, and off-site storage, backup, retrieval, and disaster recovery strategies.

Actions:

a) Complete implementation of AI-based quality control in the preservation pipeline by FY21. Reduce labor-intensive QC processes for digitized documents by 90% and reinvest efficiencies to increase rate of document digitization.
b) Obtain required labor and technological infrastructure to complete digital preservation of OSTI's ~580,000 legacy paper-based records by FY25 to reduce risk of information/data loss. OSTI's largest focus will be the immediate digitization of legacy records before applying additional enrichment techniques.
c) Demonstrate OSTI's commitment to meeting the OMB 19-21 requirement to move to all government records in digital formats before 2022 by developing and presenting an updated OSTI Records Management Plan for approval by Q4 2020. This plan will drive the restructuring of OSTI's internal file server by managing cross-organizational cleanup of current records during FY21 and will identify collections of paper administrative records which will be prioritized for digitization.
d) Adopt and implement strategies for the most cost-effective and operationally-conducive mix of cloud and physical solutions for backup, retrieval, and disaster recovery for the entire electronic STI collection.

Strategic Objective 2.2

Expand intake and preservation services for new open science research outputs (data, source code, etc.)

Evolution in modern science has created new research outputs, particularly numeric data and scientific software. OSTI recognizes the value of these outputs and will widen its intake and curation processes to preserve and present them to future generations of scientists as well as to the public. OSTI will ensure that these outputs are uniformly categorizable, citable, and reproducible as they are made available in an enriched form. These enrichments will include both technical and administrative procedures, while remaining transparent on processes involved, to ensure integrity of the open science data made available.

Actions**:**

a) Consolidate the three supporting applications which manage the document metadata and sensitivity review and release process as described in Strategic Objective 2.3 into a single streamlined application to support a more efficient and expedited review and accessibility of sensitive records for cleared researchers by FY23.
b) Partner with labs and the Office of Science to identify additional scientific data artifact categories requiring preservation and access workflows by FY25. Knowledge of and engagement with these stakeholders will continue to broaden OSTI's open science portfolio and will ultimately provide researchers an enriched set of usable, high-quality preserved R&D outputs to advance science.
c) With stakeholder input, develop a data preservation and access service, addressing long-term preservation of data for which OSTI is the optimal solution.
d) Expand archiving of DOE-funded scientific software and ensure version-level preservation.

Strategic Objective 2.3

Enhance infrastructure for classified, UCNI, and CUI research to deliver a reliable and regulatory-compliant preservation and delivery network

OSTI's mission to make STI widely available to scientific researchers and the public presents a unique challenge when the information has access limitations. To confirm the external party's need-to-know (or approved requisite access) before providing information, OSTI relies heavily on a network of technical and administrative processes to deliver information securely only to the required users. This review/release network will be improved and optimized in the coming years by applying technical advances and increased

human capital efforts.

Actions**:**

a) Add subject matter expertise to OSTI's sensitivity review process by FY22 to maximize public access to unclassified content and secure access to classified and sensitive content.
b) During FY23, increase technical support for the classified information system to provide increased depth and security.

Goal 3: Broad Discovery, Use, and Visibility of DOE R&D Results

OSTI will work to maximize the use and further increase discovery of DOE R&D results to support DOE goals and objectives to ensure advancement of America's leadership in science and energy. OSTI will provide a leading-edge search experience by linking together research objects, encompassing awards, organizations, instruments, equipment, collections, and other objects throughout the research lifecycle. As a result, a more complete picture of research is presented. OSTI search tools will offer users new ways to interact with research outputs and search results, and, through the implementation of artificial intelligence, offer improved and contextualized searching. By leveraging current and emerging partnerships with industry leaders in digital identification and interlinking of research objects, OSTI will showcase its leadership in the open science community with innovative approaches for increasing access to DOE-funded R&D results.

Strategic Objective 3.1

Increase role in open science through persistent identifier services

The concept of open science is making research objects openly accessible to provide more transparency, reusability, and reproducibility to scientific findings. Making DOE-funded science more open has the potential to increase the pace of scientific discovery, promote more efficient and effective use of government resources, and protect the integrity of science by facilitating validation of results. Through OSTI's implementation of the DOE Public Access Plan, development of DOE CODE, interlinking of research results, and persistent identifier services, OSTI has become an open science leader in the U.S. government and internationally. OSTI's current persistent identifier services provide Crossref and DataCite digital object identifiers (DOIs) to text-based documents, datasets, instruments, and software and provide ORCID iD services through the U.S. Government ORCID Consortium led by OSTI. OSTI's persistent identifier services allow for increased visibility of DOE research outputs and interlinking of research objects throughout the research lifecycle.

Actions:

a) Continue OSTI's leadership in the use of persistent identifiers, offering services assigning DataCite and Crossref DOIs to text-based documents, datasets, instruments, and software.
b) Grow the usage of our persistent identifier services with researchers, new DOE funding offices, user facilities, and other federal agencies.
c) Evaluate current DOI services for potential updates and functionality useful to the user community and implement beneficial changes.
d) Evaluate expansion of persistent identifier services to other federal agencies.
e) Evaluate and deploy DOI assignment to other types of research outputs.
f) Implement new service for assigning Crossref DOIs to conference posters and presentations.

g) Implement new Award DOI Service assigning Crossref Grant ID DOIs to awards. The service will be piloted with the Environmental Molecular Sciences Laboratory (EMSL) and Joint Genome Institute (JGI) user facilities, assigning DOIs to awarded proposals at their facilities.
h) Increase membership in the U.S. Government ORCID Consortium to promote broad integration with ORCID Application Programming Interface (API) services.
i) Evaluate and implement the use of Research Organization Registry (ROR) IDs as persistent identifiers for organizations.
j) Support additional linkages of related research objects throughout the research lifecycle, connecting funding to researchers to organizations to research outputs, through the use of persistent identifiers.
k) Track and evaluate emerging persistent identifier services for potential implementation – including IGSNs (persistent identifiers for samples) and RRIDs (persistent identifiers for antibodies, cells, organisms, and tools).
l) Establish a Data ID Service Community of Practice, hosting workshops and communications forums.
m) Evaluate the need for establishing a broad persistent identifier community of practice.

Strategic Objective 3.2

Provide and Develop Leading-Edge Discovery Tools

OSTI is committed to providing and developing leading-edge search tools enabling broad discovery of DOE-funded R&D results. OSTI will continue to engage our users and identify opportunities to add innovative functionality to effectively disseminate R&D results. All the OSTI search tools have been updated with a consistent design to show cohesiveness and branding as DOE OSTI search tools. With the redesigns, functionality was added to streamline discovery and information navigation for users. Functionality includes the addition of related identifiers and persistent identifiers, interlinking research objects. OSTI is working to identify, connect, and add many forms of related research objects in the search tools – connecting funding to researchers to organizations to research results through the use of persistent identifiers. Achieving these linkages and presenting them in a structured format to users of OSTI tools improves the accessibility, transparency, and reproducibility of federal research. OSTI has also implemented new functionality providing figures and tables extracted from journal article accepted manuscripts. As users look for visual representations of information, OSTI.GOV offers search results of the extracted images and reference and citation traversal. OSTI is committed to adding further functionality and implementing user feedback to guide development and changes to OSTI search tools.

OSTI search tools include:

- **OSTI.GOV**, the primary search tool for DOE science, technology, and engineering research and development results and the organizational hub for information about OSTI.
- **Department of Energy Public Access Gateway for Energy and Science (DOE PAGES)**, the search tool that makes peer-reviewed scholarly scientific publications resulting from DOE research funding publicly accessible to read, download, and analyze.
- **DOE CODE**, the DOE software service platform and search tool for DOE-funded code. It serves as the DOE software catalog with a comprehensive inventory of DOE-funded custom developed scientific and business software that is provided to the government-wide Code.gov inventory.
- **DOE Data Explorer (DDE)**, the search tool for discovering DOE-funded scientific data objects submitted by data centers, repositories, and other organizations funded by the DOE.
- **DOE ScienceCinema,** a collection of videos produced by the DOE national laboratories, other DOE research facilities, and the European Organization for Nuclear Research (CERN). Using innovative, state-of-the-art audio indexing and speech recognition technology from IBM Watson Bluemix, DOE

ScienceCinema allows users to search for specific words and phrases spoken within video files to deliver precision searching already common in text-based databases.

- **DOE Patents**, the comprehensive search tool for DOE-funded patent information. OSTI uses USPTO APIs to automate the retrieval of DOE-funded patent information to streamline patent ingest and curation.
- **Science Research Connection**, search tool exclusively for the DOE community that provides research information integrated from various OSTI databases, including both unclassified/unlimited and statutorily controlled information.
- **Science.gov**, a federated search engine maintained and stewarded by OSTI, offering a single search interface across 13 federal science agencies.
- **World Wide Science**, a global federated search engine maintained and stewarded by OSTI, comprised of national and international scientific databases and portals.

OSTI's communications program focuses on expanding awareness of OSTI products and services to key stakeholders, leading to increased usage of DOE-funded R&D results. This program emphasizes the wide range and usefulness of DOE R&D results. OSTI communication efforts employ tools such as OSTI.GOV News, news round-up emails, videos to promote usage of and instruct users about DOE OSTI discovery tools, and social media. Through communications, workshops, conferences, and other meetings, OSTI works to bolster outreach to key stakeholders including DOE-affiliated researchers; DOE program, field, site, and procurement offices, national laboratories, and research facilities; research universities and libraries; scientific professional societies; other federal science agencies and international science organizations; and other STI community partners.

Actions:

a) Explore, evaluate, and implement search tool best practices, new functionalities, and cutting-edge technologies in OSTI search tools.
b) Through metrics and analysis efforts, better understand OSTI users and develop targeted ways to reach and engage with various user communities.
c) Evaluate new and increased communications efforts to support increased visibility of DOE R&D results, provide customer service, support news announcements, and further OSTI's mission to advance science and sustain technology.
d) Explore and evaluate opportunities to use artificial intelligence to provide search tool enhancements, including returned search results, addition of subject keywords, automatic recognition of abbreviated or varied metadata, and connections/relationships between R&D results.
e) Continue work of extracting images, figures, and tables to provide additional STI record features and search capabilities.
f) Develop and implement visual representations of information provided in OSTI search tools. Specific exploration will include visualization of related identifiers and persistent identifier relationships, and search results and filtering.
g) With the deployment of the new Award DOI Service, provide the award DOI landing pages. As the service moves from pilot to production, OSTI will develop and implement a new award search tool.
h) Assess and implement new functionality in DOE CODE, including potentially adding software with access restrictions (CUI, encompassing Official Use Only (OUO) and Export Controlled Information (ECI) software); explore support of provisioning and execution of code; and adding other community driven features.
i) Evaluate the project, collection, and dataset structure within DDE and transition to a less hierarchical related identifier-based model.
j) Investigate adding DOE-funded datasets with DOIs assigned by other entities into the DDE collection.

Strategic Objective 3.3

Provide Quality Collection Curation

A key aspect of providing leading-edge discovery tools for any collection depends on the quality of its content, including metadata. OSTI recognizes searching and finding useful research results is based on quality metadata. A key component of OSTI's metadata strategy is the addition and curation of related identifiers, such as URLs or DOIs that serve as links to related research objects (e.g., journal articles, datasets, software, patents). Combining high-quality metadata with innovative technologies such as AI to improve curation efforts, OSTI will provide more targeted and comprehensive searching for OSTI users.

Actions:

a) Prioritize accepted manuscript curation to provide users with quick access to manuscripts and allow for evaluation of public access performance across DOE.
b) Evaluate current curation processes to streamline and implement efficiencies.
c) Continue to curate metadata for new STI records submitted to OSTI.
d) Continue to curate metadata for STI records previously submitted to OSTI that did not initially undergo high-quality metadata review.
e) Explore and develop processes and technologies for curation of new metadata types.
f) Explore and evaluate options for automated curation of STI and related identifier metadata, including using artificial intelligence, to improve metadata quality and allow for manual curation effort to focus on other needs.
g) Evaluate and implement options for increasing efficiencies in the image extraction and curation processes.

Strategic Objective 3.4

Provide Secure, Need-to-Know-Based Search and Retrieval for Classified and UCNI STI

In support of the Department's strategic national security objectives, OSTI maintains an extensive and expanding collection of classified and UCNI metadata, in addition to a collection of full-text classified R&D reports. OSTI is committed to protecting and securely sharing this collection which has been a longstanding resource within DOE and NNSA. In support of this commitment, and in coordination with Departmental stakeholders, OSTI developed a metadata search functionality, the Classified STI Metadata Search (CSMS), where full text can be obtained from the originating site by those with the proper clearance and need-to-know. For full-text reports in the OSTI collection, OSTI coordinates with authorized requestors in accordance with established guidelines. Secure access and protection of information are the cornerstones of this need-to-know-based search product for classified and UCNI STI metadata. OSTI will maintain and continue to develop functionality for the need-to-know-based search product, including the incorporation of formal application development and quality assurance processes into product development.

Actions:

a) Further develop processes and procedures for continuous monitoring of application health and security based on well-defined standards.
b) Develop the Complex Classified Document Request Tracking System, an application that OSTI and participating Enterprise Secure Network (ESN) sites could use to track and update the status of each document on their site, incorporating stakeholder input in the development process as feasible to enhance utility.

c) Add functionality for users to request multiple documents in one transaction and for those requests to appear in the Complex Classified Document Request Tracking System and incorporate capability of requesting users to see the status of individual documents.
d) Develop the Complex Hierarchical Need to Know (NTK) Code System, incorporating enhanced functionality to manage NTK hierarchical data.

Goal 4: Provide Core Technologies, Expertise, and Specialized Tools and Services to Maximize the Value and Impact of DOE's R&D and Information Capabilities

Technology undergirds and enables OSTI's core functions to collect, preserve, and disseminate DOE's R&D results. This includes both the applications development process that creates and maintains OSTI's ingest and dissemination products and the infrastructure that stores, hosts, preserves, and protects the information and products, in addition to accommodating the millions of incoming and outgoing transactions of STI content. OSTI will commit significant resources to leveraging the power of machine learning and AI to improve mission performance. As a key component of the Office of Science, OSTI will contribute to SC's mission through specialized services and support and will provide services and expertise to other DOE offices, other federal agencies and interagency bodies, and international and public-private partners.

Strategic Objective 4.1

Provide a leading-edge technology infrastructure to enable responsive and secure collection, preservation, and dissemination of DOE's R&D results

OSTI's multi-terabyte data collection encompasses numerous formats and categories of scientific and technical information which continue to expand as new artifacts of research are recognized. This expanding STI landscape requires an equally evolving technical infrastructure to continue to provide best-in-class data access to scientific researchers and the scientific public in a reliable and secure manner. OSTI will use emerging technologies to improve workflows while minimizing the risks of an open science environment.

OSTI supports rapid application deployment pipelines by investing in thin-client terminals to replace current developer workstations with more scalable virtual endpoints. In addition to providing increased energy efficiency, these clients lower security risk to OSTI's application development pipeline while simultaneously reducing organizational budget overhead. For externally-facing assets, OSTI's infrastructure and development staff will migrate existing workflows to a container-based environment. Containers are a highly scalable and portable virtual environment allowing applications to respond in real-time to demand increases without requiring administrator intervention. They also use improved security controls to ensure integrity of scientific data being delivered while remaining transparent to the end user. Finally, containers provide OSTI with a completely resilient platform capable of migrating offsite in hours (rather than days) to ensure ultimate corpus presentation survivability in disaster or contingency scenarios.

Actions:

a) By FY24, complete technical evaluation and begin lifecycle acquisition of thin-client terminals, supporting infrastructure upgrades and software licenses necessary to replace all existing

workstations, thereby increasing security of the OSTI development pipeline and removing OSTI's reliance on and organizational budgetary expense of refreshing physical desktop workstations.

b) Complete installation of the DEVLAN development network during FY21 to allow OSTI's development workforce to begin migrating applications to virtualized distributed containers. No later than FY22, produce a fully functional containerized production environment ready to respond to the advancing needs of scientific researchers.
c) Develop an OSTI Technical Vision 2025 Roadmap in FY2021 to map organizational priorities and projects to upcoming IT initiatives and allow for strategic alignment of purchases and technological advances to increase the quality of OSTI's search products.
d) Continue coordination with ESnet and expand network bandwidth capabilities to support increasing demands due to data preservation services and increased use of external OSTI services. Update capabilities within OSTI network to align with provisioned ESnet capabilities to enable seamless access to curated science data with minimal speed and capacity bottlenecks to ensure optimal access.

Strategic Objective 4.2

Bolster Product Development Processes

OSTI is committed to implementing process improvements in the development and operation of OSTI products and STI workflows. Core functions and features of OSTI ingest technologies and underlying micro-services will be refactored to achieve overarching improvements to the technical architecture and subsequent customer experience, both internally and externally to OSTI. OSTI will continue to evaluate and upgrade the technology stack, as well as evaluate the technical implementation of third-party interactions, such as with Crossref/CHORUS and DataCite, to improve efficiencies. Quality assurance during the application development process is vital to providing leading-edge products. We will continue to emphasize a team approach to quality assurance in agile application development with defined processes that include requirements gathering and analyses, planning, security and in-depth performance testing, analysis of usage data, and documentation. This ongoing effort will include a continual assessment of utilized libraries, dependencies, and potential vulnerabilities and resolution strategies.

Actions:

a) Reengineer multiple existing backend systems including middle-tier code, ingest APIs, framework structures, database tables, and media processing to allow more efficient and less time-intensive submittal workflows.
b) Implement real-time release processing of submitted STI including real-time error reporting.
c) Implement a series of shared libraries for interacting with third-party services such as Crossref, DataCite, and Web of Science in order to increase efficiency across the corpus of mission and cost-recovery applications.
d) Further develop processes and procedures for continuous monitoring of application health and security based on well-defined standards.
e) Continue to standardize and modernize all instances of our indexes with the latest version of Solr. This long-term project is performed in a phased approach to align all products (discovery and ingest). Once all products have been updated to Solr 8, additional search facets and API tools will be leveraged to support findability and improve overall performance.
f) Improve and formalize process for continual assessment of utilized libraries, dependencies, and potential vulnerabilities and resolution strategies.

Strategic Objective 4.3

More Broadly Deploy Artificial Intelligence

As OSTI broadens the landscape of STI beyond publications to data, software, and other research objects, traditional human-performed functions become implausible with limited resources. OSTI will both develop internal AI resources and capabilities and partner with DOE labs and the private sector to address challenges.

Actions:

a) Use container technology to fully utilize resources that would otherwise sit idle in OSTI's virtual infrastructure for resource-intensive AI processing.
b) Procure dedicated AI hardware, including graphical processing units.
c) Leverage artificial intelligence and machine learning in search innovations (e.g., develop semantic search capability to interlink related research objects by meaning in addition to keywords).
d) Leverage AI to make Export Controlled Information (ECI) determinations a faster, more efficient and less expensive process.
e) Leverage AI capabilities for system and applications cyber security.

Strategic Objective 4.4

Support Office of Science Mission

As a component of the Office of Science (SC), OSTI will leverage information management skills and web-based capabilities to advance the SC mission. Specifically, this support will focus on the corporate needs of SC, including senior SC leadership and program managers.

Actions:

a) Manage the bulk of SC's public-facing web presence through the SC Content Repository.
b) Contribute to the development of SC's portfolio analysis and evaluation capabilities, and perform analyses using OSTI tools and information collections.
c) Develop solutions for supporting SC program managers' ability to measure and evaluate programmatic outputs.
d) Manage SC electronic Field Work Proposals, providing front-end submission and back-end search functionality.

Strategic Objective 4.5

Provide Services and Expertise to DOE, Federal, and International Partners

DOE, federal, and international partners can benefit from OSTI's information collections and expertise in meeting their own programmatic needs. From a policy and leadership perspective, OSTI will contribute to national and international stakeholder communities to shape implementation of open science practices. On a cost-recovery or cost-sharing basis, OSTI will also provide specialized products and innovative services and tools for information management and access.

Actions:

a) Leverage OSTI's classified R&D information collection and capabilities to provide specialized products to the DOE/NNSA classified R&D community.

b) Contribute DOE's knowledge and best practices to national and international open science communities and venues, including government-to-government and public-private activities.
c) Develop innovative federated search functionality to enable simplified access to distributed federal and international STI collections through tools such as Science.gov and WorldWideScience.org.
d) Partner with other federal agencies in supporting their public access models, including the front-end ingest of research objects and back-end search functionality.