U.S. Department of Energy

Advanced Scientific Computing Advisory Committee (ASCAC)

Subcommittee on Scientific and Technical Information

First Report, September 4th, 2015

Executive Summary

Introduction

The new Office of Scientific and Technical Information (OSTI) standing subcommittee of ASCAC was charged to assess the quality and effectiveness of OSTI's recent and current products and services and to comment on its mission and future directions in the rapidly changing environment for scientific publication and data.

Overview of Responses to the Charge Questions

Charge question (a): Are the current OSTI products and services best in class and are they the most critical for the OSTI mission given the present constrained budget environment?

The subcommittee examined the vision and mission of OSTI and welcomed its recent re-organization and its new strategic plan. We find that OSTI's products and services are professional and generally well done, and the subcommittee welcomed their increased focus on the search and discovery of DOE STI R&D.

OSTI is playing an important role among government agencies, for example by operating the flagship Science.gov site for DOE and by providing key data ID services through the DataCite organization.

Many of OSTI's services incorporate leading-edge technologies and, in this sense, can be regarded as 'best in class'. These include: SciTech Connect's powerful semantic search capability, the ScienceCinema video service's automated audio-indexing technology, and WorldWideScience.org's language translation technology. OSTI also offers an innovative federated search capability in NLE^{Beta}, Science.gov, and WorldWideScience.org. OSTI's rapid development of the DOE PAGES^{Beta} service and the associated further development of their E-Link service for submission have been impressive.

One OSTI service that was found <u>not</u> to be best in class is the ESTSC software service. This service and its software inventory seem very outdated and out of touch with the leading DOE research software developers. The ESTSC model of charging for software also seems unlikely to be an attractive offer given the open source culture of the scientific research community. In addition, some of the datasets stored in the Data Explorer tool prior to OSTI's establishing their Data ID service need to be reviewed.

Overall, the subcommittee found that OSTI has made great strides recently and most of their latest product releases appear to have elements that are best in class. However, a focus on unifying the product set, reducing redundancy, and improvements in content coverage will be required before they can claim to be fully best in class.

Charge question (b): Do OSTI products and services fulfill customer needs now?

We first note that OSTI has multiple kinds of 'customers' including librarians and public and commercial information services, DOE data program managers, as well as the DOE researchers called out in its mission statement.

There appears to be significant take-up of OSTI services by the public and by commercial services. OSTI provided comparative evidence for the quality of its services.

Certain classes of customers (such as major DOE data program managers in ARM and CDIAC) are satisfied with existing OSTI products and services and see OSTI as one of the few solutions (and the only DOE supported one) for their needs. They also see OSTI as being open to tailoring and improving their services based on their needs.

To researchers, OSTI services seem cumbersome by comparison to existing domain-specific solutions and to leading community archives and public sharing sites, which offer very different and more interactive ways of sharing knowledge.

Researchers see the need for more integration of the different services and an improved user interface. Using user-interface simplification and best-practices to unify and reduce redundancy in the toolset would improve the user's experience of OSTI services.

Charge question (c): Are the OSTI products and services positioned to evolve to fulfill customer needs in the future? Has the OSTI strategic plan appropriately addressed the rapid evolution of technologies, research product types, and ways in which research results are communicated and shared?

The emerging challenge of collecting electronic versions of graphs, tables, and images in papers does not seem to be currently addressed in detail in OSTI's plans. Although data, software, images and video are all part of OSTI's larger data vision it must be emphasized that OSTI and the DOE STIP community must work closely with the DOE research community if they are to develop new useful services for today's researchers. A detailed implementation plan showing how OSTI will achieve the goals set out in their strategic plan would be helpful.

OSTI staff showed a good awareness of the likely evolution of the services with respect to linking publications to data. OSTI's Data ID Service is a useful start and OSTI's involvement with other organizations working in this area will be valuable.

OSTI's vision for providing '*named user'* functionality could clarify the issue of which part of their customer base they are addressing. Adding some *'social functionality'* could also help bring the interactivity of OSTI services up to the level of the best research community tools.

Charge question (d): What is the national and international standing of OSTI with respect to similar organizations whether at other U.S. Federal Agencies, DOE Laboratories, or universities? In what areas must OSTI be a clear leader to fulfill its mandated responsibilities to the DOE?

OSTI is in a leadership position among Federal agencies. OSTI services employ a range of innovative technologies not uniformly available from their peer international scientific information organizations.

OSTI has a leadership role with the CENDI interagency group in operating the Science.gov gateway to government science information. This offers a federated search service across 60 scientific databases and 200 million pages of scientific information. The recently developed DOE PAGES^{Beta} service for access to research journal articles has won the respect of the NSF and the DOD (DTIC) who are considering using the OSTI system for delivering their public access plans. This is an area in which OSTI must be a clear leader to fulfill its mandated responsibilities.

In terms of international leadership and recognition, OSTI is a founder member of the WorldWideScience Alliance and is responsible for providing novel real-time searching and translation service over globally-dispersed multilingual scientific literature to the other Alliance members.

OSTI products and services compare well with those delivered by similar organizations in Canada (CISTI), France (INIST-CNRS) and the German National Library of Science and Technology (TIB).

Because of its recognized expertise in information management, OSTI was invited to chair the Technical Activities Coordinating Committee of the International Council for Scientific and Technical Information (ICSTI).

Overview of Responses to supplementary questions

Is the mission statement sensible in the light of the statutory authorities?

The OSTI mission statement is entirely appropriate in targeting DOE researchers and the public.

Is OSTI organized and staffed to accomplish today's mission?

The recent re-organization of OSTI in terms of its three core functions has given OSTI clearer focus on DOE research results.

It is likely that some changes to the mix of technical expertise at OSTI will be required to design and develop services suitable for modern science environments. This could be acquired either through new hires or by collaboration with existing DOE lab researchers and librarians.

If OSTI is to take on a larger role with respect to data it needs to expand its expertise in this area.

Are the current and planned OSTI products and services the correct ones?

The products and services need to be targeted for at least three different communities – the traditional library and information management community, the DOE research community, and the general public.

The Data ID Service is a critical first step towards making datasets citable and linking data to publications. Other data collection, federating, and brokering services may be the next step.

The automated collection of publications and provision of public access versions should remain a top priority for OSTI. Optimization of the publication collection method could significantly reduce the burden on the data submitter (e.g. requesting only the DOI, organization, funding info, and a pdf) and significantly increase collection completeness.

The start on collecting multimedia content is valuable but improvements in metadata and consideration of the priorities of different types of multimedia should be undertaken.

What suggestions would the subcommittee make for the next steps?

We suggest that OSTI undertake more to understand and collaborate with its various customers, including:

Discuss approaches to **partner** with the DOE labs and researchers to improve content completeness and help reach the DOE goal. Initiate some serious two-way outreach and dialog with the DOE Labs research communities to better understand what services they would like and use. Discuss tool usability issues with the DOE research community with a view to developing an integrated 'one stop shop' approach to STI services;

Enlarge the STIP management by 'researcher champions' from each Lab. The lab library staff need to work with researchers to understand the issues of research reproducibility and open science that require linking data and software to research publications. With the STIP management, OSTI needs to develop the necessary skills to advise researchers about the required Data Management Plans. This could include discussion about possible data repositories for long-term storage of large data sets as well as how to release sufficient data to support the conclusions of the journal article.

A more detailed analysis of Google/Bing search results on DOE R&D could help determine those areas OSTI should focus to deliver complementary functionality.

Subcommittee Recommendations

To OSTI:

1. If OSTI is to truly fulfill its mission to create products and services to make '*R&D findings available and useful to DOE researchers'*, it needs to initiate a vigorous outreach program with the DOE Lab researchers. This must involve listening to researchers needs and understanding the strong and weak points of existing community sites.

2. OSTI should work with the DOE research community to re-invent the ESTSC software service. In addition, releasing software in support of a research publication needs to be supported. In respect of the Data Explorer tool, OSTI should undertake a review of the datasets currently included in the tool, in partnership with Lab researchers.

3. Work with the Labs to identify 'researcher champions' who can work with the STIP community to strengthen the link to researchers. This could include advice on Data Management Plans and target data repositories.

4. OSTI should work aggressively to continue toward a unified user environment with a limited number of, clearly delineated, non-redundant tools and develop a master plan for future development and areas of expansion through community input.

5. Through partnership with the national lab librarians and researchers identify and address publication content gaps and develop clear instructions and guidelines regarding content submission requirements. Significantly improving the completeness of coverage of the publications collection will require creative solutions. In addition, OSTI needs to work with the Labs to correct present inaccuracies in the record and also to devise practical methods to check that research papers are made openly available after the 12 month embargo time.

To the Office of Science:

1. To promote a successful implementation of the public access requirement issued by OSTP, OSTI needs top-down support from DOE in clearly communicating that this is not a requirement/burden imposed by OSTI but rather a government-wide and DOE-wide requirement meant to share federal research results and accelerate scientific progress. In this regard, labs, grantees, and their authors need to be incentivized to comply with this requirement, which partnership with OSTI staff can help them to fulfill, and one such incentive could be a measurable expectation expressed in labs' annual performance plans.

2. The Office of Science should consider defining a useful role for OSTI and the STIP management team in managing DOE data. Sharing and preservation of data are central to protecting the integrity of science, facilitating validation of results, and advancing science by broadening the value of research data to disciplines other than the originating one and to society at large. Possible roles for OSTI include:

    - Following the example of major journals and collecting digital versions of tables, graphs, and images from papers.
    - Working with all of the Office of Science Programs and the different research communities in the DOE labs to develop better solutions for linking data and software to publications.
    - Coordinating reviews of the data needs by discipline to identify explicit commonalities and differences between disciplines.
    - Participating in collaborative pilots that establish the open data and open science end-to-end infrastructures (data provenance, data workflows, experiment integration).
    - Assisting in the development of an evaluation plan to assess how well the DMP and OSTI services support the community.
    - Developing cost models for manageable and cost-effective data solutions.

However, if OSTI is to play a major role in data management, they would need to acquire significant new expertise in research data and would need to be resourced appropriately. It is also essential that OSTI collaborate closely with the DOE research community from the beginning.

1. Historical Background

The origin of the DOE's Office of Scientific and Technical Information (OSTI) can be traced back to Vannevar Bush's 1945 seminal report *'Science: The Endless Frontier.'* President Roosevelt had asked Bush to apply the experience of the war-time scientific R&D efforts, which were mostly still secret, to the '*days of peace ahead.*' In particular, Roosevelt had asked *' … what can be done, consistent with military security and with the prior approval of the military authorities, to make known to the world as soon as possible the contributions which have been made during our war effort to scientific knowledge. The diffusion of such knowledge … should help us stimulate new enterprises, provide jobs for our returning servicemen and other workers, and make possible great strides for the improvement of our national well-being.'* Bush's report recommended that the R&D results should be made openly available, consistent with national security requirements.

The Atomic Energy Acts of 1946 (P.L. 79-585) and 1954, as amended (P.L. 83-703), established a program for the dissemination of unclassified scientific and technical information and for the control of classified information (42 U.S.C. Sec. 2013, 2051, and 2161). With this mandate, in 1947 General Groves, the war-time leader of the Manhattan project, set up the organization that later became OSTI with the mission to make non-classified scientific R&D openly available. The responsibility of the OSTI organization was explicitly called out in the Energy Policy Act of 2005 (P.L. 109-58), Section 982: *'The Secretary, through the Office of Scientific and Technical Information, shall maintain within the Department publicly available collections of scientific and technical information resulting from research, development, demonstration, and commercial applications activities supported by the Department.'*

2. Context for this Report

The America COMPETES Acts of 2007 (P.L. 110-69) and 2010 (P.L. 111-358) required that Federal agencies that conduct scientific research develop agency-specific policies and procedures regarding the public release of data and results of research. On February 22, 2013, the White House Office of Science and Technology Policy (OSTP) issued a memorandum to Heads of Executive Departments and Agencies, titled '*Increasing Access to the Results of Federally Funded Scientific Research*' (). This required all the major federal R&D funding agencies to develop a plan to support increased public access to the results of their research. The OSTP memo specifically defined research results to include both peer-reviewed publications and digital data.

https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

For refereed publications, a 12-month post-publication embargo period was to be used as a '*guideline'* for making research papers publicly available. In the DOE's Public Access Plan approved by OSTP and the Office of Management and Budget (OMB), OSTI is 'responsible for operating and maintaining DOE's public access system and network' for scientific publications. In addition, the Plan states that the '*submission of accepted manuscripts and publication metadata to DOE will be a condition of funding'* and that the DOE will be responsible for ensuring compliance.

http://www.energy.gov/sites/prod/files/2014/08/f18/DOE_Public_Access%20Plan_FINAL.pdf

In the OSTP memo, digital data is defined as *'the digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings including data sets used to support scholarly publications …'* As regards public access to digital data, the DOE Plan requires an agency-wide policy for data management planning to be agreed by October 1st 2015. In addition, all research proposals selected for funding will be required to have a Data Management Plan (DMP). These DMPs will provide details '*for making all research data displayed in publications resulting from the proposed research open, machine-readable, and digitally accessible to the public at the time of publication. … Individual research offices will encourage researchers to deposit data in existing community or institutional repositories or to submit these data to the article publisher as supplemental information.*' The DOE plan also acknowledges that in some instances additional steps may be needed to collect sufficient research data to meet the requirements of the OMB memo on Open Data Policy.

https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf

The DOE commits to consulting with the relevant research communities via a range of public forums. In an explicit reference to OSTI with regard to research data, the Public Access Plan states that '*OSTI can provide digital object identifiers (DOIs) to data sets resulting from DOE-funded research.*' The Plan explicitly encourages the use of such identifiers to enhance the discoverability and attribution of these data sets.

3. OSTI Vision and Mission

The OSTI Vision and Mission are detailed in OSTI's *2015-2019 Strategic Plan*, available from the OSTI website at http://www.osti.gov/home/about.

The OSTI vision is as follows:

> *The Office of Scientific and Technical Information will fulfill a critical U.S. Department of Energy mission to ensure long-term preservation of and access to the results of DOE research and development (R&D) investments. Across the full spectrum of DOE R&D*

programs, OSTI will provide accountability for all DOE scientific and technical information – in its many forms – through electronic, efficient, and user-friendly tools and technology.

Its mission is:

The mission of the Office of Scientific and Technical Information is to advance science and sustain technological creativity by making R&D findings available and useful to Department of Energy researchers and the public.

OSTI also manages the agency-wide Scientific and Technical Information program (STIP) for DOE. This is a collaboration across the entire DOE complex to ensure the results of DOE-funded R&D are identified, disseminated, and preserved. On the OSTI STIP website (http://www.osti.gov/stip/) 'Scientific and Technical Information' (STI) is defined as:

Information products deemed by the originator to be useful beyond the originating site (i.e., intended to be published or disseminated), in any format or medium, which contain findings and technological innovations resulting from research and development (R&D) efforts and scientific and technological work of scientists, researchers, and engineers, whether Federal employee, contractor, or financial assistance recipient. STI also conveys the results of demonstration and commercial application activities as well as experiments, observations, simulations, studies, and analyses.

Scientific findings are communicated through various media – e.g., textual, multimedia, audiovisual, and digital - and are produced in a range of products such as technical reports, scientific/technical conference papers and presentations, theses and dissertations, scientific and technical computer software, journal articles, workshop reports, program documents, patents, publicly available scientific research datasets, or other forms of STI. (See Attachment 3, STI Product Types Made Available Through DOE STI Program.) STI may be classified, Unclassified Controlled Nuclear Information (UCNI), controlled unclassified information (CUI), or unclassified with no access restrictions. DOE-funded STI originates primarily from research and other activities performed by site/facility management contractors, direct DOE-executed prime procurements, DOE-operated research activities, and financial assistance recipients, in addition to DOE employees.

To meet its mission, OSTI provides a set of services and tools for use by both the DOE community and the general public. Since the 2009 Committee of Visitors (COV) Report, OSTI has reviewed its services and products and eliminated or consolidated several of them. The Office has been re-organized and resources re-balanced to reflect OSTI's three core STI functions:

- Acquisition and Information Programs
- Preservation and Technology
- Access and Operations

The 2015 – 2019 OSTI Strategic Plan makes clear that OSTI has a renewed focus on providing comprehensive access to the results of DOE R&D investments. The strategic plan then organizes the work of OSTI around five goals:

- Goal 1: Accountability for DOE Unclassified R&D Results – Collection, Acquisition
- Goal 2: Long Term Preservation of DOE R&D Results
- Goal 3: Collection, Protection, Preservation, and Secure Access to Classified R&D Results, Unclassified Controlled Nuclear Information (UCNI), and Controlled Unclassified Information (CUI).
- Goal 4: Maximum Use of and Visibility for DOE R&D Results
- Goal 5: Strong Foundations, Partnerships, and Agility.

An exciting element of the strategic plan is the implementation of public access to the peer-reviewed scholarly publications resulting from DOE R&D funding. DOE funding results in about 25,000 manuscripts per year being accepted for publication by leading scientific journals and conferences. OSTI has the mandate to implement the DOE's Public Access Plan for these refereed publications and to work with the DOE author community, the publishers and other stakeholders to provide free, public access to the full text of these papers after a 12-month *'administrative interval.'*

The DOE Public Access Plan makes reference to a new portal and search interface tool called DOE PAGES^{Beta} – Public Access Gateway for Energy and Science – being developed by OSTI to meet the OSTP requirement for better public access to scientific scholarly publications. DOE PAGES^{Beta} provides abstracts and metadata for these publications and also a link to the full text 'version of record' (VoR) hosted by the publisher when the article is available on the publisher's site openly and without charge. When such a version is not available, DOE PAGES^{Beta} will link to a full text version of the accepted manuscript 12 months after the article's publication date. A publicly accessible beta version of DOE PAGES^{Beta} was launched in August 2014. In October 2014 the Department began to include requirements for the submission to OSTI of accepted manuscripts and publication metadata in award agreements as well as in national labs' STI submission requirements.

4. Summary of OSTI Products and Services

After streamlining their products and services, and increasing the emphasis on DOE R&D results, eight dissemination products are now supported. These are summarized below:

- **DOE PAGES[Beta]** – A portal to journal articles and accepted manuscripts resulting from DOE-funded research (http://www.osti.gov/pages/).
- **SciTech Connect** – A search tool to find technical reports, journal articles, accepted manuscripts, conference papers, patents, theses, books, multimedia, and data information (http://www.osti.gov/scitech/).
- **DOE Data Explorer** – A search tool for finding scientific research data resulting from DOE-funded research (http://www.osti.gov/dataexplorer/).
- **ScienceCinema** – A collection of multimedia videos highlighting the U.S. Department of Energy's most exciting research (http://www.osti.gov/sciencecinema/).
- **DOEpatents** – A searchable database of patent information resulting from research sponsored by DOE and its predecessor agencies (http://www.osti.gov/doepatents/).
- **E-Print Network** – A gateway to millions of e-prints in basic and applied sciences (http://www.osti.gov/eprints/).
- **DOE R&D Accomplishments** – A listing of remarkable accomplishments in science resulting from past DOE research and development (http://www.osti.gov/accomplishments/).
- **ESTSC Energy Science and Technology Software Center** – A collection of DOE-sponsored scientific and technical software (http://www.osti.gov/estsc/).

OSTI also supports a **Data ID Service** to DOE researchers that offers free assignment and registration of a Digital Object Identifier (DOI) for DOE datasets. The intent of the service is to increase the acquisition, registration, and usability of DOE R&D datasets (http://www.osti.gov/home/document/data-id-service).

Three other OSTI services allow federated searching across a wide range of resources:

- **NLE National Library of Energy[Beta]** – Science resources and other information from across the DOE complex (http://www.osti.gov/nle/).
- **Science.gov** - A gateway to government science information and research results consisting of over 60 scientific databases and 200 million pages of science information (http://www.science.gov/).

- **WorldWideScience.org** – A global science gateway comprised of national and international scientific databases and portals (http://worldwidescience.org/).

In addition to these public-facing services, OSTI develops and maintains two services for the DOE research community:

- **E-Link Energy Link** – This tool is developed and maintained by OSTI for the DOE research community to submit their STI products to the DOE (https://www.osti.gov/elink/).
- **Science Research Connection (SRC)** – This tool provides the DOE community with access to research information integrated from various OSTI databases, including both unclassified/unlimited and statutorily controlled information, which is accessible on a need-to-know basis (https://www.osti.gov/src/nag.jsp?nextURL=https://www.osti.gov/src/index.jsp)

The purpose of the E-Link tool is to collect results of DOE's annual R&D investment in the form of 30,000 – 40,000 STI products per year. There are plans for the integration of the E-Link service with the Office of Science's PAMS – the Portfolio Analysis and Management System.

5. The Charge Letter for the ASCAC Subcommittee

The charge letter (see Appendix A) requested ASCAC establish a standing subcommittee for an initial period of two years to advise the Office of Science on matters associated with the DOE Office of Scientific and Technical Information (OSTI). As its first activity, the ASCAC-STI subcommittee was asked to examine the following four questions:

a. Are the current OSTI products and services best in class and are they the most critical for the OSTI mission given the present constrained budget environment?

b. Do OSTI products and services fulfill customer needs now?

c. Are the OSTI products and services positioned to evolve to fulfill customer needs in the future? Has the OSTI strategic plan appropriately addressed the rapid evolution of technologies, research product types, and ways in which research results are communicated and shared?

d. What is the national and international standing of OSTI with respect to similar organizations whether at other U.S. Federal Agencies, DOE Laboratories, or universities? In what areas must OSTI be a clear leader to fulfill its mandated responsibilities to the DOE?

Additional guidance was later provided to the subcommittee by the Office of Science in the form of four succinct questions:

- Is the mission statement sensible in the light of the statutory authorities?
- Is OSTI organized and staffed to accomplish today's mission?
- Are the current and planned OSTI products and services the correct ones?
- What suggestions would the subcommittee make for the next steps?

This report will address the four charge questions in detail and comment as appropriate on this additional guidance.

It will be helpful to explain the rationale for selecting members of the subcommittee. It is clear that OSTI is well-known by significant parts of the library community both inside DOE and outside, and both nationally and internationally. However, the OSTI mission statement explicitly refers to its users as '*Department of Energy researchers and the public*.' In addition, the 2009 COV (see next section) had a serious concern that '*OSTI was not well known within DOE*.' For these reasons, the majority of the subcommittee members were deliberately chosen to be working DOE research scientists from a number of DOE Labs who were working with several different scientific communities (see Appendix B). The remaining subcommittee members were from the non-DOE, university library community, from academic researchers in information science and policy, and from industry, drawn from the US and the UK.

Finally, the subcommittee will review only the public, unclassified material, and will not review activities of OSTI dealing with classified and controlled DOE R&D material.

6. Issues noted in the Report of the 2009 COV

A Committee of Visitors review of OSTI reported their recommendations, findings, and observations in 2009. The report began by commending the leadership of OSTI on its motivated and capable workforce and its spirit of excellence and entrepreneurship. However, the COV also had a number of concerns and suggestions for improvement.

- **Balance**
  The major concern of the COV was about the *'balance between its mission to provide ready access to DOE R&D results and its more entrepreneurial mission of making all scientific information available to the world.'*

- **Completeness**
  Another concern was about the completeness of the collection of DOE R&D results and the COV noted *'the existence of less than optimum capture of DOE R&D output.'*

- **Customers**
  One serious concern of the COV was that *'OSTI was not well known within DOE.'* One COV member recommended that OSTI should undertake a new action: *'listening to what DOE staff do, how they do it, and the challenges they face.'*

- **Digitization**
  The COV recommended that OSTI should expand digital access to its non-digitized legacy collection of historical reports. It was felt that DOE should incur a one-time cost to place this important heritage resource in the public domain. It was also recognized that OSTI had '*demonstrated a clear financially efficient plan to achieve this.'*

- **Opportunities**
  One reviewer noted that *'OSTI is poised to be able to provide access to the primary literature, and to ensure its interoperability with the other publicly accessible databases it currently curates, providing a rich new resource that will facilitate new kinds of search and enable new kinds of computational research to take place.'*

We first look at the extent to which these concerns of the 2009 COV have been addressed in the interim.

7. Progress since the 2009 COV Report

It is clear that OSTI has made considerable progress since the COV in 2009. In 2014, the new OSTI Director, Brian Hitson, led an effort to refocus the Office and to develop a credible strategic plan for the next five years. In addition, with the two OSTP memos in 2013 on increased public access to DOE research results and on an Open Data Policy (see Appendices A and B), OSTI has taken on a key role for DOE in their Public Access Plan (see Appendix C).

We assess OSTI's progress on the areas identified by the COV under the five headings: Balance, Completeness, Customers, Digitization and Opportunities.

Balance

The concern of the 2009 COV about the balance between capturing DOE STI and a broader activity concerned with STI from other national agencies and international organizations seems to have been resolved. OSTI now has a greater focus on capturing and improving the completeness of DOE material, and in addition it provides a useful

broader function for STI with services such as Science.gov and WorldWideScience.org, with their federated search capabilities. The addition of the new public access requirement for DOE STI and the introduction of DOE PAGES^{Beta} service further reinforce the DOE focus. With both NSF and DOD partnering with OSTI in the implementation of their public access plans, DOE is likely to gain from an increased cross-agency alignment, and the PAGES model could ultimately embrace all of DOE, NSF, and DOD's refereed publications.

Completeness

A second, related concern of the 2009 COV was '*the existence of less than optimum capture of DOE R&D output*.' OSTI has now put in place a program to remedy the omissions of the past and fix the problem of a *'leaky pipeline'* and incompleteness of coverage for DOE publications. The new public visibility of the full text of research journal articles and conference papers in addition to the other DOE research output previously captured and made publicly available by OSTI, will constitute very significant progress towards the goal of the OSTP memos.

Some of the '*incompleteness problem'* was directly attributable to budget reductions in the 1990s, leading to reductions in the quality and quantity of OSTI's STI collection. According to Director Brian Hitson:

> *'OSTI had historically employed physical scientists to perform abstracting and indexing, which resulted in complete and very high-quality metadata, including journal article metadata for articles both produced by, and of interest to, DOE scientists. With the budget cuts, OSTI had to eliminate these contracts and put more onus on the DOE Labs and grantees to prepare metadata submissions. Because of the "burden" issue, the required metadata elements were reduced to a bare minimum with abstracts now optional, based on the belief that an increase in indexing of full-text electronic documents would aid in search. During this time, individual lab performance varied widely, and certain uncontrollable events, such as 9/11 (due to more stringent review and release practices) and security lapses at several labs, adversely affected submissions. In 2001, submissions reached a low point of ~10K records/year. By 2004, most of the 9/11 concerns had been resolved, and submissions held steady at 15-18K records/year through 2010. In the last 2-3 years, the number of submissions has ranged from 22K-30K records/year, with 2014 being a banner year at 40+K (due to finding pockets of legacy STI and new sources not previously tapped). It is important that OSTI maintain steady and increasing focus on working with labs and*

grantees so that we see steady and increasing submissions, especially in light of public access requirements.'

Aside from publications, data and software remain two areas where the coverage is at best sporadic. However it is good to see that OSTI is engaged in efforts to establish policies at the DOE level to gain more support going forward. One remaining question is whether there should be an effort to recover key STI of the past such as publications, data and software that had not been registered with OSTI, but would greatly add to the overall knowledge collection.

Customers

The third area of concern of the 2009 COV was the lack of awareness of OSTI within the DOE Labs, other than the library and information management community. It is apparent that OSTI relates well to information management communities both inside and outside of DOE but it is also abundantly clear that the role of OSTI was not well known to most DOE researchers on this subcommittee. Since the OSTI mission statement specifically identifies DOE researchers as one of the core target communities for its services, this seems to be a very serious problem. There seems to have been little or no progress in addressing this issue since the last COV and certainly no major action aimed at '*listening to what DOE staff do, how they do it, and the challenges they face*.'

However, it seems clear that not all the blame for this ignorance of OSTI's role should be directed at OSTI. Library staff, management and STIP representatives at the labs should all be working with OSTI staff to help make strong connections with the lab researchers and provide OSTI with useful feedback about its services. It is essential that the needs of the researchers are known and included in the development process for OSTI services if these services are to realistically make '*R&D findings available and useful to DOE researchers*.' DOE Program Offices, including OSTI's parent organization, the Office of Science, through federal research program managers, can also facilitate and enable more direct dialog between OSTI and the DOE research community.

Digitization

The digitization program for legacy DOE reports continues to make steady progress. There is still the opportunity for DOE to make a one-time investment to complete the process of putting this important heritage resource in the public domain.

Opportunities

With respect to opportunities for '*new kinds of computational research to take place'*, the subcommittee believes that OSTI, with its new capability to collect and allow public access to the full text of research journal and conference papers, is now at the start of

such a transformation. OSTI's Data ID Service will facilitate linking publications to the underlying data as well as permitting citation of research datasets. Finally, in order to permit research reproducibility and further the '*open science'* agenda, there will frequently be a requirement to link publications to research and/or analysis software. As we discuss below, this is something that OSTI could undertake with an expanded/re-developed ESTSC software service. With such linked tools and publications, OSTI will have access to a unique set of statistical data about DOE publications, usage and research funding. Such data will be helpful for guiding the development of existing and future services and tools, and identifying economic and societal impacts of DOE research.

8. Discussion of visit to OSTI and follow-up

The detailed agenda for the visit of the ASCAC-STI subcommittee is given in Appendix C. A brief summary of the presentations follows:

- The Director of OSTI, Brian Hitson, opened the meeting by giving us an overview of the '*broad landscape'* for OSTI's work. In particular he began by focusing on keywords in the charge questions such as products, services, customers, best in class, and research product types. Hitson then explained how the OSTI organization had addressed the '*balance'* issue and he identified near, mid, and long term priorities for OSTI.

- Associate Director, Mark Martin, described OSTI's strategic goals and objectives. Implementing DOE PAGES^{Beta} to fulfill the acquisition aspects of the DOE Public Access Plan was a high priority with the goal that capturing the research publication content for all DOE Labs should be 90% comprehensive by 2018. He noted a shift in the traditional article format and an increase in searchable multimedia forms of STI. The OSTI Data ID Service assigns Digital Object Identifiers (DOIs) to research datasets and this will facilitate linking publications to the supporting data. Goal 4 of the OSTI Strategic Plan was to enable '*maximum use of and visibility for DOE R&D results*' and he outlined better communication and outreach as the key drivers for visibility.

- Assistant Director, Judy Gilmore, described the DOE's Scientific and Technical Information Program (STIP). This program is required through a DOE directive and supports the '*key tenets of DOE's STI responsibilities and objectives.*' All labs are represented in STIP so that it can provide a framework for '*routine communication and information exchange*.' There is an annual meeting and OSTI's leadership of STIP is clearly appreciated by the library and IT communities in the DOE labs.

- Project Manager, Lance Vowell, took us through the E-Link electronic collection tool capabilities and their focus on '*minimizing the research/administrator burden*.' E-Link is evolving and will be linked to the DOE's Portfolio Analysis and Management System (PAMS).

- Jannean Elliott discussed new/emerging types of STI and the new Data ID service launched in August 2011. OSTI was the third US member of the DataCite international organization and OSTI can now allocate DOIs for datasets to DOE researchers at no charge. There is an impressive list of clients for the Data ID Service and all the datasets are visible in OSTI's Data Explorer tool. OSTI is supporting the DOE Office of Science in its implementation of the Digital Data Management component of the Public Access Plan. OSTI is also working towards integration and validation of ORCID researcher identifiers.

- Assistant Director, Judy Gilmore, discussed the new public access requirements. The new OSTP mandate will allow the DOE to bridge a key gap in OSTI's STI collections by requiring the collection of the 'gold standard' of scientific communication – refereed research journal articles. Furthermore, after a 12-month '*administrative*' delay, the full text of these research papers will either be available from lab and grantee institutional repositories, the CHORUS publisher consortium sites, or from a '*Dark Archive'* maintained by OSTI.

- Lorrie Johnson described the steps that OSTI is taking to measure and increase the comprehensiveness of its DOE collections. The DOEpatents service is now OSTI's first comprehensive collection.

- Mark Martin talked about Information Dissemination and OSTI's product philosophy. The goal was not to compete with Google and Bing search engines but rather to offer relevant search technology enhancements for searching scientific information. Clear examples of innovation and leadership are:

    - In SciTech Connect, OSTI has implemented a keyword-to-concept mapping semantic search capability.

    - In ScienceCinema, OSTI has partnered with Microsoft Research to implement automated audio indexing of videos.

    - With searches across multiple websites and databases in the NLE service, Science.gov and the WorldWideScience.org service, OSTI offers an innovative form of federated search.

    - In collaboration with Microsoft Research, OSTI has implemented automated multilingual translation in the WorldWideScience.org service.

- Catherine Pepmiller looked at '*Future Product Directions*' and discussed three example pilot projects:
    - Citation of the future: interlinking publications, data and people – using CrossRef and DataCite APIs to link publications and data
    - '*Named User'* functionality – enhanced product functionality for named user needs
    - Social functionality – allowing interactive user engagement with content
- Debbie Cutler talked about '*Content Quality and Curation*.' Actions included the formation of a '*Quality Team'* and establishing metrics for demonstrating quality improvements.
- Product Manager and Librarian, Lynn Davis, gave a presentation on '*Targeted Dissemination and Outreach.*' One approach was to enhance the discoverability of DOE's '*deep web*' material by 'surface web' search engines such as Google and Yahoo. The inclusion of Microdata via schema.org was one way of doing this. Academic libraries using OSTI's MARC system records were one target community as were other library commercial discovery systems. Other outreach targets were other federal agencies, international organizations, and community sharing platforms like ResearchGate.
- The next presentation was by Sara Studwell on '*Product Usage and Metrics*.' Visits, page views, and a number of other visit parameters were tracked using Google Analytics.
- The final presentations were on external comparisons of OSTI services. Joanna Martin discussed national comparisons and how OSTI services compared to those of other federal agencies. Lorrie Johnson then discussed some international comparisons with services from organizations such as CNRS in France, CISTI in Canada, and TIB in Germany.

The visit ended with a brief visit to the Information Preservation service and a short look at OSTI's technology infrastructure.

The discussion with the subcommittee members during the presentations was interactive and wide-ranging. The major points that arose are discussed below:

OSTI Products and Services

The services that OSTI provide look robust and professional. The development of the DOE PAGES^{Beta} service in particular, and the collaboration with both NSF and DOD, are to be commended. In discussion, the following suggestions and concerns were raised by members of this subcommittee:

Customers

- It is necessary to distinguish the different user communities in assessing the OSTI services – librarians, researchers, and the public, for example.
- The majority of subcommittee members felt that the present set of OSTI services are not widely used by DOE researchers.
- It is necessary to compare the OSTI tool experience to that of individual research communities and their tools. For example, the particle physicists' INSPIRE system is widely liked and used by the community for its one-stop-shop service.

Unification

- Researchers would value the tools if they offered a more fully integrated experience to search for information. The current set of tools is confusing and contains many redundancies.
- It could be valuable to engage a user experience expert to better understand and fit the set of tools and interfaces to the needs of DOE researchers.
- It could be worthwhile to engage a software QA/QC test team to perform more thorough testing of tools before release. The committee members encountered several bugs in evaluating the tools.

The ESTSC Software Service

- There was no information – or usage information - presented on the Energy Science and Technology Software Center (ESTSC) service for making available DOE software. Further examination of this service by members revealed that ESTSC is in need of a re-evaluation.
- In some cases software is as important as data for '*research reproducibility*' and re-use. There was no focus on software associated with research publications. Software used in publications often needs to be linked to the text just as much as the data.
- The ESTSC site seems to function by charging customers for downloading DOE software. In this age of public access and with the prevailing DOE requirement for open source software development, this seems like an anachronism.
- Among the '*best-selling*' DOE software products are the ScaLAPACK linear algebra libraries (www.netlib.org), and the Chombo software for adaptive solutions of partial differential equations (https://commons.lbl.gov/display/chombo/). Unfortunately, although these software are developed and maintained at DOE labs, they are not available through OSTI, demonstrating the significant gaps in coverage

in the OSTI software collection, as well as the lack of interaction with the research community.

- There should be a complete re-think of the ESTSC service – but this will need to involve researchers. A dialog with the DOE Lab research software developer community seems overdue.

Outreach

- Neither OSTI nor STIP seem to be well-known outside the DOE laboratory library and management community. Perhaps the STIP team needs to include a researcher from each DOE Lab as well as a librarian or manager.
- The STIP teams at many of the labs seem to be missing opportunities given by these new public access developments, to truly engage with researchers.
- A program of '*listening*' sessions with researchers – as recommended in the 2009 COV report – would seem to be a good first step.
- At present the outreach is not focused on the DOE research community and appears to be driven by '*information push'* techniques such as newsletters that have clearly been ineffective.

Usage, statistics, and evaluation

- The data on usage was puzzling since there was only a tiny amount of usage of OSTI services by the .edu and .gov domains. Of course, these two domains constitute a very small fraction of all the commercial and other public domains so this may not be unreasonable. However, given the range of scholarly research on energy topics, the very low amount of access to OSTI services from the .edu domain still seems somewhat surprising.
- OSTI should work with its stakeholders to first define success criteria for the services and then devise appropriate metrics and capture measures to see if they are achieving success.
- More detailed usage figures and statistical analysis would allow prioritization and provide guidance on future product development.
- User interface studies and other forms of evaluation of systems and services should be conducted.

Director Brian Hitson commented:

> *'While we're probably best known by librarians (whose job it is to be experts on information sources), we're most used by a wide range of public constituencies; that is, it's probably a very small percentage of librarians who account for 40M pageviews/year of OSTI-held content and the (so far) uncounted usage of lab-hosted content (which bypasses OSTI as the pass-through agent, but which OSTI is responsible for as the manager of the distributed DOE STI network). Knowing that 80+ percent of our usage comes through Google referrals, we are fulfilling the public dissemination aspect of DOE's legislative mandate, but the distributed and far-flung nature of this public user base doesn't know or retain that it is OSTI making this content accessible to them. We want to find the right balance in delivering products that meet public dissemination needs/requirements and lend themselves to reaching the public through Google-type search engines while also targeting specific scientific communities with specialized features and functionality. We're keen to meet this latter need, which will require careful prioritization, as each scientific discipline is likely to have unique needs. This was the "named user" aspect to the product roadmap discussed by Pepmiller and Studwell in the on-site review.'*

Data issues

- The subcommittee welcomed OSTI's forward thinking on data issues and their work with DOIs and support of the DataCite organization. Linking publications to the supporting data is clearly an important next step for public access and reproducible research.
- The Data ID Service seems very valuable and it seemed curious that this service was not available as a '*top tier*' service on the OSTI website and required some effort to find.
- OSTI clearly recognizes the importance of unique researcher IDs and is a member of the ORCID organization. However, OSTI has no mandate to insist that all lab researchers obtain ORCID IDs.

9. Response to the Charge Questions

Before responding in detail to the Charge questions, it is appropriate here to acknowledge the professionalism of the OSTI organization. All of the presenters were knowledgeable and enthusiastic about their subject area. From the Director down, all the staff in OSTI were very motivated and committed to delivering excellence. It was particularly impressive to see how

OSTI had stepped up to the challenge of increased public access to research journal and conference papers. The staff had also thought innovatively about future developments and challenges.

Charge question (a): Are the current OSTI products and services best in class and are they the most critical for the OSTI mission given the present constrained budget environment?

The subcommittee found the phrase 'best in class' vague and difficult to interpret. The utility of any given service also depends on the customer audience being targeted – the library community, researchers or the public. However, we can make some definite statements:

- OSTI's products and services are professional and generally well done, and it is welcome that they are now more focused on the search and discovery of DOE STI R&D.
- OSTI is a charter member of CENDI – a US interagency working group of senior STI managers. OSTI operates Science.gov, CENDI's flagship, cross-agency STI product for searching US Government scientific and technical information.
- OSTI was the first US federal agency to be a member of the DataCite organization. With its Data ID Service, OSTI can now provide researchers with DOIs for their datasets. These datasets are then made available to users via OSTI's Data Explorer database. The Data ID Service is clearly a valuable and forward-looking service offered by OSTI.
- Many of OSTI's services do incorporate leading-edge technologies and, in this sense, can be regarded as 'best in class':
    - SciTech Connect offers a powerful semantic search capability
    - ScienceCinema video service integrates automated audio-indexing technology
    - WorldWideScience.org uses automated language translation technology.
    - OSTI also offers an innovative federated search capability in NLE^{Beta}, Science.gov, and WorldWideScience.org.
- In terms of providing public access to the full text of journal articles, the NIH, with its National Library of Medicine and its PubMed Central repository, has a significant head start on all other federal agencies. The NCBI PubMed Central site also provides access to a large number of specialist biomedical databases. Their Entrez cross database search engine can then find supplementary information relevant to the

PubMed Central article. At present then, the NIH PubMed Central service must be regarded as best in class among the federal agencies.

- OSTI's rapid development of the DOE PAGES^{Beta} service and the associated further development of their E-Link service for submission have been impressive. This role in implementing the DOE's Public Access Plan is clearly critical for the fulfillment of OSTI's mission.

- The DOE PAGES^{Beta} service must be regarded as a promising newcomer to the public access agenda. Perhaps surprisingly, it already seems clear that OSTI's DOE PAGES^{Beta} service could prove to be an attractive alternative solution to that offered by the NIH. Both the NSF and DOD funding agencies are working with OSTI and are committed to following the PAGES solution to implement their public access plans.

- OSTI has taken the lead role for DOE in engaging with CHORUS, a consortium of scholarly publishers. The two academic librarians on the subcommittee both regard this commitment to CHORUS as a highly controversial move by DOE. However, OSTI is also responsible for implementing safeguards against non-compliance by the publishers and these concerns should be mitigated by the option of providing access to the full-text of accepted research papers via a dark archive maintained by OSTI.

- Overall, the subcommittee found that OSTI has made great strides recently and their latest product releases appear to have elements that are best in class. However, a focus on unifying the product set, reducing redundancy, and improvements in content coverage will be required before they can claim to be fully best in class.

One OSTI service that was found not to be best in class is the ESTSC software service. This service and its software inventory seem very outdated and out of touch with the leading DOE research software developers. The ESTSC model of charging for software seems unlikely to be an attractive offer given the open source culture of the scientific research community. GitHub is an interesting example of a modern software repository with tools that support the open source community software development process. In addition, some of the datasets stored in the Data Explorer tool prior to OSTI's establishing their Data ID service need to be reviewed.

Comments

- The ScienceCinema video indexing service is an interesting addition to the SciTech tool and contains a significant collection of videos and has implemented a nice search capability. However, additional work on improving metadata such as source credits and information is needed. It was also not clear what analysis was done to determine that video was the most critical priority for collection versus images, audio, etc. OSTI might possibly explore a collaboration with Google's Youtube® since

this has become a *de facto* destination for a large amount of video content created by the DOE and non-DOE R&D researcher community.

- There are now signs that Elsevier is attempting to extend the embargo period from 12-months to 48-months, a major change from their 2004 policy. The Elsevier proposal is also sowing much confusion in the academic library and institutional repository communities and in the scientific research community. It is therefore important that OSTI implement the safeguards against non-compliance by the publishers as specified in the DOE Public Access Plan:
    - '*During an "administrative interval" of up to twelve months, PAGES will not provide access to the full-text manuscripts. During this time, metadata including links to the publishers' VoR will be discoverable through the PAGES search interface and via PAGES APIs*'
    - *'PAGES will automatically reconcile DOIs submitted by DOE authors and by publishers to determine whether the VoR is accessible by the end of the administrative interval. In cases where the VoR is not accessible, PAGES will display a link to the accepted manuscript.*
    - *'In all cases, OSTI will maintain a dark archive of manuscripts to be used in the event links become broken or full text access is otherwise interrupted or discontinued. The dark archive will be part of the Department's Enterprise Data inventory.'*
- There is as yet an unfulfilled potential to provide services that leverage OSTI's unique position in holding references to all DOE STI products. It is possible that OSTI could offer tremendous value to the DOE research community by identifying and publishing those linkages. This is something that no one else could easily do today.
- The ESTSC service needs to be re-envisioned in a dialog with the DOE researchers. One possibility is for a new service that provides a home for software supporting research publications. As mentioned earlier, in discussions of research reproducibility, access to the software used by the authors can be as important as access to supporting data.

Charge question (b): Do OSTI products and services fulfill customer needs now?

- OSTI provided comparative evidence for the quality of its services. However, OSTI clearly sees its natural peer organizations to be other national library services, rather than community archives (e.g. Earth Systems Grid Federation, High Energy Physics INSPIRE/arXiv and Astronomy ADS), or public sharing sites such as ResearchGate,

GitHub or NanoHub, which offer very different and more interactive ways of sharing knowledge.

- In order to answer this Charge question, it is necessary to separate the different types of customers being served by OSTI's products. We note that:
    - From the usage statistics there appears to be significant take-up of OSTI services by the public and by commercial services.
    - Unfortunately, OSTI's services do not appear to be widely used by the DOE researcher community, a community that is specifically called out in OSTI's mission statement.
    - In respect of outreach to the DOE researchers, very little seems to have changed since the COV of 2009.
    - As a result the services are more targeted towards librarians than researchers and the OSTI services seem cumbersome by comparison to existing domain-specific solutions (e.g., the INSPIRE system at SLAC).
- We should note that certain classes of customers (such as major DOE data program managers in ARM and CDIAC) are satisfied with existing OSTI products and services, and see OSTI as one of the few solutions (and the only DOE supported one) for their needs. They also see OSTI as being open to tailoring and improving their services based on their needs.
- Researchers see the need for more integration of the different services and an improved user interface. Using user-interface simplification and best-practices to unify and reduce redundancy in the toolset would improve the user's experience of OSTI services.

Comment

- Further development of OSTI products targeted at DOE researchers must involve dialog with the DOE research scientists.

Charge question (c): Are the OSTI products and services positioned to evolve to fulfill customer needs in the future? Has the OSTI strategic plan appropriately addressed the rapid evolution of technologies, research product types, and ways in which research results are communicated and shared?

- The emerging challenge of collecting electronic versions of graphs, tables, and images in papers does not seem to be currently addressed in detail in OSTI's plans.

Although data, software, images and video are all part of OSTI's larger data vision it must be emphasized that OSTI and the DOE STIP community must work closely with the DOE research community if they are to develop new useful services for today's researchers. A detailed implementation plan showing how OSTI will achieve the goals set out in their strategic plan would be helpful.

- In their presentations, the OSTI staff showed a good awareness of the likely evolution of the services with respect to linking publications to data. The Data ID Service is a useful start and OSTI's involvement with CrossRef, FundRef, DataCite, and the ORCID organizations will be valuable.

- OSTI's vision for providing '*named user*' functionality could clarify the issue of which part of their customer base they are addressing. Adding some *'social functionality'* could also help bring the interactivity of OSTI services up to the level of the best research community tools.

Comment

The evolution of libraries in an era of search engines and social computing technologies means that both OSTI and the DOE library community must involve DOE researchers to develop new services and tools that support the changing nature of scholarly communication.

Charge question (d): What is the national and international standing of OSTI with respect to similar organizations whether at other U.S. Federal Agencies, DOE Laboratories, or universities? In what areas must OSTI be a clear leader to fulfill its mandated responsibilities to the DOE?

- OSTI has a leadership role with the CENDI interagency group in operating the Science.gov gateway to government science information. This offers a federated search service across 60 scientific databases and 200 million pages of scientific information.

- The recent development of the DOE PAGES^{Beta} service for access to research journal articles has won the respect of the NSF and the DOD (DTIC) who are considering using the OSTI system for delivering their public access plans. This is an area in which OSTI must be a clear leader to fulfill its mandated responsibilities.

- The OSTI services employ a range of innovative technologies not uniformly available from their peer international scientific information organizations.

- In terms of international leadership and recognition, OSTI is a founder member of the WorldWideScience Alliance and is responsible for providing novel real-time searching and translation service over globally-dispersed multilingual scientific literature to the other Alliance members.
- OSTI products and services compare well with those delivered by similar organizations in Canada (CISTI), France (INIST-CNRS) and the German National Library of Science and Technology (TIB).
- Because of its recognized expertise in information management, OSTI was invited to chair the Technical Activities Coordinating Committee of the International Council for Scientific and Technical Information (ICSTI).

10. Response to supplementary questions

- **Is the mission statement sensible in the light of the statutory authorities?**
    - The OSTI mission statement is entirely appropriate in targeting DOE researchers and the public.

- **Is OSTI organized and staffed to accomplish today's mission?**
    - The recent re-organization of OSTI in terms of its three core functions has given OSTI clearer focus on DOE research results.
    - It is likely that some changes to the mix of technical expertise at OSTI will be required to design and develop services suitable for modern science environments. This could be acquired either through new hires or by collaboration with existing DOE lab researchers and librarians.
    - If OSTI is to take on a larger role with respect to data it needs to expand its expertise in this area.

- **Are the current and planned OSTI products and services the correct ones?**
    - The products and services need to be targeted for at least three different communities – the traditional library and information management community, the DOE research community, and the general public.

- The automated collection of publications and provision of public access versions should remain a top priority for OSTI. Optimization of the publication collection method could significantly reduce the burden on the data submitter (e.g. requesting only the DOI, organization, funding info, and a pdf) and significantly increase collection completeness.
- The start on collecting multimedia content is valuable but improvements in metadata and consideration of the priorities of different types of multimedia should be undertaken.
- The Data ID Service is a critical first step towards making datasets citable and linking data to publications. Other data collection, federating, and brokering services may be the next step.

- **What suggestions would the subcommittee make for the next steps?**
  - Initiate some serious two-way outreach and dialog with the DOE Labs research communities to better understand what services they would like and use.
  - A more detailed analysis of Google/Bing search results on DOE R&D could help determine in which areas OSTI should focus to deliver complementary functionality.
  - Discuss tool usability issues with the DOE research community with a view to developing an integrated 'one stop shop' approach to STI services
  - Enlarge the STIP management by 'researcher champions' from each Lab. The lab library staff need to work with researchers to understand the issues of research reproducibility and open science that require linking data and software to research publications.
  - With the STIP management, OSTI needs to develop the necessary skills to advise researchers about the required Data Management Plans. This could include discussion about possible data repositories for long-term storage of large data sets as well as how to release sufficient data to support the conclusions of the journal article.
  - Discuss approaches to **partner** with the DOE labs and researchers to improve content completeness and help reach the DOE goal.

11. Recommendations

Recommendations to OSTI

1. If OSTI is to truly fulfill its mission to create products and services to make '*R&D findings available and useful to DOE researchers'*, it needs to initiate a vigorous outreach program with the DOE Lab researchers. This must involve listening to researchers needs and understanding the strong and weak points of existing community sites.
2. OSTI should work with the DOE research community to re-invent the ESTSC software service. In addition, releasing software in support of a research publication needs to be supported. In respect of the Data Explorer tool, OSTI should undertake a review of the datasets currently included in the tool, in partnership with Lab researchers.
3. Work with the labs to identify 'researcher champions' who can work with the STIP community to strengthen the link to researchers. This could include advice on Data Management Plans and target data repositories.
4. OSTI should work aggressively to continue toward a unified user environment with a limited number of, clearly delineated, non-redundant tools and develop a master plan for future development and areas of expansion through community input.
5. Through partnership with the national lab librarians and researchers identify and address publication content gaps and develop clear instructions and guidelines regarding content submission requirements. Significantly improving the completeness of coverage of the publications collection will require creative solutions not mandates. In addition, OSTI needs to work with the Labs to correct present inaccuracies in the record and also to devise practical methods to check that research papers are made openly available after the 12 month embargo time.

Recommendations to the Office of Science

1. To promote a successful implementation of the public access requirement issued by OSTP, OSTI needs top-down support from DOE in clearly communicating that this is not a requirement/burden imposed by OSTI but rather a government-wide and DOE-wide requirement meant to share federal research results and accelerate scientific progress. In this regard, labs, grantees, and their authors need to be incentivized to comply with

this requirement, which partnership with OSTI staff can help them to fulfill, and one such incentive could be a measurable expectation expressed in labs' annual performance plans.

2. The Office of Science should consider defining a useful role for OSTI and the STIP management team in managing DOE data. Sharing and preservation of data are central to protecting the integrity of science, facilitating validation of results, and advancing science by broadening the value of research data to disciplines other than the originating one and to society at large. Possible roles for OSTI include:

    - Following the example of major journals and collecting digital versions of tables, graphs, and images from papers.
    - Working with all of the Office of Science Programs and the different research communities in the DOE labs to develop better solutions for linking data and software to publications.
    - Coordinating reviews of the data needs by discipline to identify explicit commonalities and differences between disciplines.
    - Participating in collaborative pilots that establish the open data and open science end-to-end infrastructures (data provenance, data workflows, experiment integration).
    - Assisting in the development of an evaluation plan to assess how well the DMP and OSTI services support the community.
    - Developing cost models for manageable and cost-effective data solutions.

    However, if OSTI is to play a major role in data management, they would need to acquire significant new expertise in research data and would need to be resourced appropriately. It is also essential that OSTI collaborate closely with the DOE research community from the beginning.

References and URLs

General references

1. Hitson - ASCAC briefing_Final.ppt, Presentation by Web Conference to ASCAC Subcommittee, March 12, 2015
2. Annotated OSTI usage metrics report, analysis, [Internal Memo conveyed via email] (6/5/2015)
3. ASCAC STI Subcommittee Notes 4.doc, Site Visit 27th May, 2015.

URLs for Relevant Memoranda and Policy Statements

- **OSTP Memo on Increasing Access to the Results of Federally Funded Scientific Research**

https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

- **DOE Public Access Plan**

http://www.energy.gov/sites/prod/files/2014/08/f18/DOE_Public_Access%20Plan_FINAL.pdf

- **OMB Memo on Open Data Policy**

https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf

- **DOE Office of Science Statement on Digital Data Management**

http://science.energy.gov/funding-opportunities/digital-data-management/

Appendix A: Charge Letter

Department of Energy
Office of Science
Washington, DC 20585

November 17, 2014

Prof. Roscoe Giles, ASCAC Chair
Department of Electrical and Computer Engineering
Boston University
8 St. Mary's Street
Boston, MA 02215

Dear Professor Giles:

Thank you for the many outstanding activities that you have led during your years as Chair of the Advanced Scientific Computing Advisory Committee (ASCAC). I am writing now to ask that ASCAC take on a new and very important task for the Office of Science. I am requesting that ASCAC establish a standing subcommittee for an initial period of two years to advise the Office of Science on matters associated with the DOE Office of Scientific and Technical Information (OSTI).

OSTI was established in 1947 to fulfill the agency's responsibilities associated with the collection, preservation, and dissemination of scientific and technical information from DOE R&D activities, both classified and unclassified. This responsibility was codified in the enabling legislation of DOE and its predecessor agencies and, more recently, was defined as a specific OSTI responsibility in the Energy Policy Act of 2005. OSTI's mission is to maintain, within the Department, publicly available collections of scientific and technical information resulting from research, development, demonstration, and commercial applications activities supported by the Department.[1]

With today's requirements for broad sharing of digital data and open access of publications, the "collection, preservation, and dissemination of scientific and technical information from DOE R&D activities" assumes a complexity impossible to have imagined when OSTI was formed nearly 70 years ago. External, independent advice will be needed as OSTI transitions its products and services to methods appropriate to the new era of information gathering and sharing.

As its first activity, I would like the ASCAC-STI subcommittee to examine the following and provide me with a report by the late-spring or summer 2015 meeting of ASCAC:

[1] EPACT 2005 (PL 109-58), SEC. 982. OFFICE OF SCIENTIFIC AND TECHNICAL INFORMATION.
http://www.gpo.gov/fdsys/pkg/PLAW-109publ58/pdf/PLAW-109publ58.pdf

Appendix A: Charge Letter (continued)

a. Are current OSTI products and services best in class and are they the most critical for the OSTI mission given the present constrained budget environment?

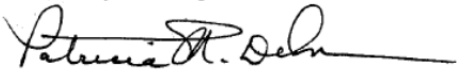
b. Do OSTI products and services fulfill customer needs now?

c. Are the OSTI products and services positioned to evolve to fulfill customer needs in the future? Has the OSTI strategic plan appropriately addressed the rapid evolution of technologies, research product types, and ways in which research results are communicated and shared?

d. What is the national and international standing of OSTI with respect to similar organizations whether at other U.S. Federal Agencies, DOE Laboratories, or universities? In what areas must OSTI be a clear leader to fulfill its mandated responsibilities to the DOE?

If you or the ASCAC-STI chair has questions, please contact Christine Chalk, the Designated Federal Official for ASCAC, at 301-903-5152 or by email at christine.chalk@science.doe.gov.

With very best regards,



Patricia M. Dehmer
Acting Director, Office of Science

Appendix B: Review Participants

Tony Hey (Chair)
Senior Data Science Fellow
eScience Institute
Campus Box 351570
University of Washington
Seattle, WA 98195-1570
tony.hey@live.com

Deborah Agarwal (LBNL)
Senior Staff Scientist
Lawrence Berkeley National Laboratory
Advanced Computing for Science Department
1 Cyclotron Road
Berkeley CA 94720
daagarwal@lbl.gov

Christine Borgman (UCLA)
Distinguished Professor & Presidential Chair in Information Studies, UCLA
405 Hillgard Avenue
Los Angeles, CA 90095
http://christineborgman.info

Concetta Cartaro (SLAC)
Experimental physicist
SLAC National Accelerator Laboratory
2575 Sand Hill Road, MS 34
Menlo Park, CA 94025
cartaro@slac.stanford.edu

Silvia Crivelli (LBNL and UC Davis)
Faculty Scientist
Computational Research Division
Lawrence Berkeley National Laboratory
sncrivelli@lbl.gov

Kerstin Kleese Van Dam (PNNL)
Data Scientist
Pacific Northwest National Laboratory
Richland WA 99352
Kerstin.KleesevanDam@pnnl.gov

Richard Luce (University of Oklahoma)
Dean, University Libraries and Associate VP for Research
University of Oklahoma
Norman OK 73019
rluce@ou.edu

Arjun Shankar, Ph.D. (ORNL)
Senior Research Scientist and Director,
Compute and Data Environment for Science (CADES)
Oak Ridge National Laboratory
1 Bethel Valley Rd.
Oak Ridge, TN 37831
shankarm@ornl.gov

Anne Trefethen (University of Oxford)
Pro-Vice-Chancellor (Academic Services and University Collections) and Chief Information Officer
University of Oxford
University Offices
Wellington Square
Oxford OX1 2JD
United Kingdom
anne.trefethen@it.ox.ac.uk

Alex Wade (Microsoft Research)
Director of Scholarly Communication
Microsoft Research, Microsoft Corporation
One Microsoft Way, Redmond WA 98052
awade@microsoft.com

Dean Williams (LLNL)
Analytics and Informatics Management Systems Project leader
Lawrence Livermore National Laboratory
Mail Stop: L-103
7000 East Avenue, P.O. Box 808
Livermore, California 94550
williams13@llnl.gov

Appendix C: Review Agenda for Subcommittee visit to OSTI May 27, 2015

1. **8:00-8:15 Welcome/Introductions**

2. **8:15-8:30 Review of charge letter questions – Tony Hey**

3. **8:30-9:15 OSTI – The Broad Landscape – Brian Hitson**
    a. In the context of charge questions:
        i. What are our products and services? STI Management and its constituent services and products
        ii. Who are our customers? Regarding "best in class," who is the class? Analogues and Initial Comparisons
    b. Strategic shifts in priorities
        i. Public access
        ii. Re-focus/Re-balance
        iii. Reorganization, Resources

4. **9:15-9:45 Strategic Goals, Objectives, Measures of Success – Mark Martin**
    a. Goals/objectives structure; measures of success
    b. Special emphasis areas and future direction

9:45-10:00 – Break

5. Drilling deeper
    a. 10:00-12:00 Information collection – Judy Gilmore, Lance Vowell, Jannean Elliott, Lorrie Johnson
        i. STIP – Scientific and Technical Information Program – Describe DOE-wide network of Labs, HQ, grantee STI managers and processes. Summarize research product types.
        ii. Electronic collection tools – E-Link, web services, Wizard, etc. Future enhancements to these processes (easy as possible for Labs, researchers; linking R&D investments to output, i.e. via PAMS)
        iii. New/Emerging types of STI (including dataset registration/discoverability) – Cover OSTI's involvement with DataCite, international membership, our clients, etc. Describe DOI registration service and process/workflow.
        iv. New public access requirements
            1. Review of OSTP and legislative mandates
            2. DOE journal article production
            3. Fulfilling requirements through STIP, complemented by publisher participation
        v. Measuring comprehensiveness – Describe efforts and methodologies to improve comprehensiveness across all STI formats. Patents success story. Future options for

analysis with public access/comprehensiveness. Role of "grey literature" (e.g. technical reports) with conventional literature

12:00-12:15- Break, Working Lunch to begin at 12:15

b. 12:15-3:15 Information dissemination – Mark Martin, Catherine Pepmiller, Debbie Cutler, Lynn Davis, Sara Studwell

i. Dissemination products – Product breakout; product philosophy/roadmap, and innovation (audio indexing, semantic search, federated searching, multilingual translations)
ii. Future product direction, modernization/innovation – Product research /investigation; formation of user/focus groups; linking publications to data, linking across STI types (technical reports to patents, etc.); meeting future customer needs, including needs expressed by SC (science of science, emerging technologies)
iii. Content quality/curation
iv. Targeted dissemination and outreach (e.g., .edu, API services) – Efforts with academic community, Google/Bing site maps, etc.
v. Product usage; OSTI efforts to drive usage – Trends in usage, increased latitude to gather data on users, new focus on understanding user behavior

2:15-2:30 – Break

vi. 2:30-3:15 Product demos – SciTech, ScienceCinema, PAGES, WWS.org – Catherine Pepmiller, Sara Studwell, Lorrie Johnson

6. 3:15-3:45 External Comparisons – Joanna Martin, Lorrie Johnson

a. "Best in Class" – Comparisons to other STI management organizations
b. "National and International standing" – OSTI's role and position in national and international STI community

7. 3:45-4:45 Facility Tour – Vaults, data centers, etc. – Jeff Given, Brian Hitson

a. Information preservation – Archival responsibilities, NARA, records management, show historic documents in vault
    i. Born digital preservation
    ii. Non-digital collection – Preservation, digitization
b. Technology infrastructure
    i. Architecture for current and future needs
    ii. Cyber security, cyber risk management

8. 4:45-5:00 Summary – Brian Hitson