A Usability Study of SciTech Connect, the Department of Energy's Public Access Portal

By Heidi A. Uphoff

Submitted in Partial Fulfillment of the Requirements for the Degree of Certificate of Advanced Study in Library and Information Science at the Graduate College of the University of Illinois at Urbana-Champaign, Summer 2016

CAS Project Committee:
Professor, Michael Twidale, Chair
Associate Professor, Jerome McDonough
User Experience Librarian, Suzanne Chapman
# Table of Contents

Introduction .......................................................................................................................... 4

Literature Review .................................................................................................................. 4

  About the Database ........................................................................................................... 4

  Government Public Access ............................................................................................. 7

  Usability Test Methods .................................................................................................. 19

Database Analysis ............................................................................................................... 24

  Personas .......................................................................................................................... 24

  Cognitive Walkthrough ................................................................................................. 28

  Heuristic Analysis .......................................................................................................... 50

  Comparisons to Similar Applications ........................................................................... 53

User Study ............................................................................................................................ 63

  Methods ......................................................................................................................... 63

  Results ............................................................................................................................ 64

Redesign Recommendations .............................................................................................. 74

  Basic Search Form .......................................................................................................... 74

  Homepage ...................................................................................................................... 75

  Advanced Search Form ................................................................................................. 76

  Search Results List ....................................................................................................... 78

  Full Record Display ..................................................................................................... 81

  Links to Items ............................................................................................................... 86

  Pop-ups ......................................................................................................................... 87

Conclusions and Future Study .......................................................................................... 87
A USABILITY STUDY OF SCITECH CONNECT

References ........................................................................................................................................89

Appendices ......................................................................................................................................97

A. Federal Public Access Plans Chart .............................................................................................97

B. User Test Protocol ......................................................................................................................107
A Usability Study of SciTech Connect, the Department of Energy's Public Access Portal

This study of SciTech Connect, the Department of Energy’s (DOE) portal to their funded research, was devised to identify usability issues in order to make redesign recommendations. The application has undergone many changes in its scope and function over the last few years in response to federal public access mandates, making it an interesting and valuable candidate for such a study. The methods for the study were two-fold: database analysis and user study. The database was analyzed through a cognitive walkthrough, heuristics, and comparisons to similar applications. In addition, a concurrent think aloud user study was conducted with twelve participants. The issues identified by those methods were utilized as the basis for redesign recommendations.

Literature Review

About the Database

SciTech Connect is a public portal to free, publicly available DOE funded products of research, also known as scientific and technical information (STI). The database is owned by DOE’s Office of Scientific and Technical Information (OSTI), responsible for collecting, preserving, and disseminating STI created by DOE laboratories, sites, and grantees as well as STI from DOE’s predecessor agencies (About OSTI, 2015). SciTech Connect was created to increase public access to STI and replace two older databases - DOE Information Bridge and the Energy Citations Database (SciTech Connect: FAQ, n.d.).

The content of the database’s collection varies greatly. It includes citations, technical reports, conference papers, presentations, posters, books, multimedia, software, and dataset citations. According to SciTech Connect’s FAQ webpage, the following subject areas are covered in the collection:
OSTI breaks up these subject categories further within the database through their controlled vocabulary set, Subject Categories for Unclassified STI Products. There are 50 subject categories in the controlled vocabulary with each assigned a two-character numeric code (OSTI, 2008). The collection spans over 65 years, from the 1940s and the days of the Manhattan project to the present (SciTech Connect; FAQ, n.d.).

SciTech Connect has many user features that would be expected in interfaces that search metadata records and electronic full text. A basic search leads to a list of results and facets to filter the results. Full record displays are accessed by clicking on a title in the search results list. Query words are bolded in the metadata. Full text can be accessed by clicking a link in either the
search results list or on a full record display. There is also an advanced search option. Further, the database has other features not commonly found in information interfaces. For example, users can browse by subject and build searches using the subject and other facets. Users can also export a single record or an entire search result list into an Excel, CSV, or XML file. Records have word clouds built from their full text and “more like this” suggestions for discovering similar records. Users can even search the full text of an individual item from its full record display. As well, there are two choices for searching: the default semantic search and the term search. The term search retrieves only exact matches for the terms a user inputs, while the semantic search maps search terms to related concepts and retrieves more results. The database allows users to create accounts to customize how they see results, to save searches, to create email alerts, and to export metadata.

Items in this collection can be accessed and discovered through other search interfaces. OSTI maintains databases with sub-collections found in SciTech Connect, including DOE patents for patents, DOE Data Explorer for datasets, ScienceCinema for audio and video items, and PAGES for peer-review publications. After items are uploaded into DOE’s databases, they are also made available through federal government federated search tools, such as Science.gov that searches research across 15 federal agencies, and WorldWideScience.org, which searches scientific research from over 70 countries (About OSTI, 2015). Major search engines, like Google and Google Scholar, also index the records in SciTech Connect’s collection.

While there are multiple options for search and retrieval of metadata and full text in SciTech Connect’s collection, there are various reasons users would choose this interface over others. With search engines, a user cannot limit their search to only free, publicly accessible full text. With the other government search interfaces, namely DOE’s sub-collections and federal
aggregate search tools, the collections as well as the tools differ and may not meet the user’s unique needs.

There are two reasons that make SciTech Connect an interesting and worthy candidate for an in-depth usability study. First, the database functions as a portal and not an archive or repository. Not all records with full text have it stored in SciTech Connect. Instead, these records store links which route users to a DOE site’s servers or to a publisher’s website to access the full text. Second, SciTech Connect has had recent additions to the scope of its collection since 2014, fully utilizing portal architecture to provide access to full text publications and datasets. Prior to October 1, 2014, the collection consisted mainly of technical reports and other grey literature. Grey literature, according to the Fourth International Conference on Grey Literature, October 1999, is “information produced at all levels of government, academia, business, and industry in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body” (What is Grey Literature?, n.d.). Essentially, “grey literature is everything but peer-reviewed journals and academically- or commercially-published books” (Von Hendy, 2014, p. 62). From the point at which the DOE Public Access Plan went into effect, more accepted manuscripts and publisher versions to peer-reviewed journal articles and conference papers, along with more citations to datasets, have been incorporated into SciTech Connect. The mix of public access concepts, like publisher embargo periods, with records management concepts, including report numbers, and the portal architecture, which links users to the full text and datasets, has the potential to be confusing for users.

**Government Public Access**

Government public access has its roots in the Open Access movement that began in the late 1990s and early 2000s with the development and increased use of the internet. According to
the Scholarly Publishing and Academic Resources Coalition (SPARC), a global coalition working towards making Open Access the standard for research,

Open Access is the free, immediate, online availability of research articles coupled with the rights to use these articles fully in the digital environment. Open Access ensures that anyone can access and use these results--to turn ideas into industries and breakthroughs into better lives (Open Access - SPARC., n.d.).

By lifting subscription barriers traditionally in place with journal publications, there is faster communication about discovery, a more even playing field for students, researchers, and industry professionals who don’t have subscription access to the publications they need, and a wider impact and audience for authors’ work. Collins (2011) described Open Access as an interest to the library profession as a “possible saving grace to the crisis in scholarly communication and a potentially viable solution to the ever-increasing, unsupportable inflation of library subscription prices” (p. 138).

Collins (2011) cited three catalysts to the Open Access movement in her review publication, “Open Access Literature Review 2008-9: A Serials Perspective.” The first was the National Institute of Health (NIH) mandate voted into law January, 2008. This required NIH-funded authors to deposit the accepted manuscript, their final version that includes peer-review content changes but does not have the publisher’s final editing and copyright, of all publications into PubMed Central within 12 months of publication. This historic mandate cleared many copyright roadblocks for health and biology publishers. Several publishers also deposit articles into PubMed Central for authors. The Nature Publishing Group made their strict self-archiving policies less harsh and now permit authors to deposit accepted manuscripts into public repositories after a 6-month embargo period (Collins, 2011, p. 140).
The second catalyst cited by Collins was university Open Access policies. The same year the NIH mandate went into effect, Harvard’s Faculty of Arts and Sciences had a unanimous vote for an Open Access policy. This was unique and important because at the time, only 12 university policies existed worldwide and this was the first to be instituted by faculty and not administration (Collins, 2011, p. 140). By 2012, over 300 institutions, programs, and funding bodies worldwide had adopted such a policy and that number continues to grow (Xia et al., 2012, p. 85).

Collins’s third catalyst, the emphasis on self-archiving institutional repositories for accepted manuscripts, is tied to university policies (Collins, 2011). Institutions built public repositories using new open system technologies and interoperability standards in the early 2000s, but were slow to accumulate items. This could be attributed to a number of factors, including lack of faculty awareness and incentives and faculty concerns over copyrights and how this would affect their tenure and promotion process. Early studies showed that universities with policies were able to collect a larger amount of items than those without policies, although this may have been caused by librarians and other institutional repository staff archiving papers on the faculty’s behalf (Xia et al., 2012).

There has been special attention paid to bringing the Open Access concept to U.S. federally-funded research. After all, “public funding of scientific, technical and medial (STM) research is undertaken with the expectation that the economic and social returns to taxpayers will exceed the amount of the research investment” (Houghton, Rasmussen, & Sheehan, 2010, p.1). In 2010, SPARC published a report that the calculated the potential impacts of the U.S. Federal Research Public Access Act (FRPAA), which was proposed three times in Congress but never passed. The act would have required all federal agencies with over $100 million in research
expenditures to make journal articles that resulted from their funding publicly available within six months of the publication date in an online repository. The reasoning was that in order to see the full potential of research funding, journal articles needed to be Open Access. New discoveries are built on past discoveries and journals are the primary vehicle for disseminating those findings. The report’s intent was to move past the anecdotal evidence of public access for a more accurate quantitative estimate of the future effect of FRPAA. The authors developed a model and decided on data sources for preliminary statistics that suggest over a period of 30 years from implementation of FRPAA, the benefits could be eight times the cost of implementation with two thirds of the benefits remaining in the U.S. and the other third spilling over into other countries (Houghton, Rasmussen, & Sheehan, 2010).

When reading about government public access, there is a vocabulary shift away from Open Access, perhaps as a consequence of its wider scope. The Open Access movement focuses mostly on journal articles, while public access at the federal level concentrates on all types of research outputs, including grey literature, much of which was already required to be publicly available, peer-review publications, including journal articles, and now scientific datasets. Recent increases in storage and processing capacity in addition to increases in the number of devices that collect and store data make sharing even large and complex datasets a reasonable goal (Bertot et al., 2014, p. 7).

In 2007, the American Association for the Advancement of Science published a report on a survey they performed of their members to review the acquisition and creation of intellectual property in the scientific community and the effects of those protections on research. 57% of survey respondents used data from public sources in their work. 25% had difficulties accessing data they needed because of delays or because they were denied access by the intellectual
property owners causing some negative effects on their research (Hansen, Kisielewski & Asher, 2007, p. 47 - 54). The scope of federal public access encompasses data, benefitting researchers like those surveyed for the “Intellectual Property Experiences in the United States Scientific Community” report.

A study released in 2014 examined how government employees were reacting to the new emphasis on data sharing. According to the authors of "Managing Scientific Data as Public Assets: Data Sharing Practices and Policies Among Full-Time Government Employees,”

A consensus is growing in the scientific community that the ability to answer complex scientific questions is contingent upon scientists managing and sharing data, challenging existing data, using existing data to address new questions, and merging existing data sets to create new data sets, especially as new computational tools come online (Douglass et al., 2012, p. 252).

The authors noted two positive trends. For one, there is interest by public employees to share their data. Two, the data already existing has advanced discovery (Douglass et al., 2012, p. 252).

A survey conducted by DataOne, a National Science Foundation (NSF) funded project focused on data preservation and sharing in the environmental sciences, found that the reasons government employees did not share their data closely matched those of scientists in academia. The top two reasons were “lack of funds” and “insufficient time,” followed by “lack of standards,” “no place to put data,” and “do not have the rights to make public” (Douglass et al., 2012, p. 258). However, while federal employees are still connected to the “communities of practice” of their discipline, those sharing practices are trumped by agency policies, practice, and infrastructure (Douglass et al., 2012, p. 253). This demonstrates a need for infrastructure funding to remove technological barriers as well as requirements to remove non-technical barriers, like
researchers not knowing if they have the right to share their data and if they should or should not share it.

While FRPAA was never passed, federal public access in its wider definition became the required default for federal agencies under the Obama administration. On his first day in office, President Obama committed his presidency to an “unprecedented level of openness in government” (Bertot et al., 2014, p. 7). Since then, he has issued two executive orders requiring federal agencies to err towards openness when evaluating Freedom of Information Act (FOIA) requests. He instructed the Office of Management and Budget (OMB) to issue an Open Government Directive to deliver openness, transparency, and accountability through guidelines and initiatives, of which “big data” is key (Bertot et al., 2014, p. 7).

On May 9, 2013, Obama issued another executive order, “Making Open and Machine Readable the New Default for Government Information.” The framework for instituting the principles of open, machine readable data was released in an OMB memorandum, entitled “Open Data Policy—Managing Information as an Asset.” The memo labeled information as a “national resource” that can “fuel entrepreneurship, innovation, and scientific discovery” (Burwell, VanRoekel, & Park, 2013). The memorandum defined data as structured information and listed seven principles for it being open: public, accessible to a wide range of users and for a wide range of purposes, described sufficiently enough for users, reusable with no license restrictions, complete, timely, and managed post-release. The memorandum laid out four areas for federal agency action:

1. Collect or create information in a way that supports downstream information processing and dissemination activities;

2. Build information systems to support interoperability and information accessibility;
3. Strengthen data management and release practices;

4. Strengthen measures to ensure that privacy and confidentiality are fully protected and that data are properly secured (Burwell, VanRoekel, & Park, 2013).

While these executive actions effect DOE policy, it was a memorandum released February 22, 2013, by the White House Office of Science and Technology Policy (OSTP) that was directly responded to by the DOE Public Access Plan. This particular memorandum directed all federal agencies with over $100 million in annual research and development expenditures to develop a plan to increase public access to their research. The requirements for the plan prioritized two different types of research output: peer-reviewed scholarly publications and scientific data in digital formats. The requirements for publications were specific and followed NIH’s well-established plans as the guideline for all others. Specifically, agencies must provide public access to the accepted manuscript or the publisher’s version of all peer-reviewed scholarly publications in a public repository within 12 months of publication date. The memorandum did not specify this to mean journal articles exclusively, but all peer-reviewed publications, including conference papers. While the publication may be embargoed for up to 12 months, the metadata to the publication should be provided in a public repository as soon as available (Holdren, 2013).

The requirements for scientific data in digital formats were less specific and as a result, vary greatly by agency in terms of their implementation. All agencies necessitate a Data Management Plan (DMP) with their funding proposals that describes how data in a project will be shared and preserved or describing why it would not be and how the results can be validated without sharing the data. All agencies must evaluate funding proposals and funded projects by their DMP, but it is up to the agencies individually to decide how to evaluate. Agencies’ plans for data should maximize public access to it while protecting personally identifiable information.
(PII), proprietary and intellectual property right interests, and national interests. Plans should encourage deposit of data in public databases, but unlike publications, this is not a strict requirement. There are no guidelines on what data should be shared or how, only that that agencies have to create plans to increase public access to it (Holdren, 2013).

DOE was the first agency to develop a plan to address the memorandum, the plan going into effect on October 1, 2014. They have since been followed by 18 other agencies, although NIH and the Institute of Education Sciences (IES) already had policies in place that met the requirements of the memorandum and their plans were not major changes to their status quo. A full comparison of agency plans is found in Appendix B: Federal Public Access Plans Chart. All plans have an up-to-12-month embargo after the publication date for accepted manuscripts. Some allow authors to choose the embargo period during submission, and other policies, like DOE’s and the Department of Defense’s (DOD), handle the embargo date themselves. Only certain plans permit authors to submit the publisher’s version when the publisher allows for unlimited distribution in their copyright statement. The rest specifically require the author’s final version.

Many agencies made use of already established repositories for publications and systems for submission. Eight agencies use NIH’s PubMed Central and its submission system: Administration for Community Living (ACL), Agency for Healthcare Research and Quality (AHRQ), Centers for Disease Control and Prevention (CDC), Food and Drug Administration (FDA), National Aeronautics and Space Administration (NASA), National Institute of Standards and Technology (NIST), Office of the Assistant Secretary for Preparedness and Response (ASPR), and the United States Department of Veteran’s Affairs (VA). Most have their own submission module in PubMed Central’s submission system. The CDC submission module also
deposits authors’ submissions into CDC Stacks, CDC’s own publication repository. Many of the CDC’s funded authors also have NIH requirements to meet, so the two agencies worked together to make the process easier on researchers. The DOE worked through their OSTI program and updated their submission systems for accepted manuscripts. Their publications are made available in SciTech Connect and a new database for publications, called PAGES. NSF will also be using DOE’s PAGES system, but with a separate interface called NSF-PAR, or NSF-Public Access Repository.

The rest of the federal agencies affected by the requirements have their own repository or one in development. IES uses ERIC, which was released in 2012. DOD is modifying DTIC, their reports repository, to handle publications. The Department of Transportation (DOT) developed the National Transportation Library (NTL) Digital Repository. The National Oceanic and Atmospheric Administration (NOAA) is creating a sub-set collection in CDC stacks. The United States Department of Agriculture (USDA) developed PubAg. The Smithsonian Institute has a repository, Smithsonian Research Online, but authorizes its grantees to choose other repositories to deposit their papers. This is a deviation from all other plans that require researchers to submit to a specific repository for that agency. Finally, the U.S. Geological Survey is expanding their current repository, USGS Publications Warehouse.

While the plans for publications are all very similar, agency data plans vary greatly. All require DMPs with funding proposals, but differ on DMP requirements. All request information on how data will be shared and preserved or why it will not be in the DMP but diverge on how that is evaluated. Generally, AHRQ, NIST, NIH, ASPR, NASA, VA, IES, and USGS expect data to be shared by the time the publications for those results are published. NOAA expects data to be shared within 2 years of production. The other agencies do not have specifics for what or
when to share. A few did emphasize metadata and documentation in their public access plans. 
CDC specifies that at a minimum, shared data should have the common core metadata used by the federal government and USGS specifies that the shared data should include metadata using the FGDC Content Standard for Digital Geospatial Metadata. IES requires enough documentation with the data to support its reuse. The Smithsonian Institute took a unique approach to data. When submitting an accepted manuscript to Smithsonian Research Online, researchers should submit links to where the underlying data is shared. This is the only agency whose system and plans directly links the data to the publications.

A number of agencies already possess or are in the process of instituting repositories for data. The requirements for using their data repositories are not as straightforward as for using the publication repositories. Many of the sub-funding categories within the environmental- or health science-related agencies have specific requirements to deposit in particular repositories, though there are no agency-wide policies for data repositories. Interestingly, two agencies put the development of data catalogs into their public access plan. DOT and NIST are creating databases where authors submit the metadata to their data along with links to where the data can be accessed.

One feature that really stands out regarding DOE’s own public access plan is that they are utilizing a portal architecture. All other agencies, with the exception of NSF who will be using DOE’s infrastructure and USGS which is developing a similar infrastructure to DOE, utilize an archive architecture. Their repositories store a copy of the accepted manuscript or the publisher’s version for users to access even if they can access the publisher’s version on the publisher’s site. DOE’s repositories, SciTech Connect and PAGES, act more like portals, linking users to the best publicly-available version of the publication. The DOE Public Access Plan
proposes that “the best version of the article is the VoR [Version of Record] hosted by the publisher.” If the best version is available, then that version will be the only one public users can access. More specifically, if a publisher will be making their official copy publicly available within 12 months of publication date, then SciTech Connect and PAGES will link to the publisher’s website and keep the accepted manuscript in a dark archive. The dark archive is inaccessible to the public but preserved and used for text analytic tools in the databases. If the publisher will not be making their official copy publicly available within 12 months of publication date, then SciTech Connect and PAGES will open up the accepted manuscript to the public within 12 months of the publication date.

In order to obtain the information necessary for this architecture, authors are required by the DOE Public Access Plan to submit either the Digital Object Identifier (DOI) to the publisher’s version or the full citation with their accepted manuscript submission. DOE systems utilize CrossRef to get accurate information about the publication’s publisher and publication date. CrossRef is a DOI registry agency for journal articles. It provides an open infrastructure for querying citation metadata and linking to full text for articles produced by participating publishers (Dylla, 2014, p. 197). CrossRef has a new metadata field, FundRef, which collects information about the article’s funding source (Dylla, 2014, p. 196). DOE is also working with the Clearinghouse for the Open Research of the United States (CHORUS). This non-profit organization uses the CrossRef infrastructure and its partnerships with participating publishers to create tools to identify publicly-available journal articles from federal funding sources (Dylla, 2014). These tools will also help DOE identify gaps between the 20,000 – 30,000 papers expected to be produced by DOE funding per year and the papers submitted for public access (Kaiser, 2015, p. 167).
The approach of limiting public users to only the best public version available within 12 months of publication date has its advantages. Users don’t reference potentially incomplete papers when the version of record is available. Publishers get more traffic on their websites, encouraging them to make their versions available within 12 months. This approach also has its criticisms. Scott-Lichter (2014) is concerned that the 12 month embargo is too rigid. Other plans allow flexibility with their embargo periods by instructing authors to choose them during submission. Heather Joseph, executive director of SPARC, says this approach “will certainly make it much harder for users to do any kind of computational analysis, text or data mining on DOE-funded articles – the kind of innovative uses the White House directive was designed to encourage” (Peterson, 2014). As SciTech Connect and PAGES are not archives, users cannot download full text in bulk for use in text or data mining projects. The links to the full text are not consistent enough to automate and collect the full text necessary to perform those projects. Links could go directly to a PDF document or to a journal article’s splash page. In addition, users may not have the rights to pull journal articles for text and data mining purposes because copyrights vary by journal. PubMed Central, on the other hand, can control the rights to the accepted manuscripts in its archive and allow users to download what they need for text and data mining.

OSTI has another data initiative associated with but not implemented by the DOE Public Access Plan that influences the contents of SciTech Connect - they provide a Data ID Service. They will assign a DOI to DOE-funded datasets on request through DataCite. When using the service, the metadata to the dataset is collected and made available with the DOI in SciTech Connect, DOE Data Explorer, Google, and other search tools (Data Services, 2015). Again, this uses portal architecture because the datasets are not stored in SciTech Connect or DOE Data Explorer. Instead, users are provided a link to where the dataset is stored. This way of
handling datasets is similar to the way DOT and NITS are handling theirs in the data catalogs they are developing.

**Usability Test Methods**

A review of the relevant literature indicates that no usability studies have been published for SciTech Connect. However, there have been a number of such studies for PubMed and applications built to work with PubMed, a comparable application with a similar purpose. Researchers have conducted these studies to expand general knowledge of the applications’ users and to identify potential improvements. According to the authors of “Studying PubMed Usages in the Field for Complex Problem Solving: Implications for Tool Design,” “recent research and innovations related to PubMed and other MEDLINE information retrieval (IR) systems have expanded [the] knowledge about scientists’ information-seeking behaviors and relevant tool-based support” (Mirel et al., 2013, p. 876). Those same authors noticed a gap in PubMed usability studies - studies done “in the field” or within the normal context of a current user’s work. To address, they developed an investigation approach that observed 14 undergraduate students whose courses and laboratory work required them to query PubMed. Participants logged into a specific workstation that logged their actions, revealing many interesting behaviors. The top activity performed was copying and pasting citations. Participants had difficulty with bibliographic formatting and spelling. They would have benefited from a tool that suggested alternative words based off of possible misspellings or synonyms, alternative spellings, and acronyms for medical terminology in the search query. Participants also had difficulty finding relevance in the retrieved results. Without indicators for why search results were relevant to their query, participants spent a lot of time skimming abstracts to see if the particular paper was what they needed. Finally, the authors noted that certain features do exist that would have assisted the
participants, but participants were unwilling to break from the flow of their task at hand and those features should be integrated within their workflow (Mirel et al., 2013).

Another PubMed study focused on users in the field as well, but employed a more naturalistic method. Tang (2007) wanted to analyze the browse feature for faceted MeSH vocabulary queries to see if it was useful and whether users preferred it over the tradition single box query form in certain instances. Nineteen participants conducted their PubMed work through a proxy server on their own workstations over a ten-week period. For each session, there was a pre-search and a post-search questionnaire. Tang found that faceted search building is more effective than traditional searching in certain information problem settings and users chose the faceted search option when the information required was vague or when they were unfamiliar with the topics. She concluded that both traditional search forms and faceted search building tools are needed (Tang, 2007). SciTech Connect has a similar browse tool for faceted searching using controlled vocabulary.

One tool that SciTech Connect does not have is a word cloud summarizing the abstracts of an entire search. For that matter, neither did PubMed, so a team of researchers created one and tested it with users. They named the tool “PubCloud” and summarized their user test findings in a presentation at the 16th International Conference on the World Wide Web. Participants were supplied with a document containing the output of a PubCloud or the traditional search results list and asked a series of questions. The study demonstrated that the PubCloud provides users a visual overview of all results of the query and enables them to locate relevant literature that is otherwise hidden too deep in the search results list (Kuo, Hentrich, Good, & Wilkinson, 2007).

“A Usability Study of PubMed on a Tap User Interface for PDAs” evaluated PubMed on mobile devices. PubMed on Tap was a prototype developed for PubMed on PDA devices to help
healthcare practitioners gain the access they need to medical information at the precise moment and place of need. Nine participants were observed performing a set of tasks on PubMed on Tap in order to identify bugs in and potential improvements for the application. A facilitator supplied participants with a list of tasks to execute who were “encouraged to speak aloud their thoughts, actions, and expectations as they progressed through each scenario” (Alexander, Hauser, Steely, Ford, & Demmner-Fushman, 2004, p. 1412). This method is known as the “think aloud” method and is that which is employed in the current study.

Nielsen wrote in his 1993 book Usability Engineering, “thinking aloud may be the single most valuable usability engineering method” and he stands by this twenty years later (p. 195). In a 2012 article, “Thinking Aloud: The #1 Usability Tool,” Nielsen wrote “thinking aloud should be the first tool in your UX toolbox, even though it entails some risks and doesn’t solve all problems” (para. 1). Facilitators of think aloud user tests recruit representative users, give them representative tasks to perform, and listen as users vocalize their impressions of the application, reasons behind their behavior, frustrations, and misconceptions. The main benefit behind this method is that researchers can discover the why behind users’ behaviors. Usage logs can describe what users do, but they do not answer why they do it. This type of user test is also inexpensive and, according to Nielsen, likely to reveal robust results even with a bad test. However, the method does have its limitations. For example, it places participants in an unnatural setting where they may not behave normally and could be biased by the facilitator (Nielsen, 2012).

The think aloud method is used to meet many different types of goals and to gather various kinds of results. Krahmer and Ummelen (2004) split the goals behind think aloud usability tests into three different categories:
(1) To find evidence for models and theories of cognitive processes...(2) To discover and understand general patterns of behavior in the interaction with documents or applications in order to create a scientific basis for designing them...(3) to test specific new documents or applications in order to troubleshoot and revise (p. 105).

The number of subjects to recruit for a usability test depends on its purpose and goal. For the third goal listed by Krahmer and Ummelen, testing specific applications to identify areas that need improvement, four to five subjects would probably suffice. In a study conducted by Nielsen (1994), “Estimating the Number of Subjects Needed for a Thinking Aloud Test,” it was found that four to five subjects identified 75% of usability problems. More subjects will indeed find more problems, but with diminishing returns. If the goal is to identify usability issues of an application during the development lifecycle, finding 75% of problems is a reasonable goal and more testing should only occur after those problems have been addressed (Nielsen, 1994).

Although, if the goal is more scholarly in nature, such as building a model for design or for understanding human-computer interaction, more participants are required. Emanuel (2013) argued in the literature evaluation, “Usability Testing in Libraries: Methods, Limitations, and Implications,” that problems with sampling is a limitation preventing some usability studies from being a valid form of scholarly research. It is difficult to acquire a representative group with only five subjects who are often self-selected. Researchers need to keep testing more subjects until a level of saturation for each target population is reached (Emanuel, 2013, p. 207 - 208).

NØrgaard and Hornbaek (2006) closely inspected how usability tests are practiced during the development cycle at seven companies in “What Do Usability Evaluators Do In Practice?: An Explorative Study of Think-Aloud Testing.” They observed 14 test sessions which included setting up the test, facilitating the test with the user subjects, and recording and analyzing the test
results. They found that test facilitators often looked for confirmation of known issues as the main focus. Generally, there is a lot of interference during the test, including technical problems and users not showing up for their scheduled time. Many of the questions asked are really reminders for the subject to think aloud and some questions are hypothetical and therefore difficult for users to understand and meaningfully answer. Further, facilitators do not normally conduct careful and systematic analysis of the results directly following the subject’s session and have a tendency to attempt laboratory-like scientific standards even though that is not called for in the situation (Nørgaard and Hornbaek, 2006).

There are many different protocols for conducting a think aloud test. Four different studies comparing the results of several protocols found virtually no differences in the number of results or the number UX problems identified, although there were differences in users' task performance and their overall satisfaction with the application. Krahmer and Ummelen (2004) compared the traditional approach developed by Ericsson and Simon, which prevents the facilitator from interfering or coaching the subject during the test, to a speech communication protocol that involves more acknowledgments from the facilitator and the possibility for subjects to ask for clarification and get encouragement to make up for the unnatural test setting. Users’ task performance was better in the speech communication protocol, but both protocols discovered a similar number of problems (Krahmer & Ummelen, 2004). Olmsted-Hawala, Murphy, Hawala, and Ashenfelter (2010) compared the same two styles from Krahmer and Ummelen’s study with a third type, a coaching protocol that had an even greater level of interaction between the subject and the facilitator. They found that participants in the coaching protocol were more satisfied by the overall experience of the application than participants in the other two protocols, but that all three discovered a similar number of problems.
van den Haak and de Jong (2003) compared the common approach of asking subjects to think aloud as they perform their tasks, or the concurrent think aloud method, to asking subjects to work through tasks in silence and then think aloud about their experience after all tasks were completed, referred to as retrospective think aloud, in “Exploring Two Methods of Usability Testing: Concurrent Versus Retrospective Think-Aloud Protocols”. Again, the results of the two protocols were similar, though participants in the traditional approach were less successful in their task performance than participants allowed to execute the tasks in silence without the distraction of thinking aloud (van den Haak & de Jong, 2003). The next year, van den Haak, de Jong, and Schellens (2004) conducted another study of those two protocols along with a third, constructive interaction in “Employing Think-Aloud Protocols and Constructive Interaction to Test the Usability of Online Library Catalogues: A Methodological Comparison”. This third protocol is also known as “co-discovery learning” because two subjects work together to complete tasks and are asked to think aloud to each other while the facilitator observes. Again, all three protocols had comparable results for number and relevance of usability problems identified. However, the researchers argued the value of the concurrent think aloud protocol over the other two because of time and cost. The retrospective think aloud protocol took longer because subjects provided their feedback following the tasks instead of during. Constructive interaction costs more because one needs twice as many participants for the study.

Database Analysis

**Personas**

Personas are models that take the thoughts, behaviors, and motivations of a set of users and apply them to a single realistic user. Personas help narrow the focus of what a product should do and how it should behave. Cooper, Reimann, Cronin, and Noessel (2014) argue in
About Face: The Essentials of Interaction Design that products should be designed for specific types of users instead of all users.

To create a product that must satisfy a diverse audience of users, logic might tell you to make its functionally as broad as possible to accommodate the most people. This logic, however, is flawed. The best way to successfully accommodate a variety of users is to design for specific types of individuals with specific needs (Cooper, Reimann, Cronin, & Noessel, 2014, p. 62).

Personas are the model representation of those specific individuals.

This argument can be extended to this usability study. While SciTech Connect was created for the general public, attempting to focus on the needs of all users and make redesign recommendations for every possible use case would only yield unfocused, non-priority results. To narrow the scope of my analysis and user study, I identified types of users who represent a larger set of key users. I created four personas which represented power users, occasional users, and users who don’t realize they are users. The four personas also represent workers in industry, academia, and secondary public schools. Industry and the academic science community are acknowledged as target populations to benefit from public access in the OSTP memorandum “Increasing Access to the Results of Federally Funded Scientific Research” (Holdren, 2013, p. 1). Young (2015) wrote about the possibilities databases like SciTech Connect and PAGES offer to secondary public schools in her opinion piece “Public Access to Federally Funded Scientific Research: Building Communities of Collaboration”. Young believes that access to scholarly scientific literature can supplement programming gaps in curriculum development and usher in new manipulatives and learning aids to further expand students’ critical thinking development. Collaborations among STEM teachers, librarians,
and other stakeholders that leverage public access to peer-reviewed scientific publications can foster innovative, cross-curriculum pedagogy that will help attract and retain middle and high school students' involvement in STEM (Young, 2015, para. 6).

It should be noted that there are few key user groups left out of the personas and therefore outside the scope of this study. Users with visual, physical, and other disabilities that require the use of special software to interact with SciTech Connect are not addressed by this study. Neither are users who prefer mobile devices for their research activities.

*Persona 1*

**ENGINEERING INDUSTRY WORKER**

“I don’t have access to major journals and databases in my field.”

Janet Montoya is a 36 year old electrical engineer at a company in Santa Fe, NM. She needs peer reviewed scholarly publications and technical reports for her work. Her company does not provide access to journals or databases and only purchases specific items on request. She relies on open access sources to keep current in her field.

**Key SciTech Connect Activities**

- Set up email alerts
- Save searches and records
- Browse subjects and search by terms
- Access full text
- Export citation metadata to EndNote
Persona 2

HIGH SCHOOL CHEMISTRY TEACHER

“I need to better prepare my AP students for college.”

Nancy Smith is a 30 year old Chemistry Teacher at a high school in Albuquerque, NM. She needs access to peer reviewed scholarly publications to enhance her AP program. The school library doesn’t have a budget for her needs so she relies on open access sources.

Key SciTech Connect Activities

• Browse subjects and search by terms
• Access full text

Persona 3

PHYSICS UNDERGRADUATE STUDENT

“I find my reading assignments with google and the library search.”

John Pearson is a 21 year old undergraduate student at the University of New Mexico. He is assigned readings and papers on various topics that he finds with Google and his library’s search discovery tool. Sometimes these searches lead him to records and full text in SciTech Connect.

Key SciTech Connect Activities

• Access records and full text through links from other search portals (library catalogs, Google, Web of Science, etc.)
• Export or copy citations
A COGNITIVE WALKTHROUGH

A cognitive walkthrough is one technique for evaluating an application without users. It can be used in situations where users aren’t available for testing or when a user test isn’t necessary yet. A cognitive walkthrough can reveal many obvious issues a user might find during a user test, so those errors can be addressed before users are brought in for testing. An evaluator steps through specific user tasks while noting possible user actions, thoughts, behaviors, and motivations (Lewis & Rieman, 1994, Ch. 4). Below are the tasks I completed in my cognitive walkthrough of SciTech Connect and the issues I identified.

Tasks for Cognitive Walkthrough

- Browse engineering subjects
- Search for recent records about a specific topic
• Find datasets about a specific topic
• Export citations from records and search results
• Make an account
• Save records
• Save a search
• Set up an alert
• Export metadata and full text for a search

Issues Identified

1.) The tabs “Subject Details” and “Sample Records” in the browse result window are on the bottom and easy to overlook. In some browser windows they are out of sight (Screenshot 1).
2.) When you browse, there are no labels for the subjects in the left column until you filter deep enough that there are no sub-categories (Screenshot 2). Then there are four labels: “no subcategories,” “narrower terms,” broader terms,” and “related terms”. The two views are inconsistent and first lacks labeling (Screenshot 3).
Screenshot 2

Scitech Connect — Explore by Subject

No label for subcategories

This category includes particle and radiation detectors and monitors (such as radiation dosimeters, nuclear spectroscopic instrumentation, high-energy physics instrumentation, and radiometric instruments) as well as other nuclear science-related instrumentation such as flowmeters, pressure gages, and heat sensors. Radiation effects on instruments and electronic systems are also categorized here.

Matching Records: 26506
Result Type: Everything
Document Type: All Document Types
Publication Date: MM/DD/YYYY to MM/DD/YYYY

Go to "Instrumentation Related to Nuclear Science and Technology" »
3.) Narrower terms, broader terms, and related terms are displayed in two places in the browse window and when selected, behave differently. The terms in the left column initiate a new search: new terms are displayed and the selection is added to the breadcrumbs at the top (Screenshot 4). The terms at the bottom narrow the number of records but new terms do not appear and the term is not added to the breadcrumbs (Screenshot 5).
A USABILITY STUDY OF SCITECH CONNECT

Screenshot 4

Screenshot 5

When "Hot-Wire Gages" is selected from the left column, the term is added to the breadcrumbs and new browse terms appear in the left column and below the search form.
4.) The clear button on the browse window only clears the date fields. I expected it to clear all of the user data in the form (Screenshot 7).
5.) There are two search options, semantic search and term search. The only explanation for the difference is buried in the FAQ page. This is a feature not common to databases (Screenshot 8).
6.) The advanced search form has a field for “Identifier Numbers” (Screenshot 9). This appears to search DOI, report number, and contract number. However, there was no explanation of what that field searches on the form or in the FAQs.
7.) In the author select feature of the advanced search, after you add an author, the “add” button is changed to text that looks hyperlinked and still says “add” (Screenshot 10). A user has to select the tab “Selected Authors” to remove the selection (Screenshot 11).
8.) On the search results view, the download item link says “Full Text Available” and is small and off to the right of the record information (Screenshot 12). Without an icon, a button, or actionable language this could be overlooked.
9.) On the search results view, authors are in a green font but are not hyperlinked (Screenshot 13). Because the font isn’t black like the text and because other text areas of the site, such as word clouds, use green font for hyperlinks, it appears to be a hyperlink to search for that author.
10.) The full record display does not show what number a record is in the search results on the breadcrumbs or anywhere else on the page (Screenshot 14 and 15).
A USABILITY STUDY OF SCITECH CONNECT

Screenshot 14

[Image of a webpage showing a search results page for superconductivity in SciTech Connect]

51. Nodal to nodeless superconducting energy-gap structure change concomitant with Fermi-surface reconstruction in the heavy-fermion compound CeCu6

52. Sensitivity of nonlinear photoresonance to resonance substructure in collective excitation

Collective behavior is a characteristic feature in many-body systems, important for developments in fields such as magnetism, superconductivity, photonics and electronics. Recently, there has been increasing interest in the optically nonlinear response of collective excitations. Here we demonstrate how the nonlinear interaction of a many-body system with intense XUV radiation can be used as an effective probe for characterizing otherwise unresolved features of its collective response. Resonant photomixing of atomic xenon was chosen as a case study. The excellent agreement between experiment and theory strongly supports the prediction that two distinct poles underlie the giant dipole resonance. Our results pave

April 2015, Nature Publishing Group
11.) The full record display does not have previous and next buttons to navigate between records in the search. A user has to go back to the search results and then select a new record.

12.) The pop-ups for when you try to do something not available (in-document search and save to my library) provide clear feedback to users, but you have to click the action, either “In-Document Search” or “Save to My library,” again to get the pop-up to disappear. There is no “x” for the pop-up and clicking elsewhere in the browser window doesn’t remove the pop-up (Screenshot 16 & 17).
13.) The filter by author and subject on the search results page only displays the top choices (Screenshot 18).
14.) DOIs to non-journal articles, older journal articles, and newer journal articles all behave differently and the labeling is inconsistent. For non-journal articles, DOI either means the DOI location of that record’s splash page in SciTech Connect or the DOI link to the full text (Screenshot 19). For older journal articles, DOI means the DOI to the publisher’s final version of the article (Screenshot 20). For newer journal articles, there is clearer labeling and links to the publisher’s DOI and to the full text accepted manuscript stored in SciTech Connect (Screenshot 21).
Screenshot 19

The DOI to this non-journal article item goes to the splash page for the record and displays the message above in the green pop-up.

Screenshot 20

The DOI to this journal article goes to the publisher's version of the item, which is different than the pre-print version that displays if you select "View Full Text".
15.) The CSV, Excel, and XML export files for search results don’t include the link to the full text item. The links in this field route the user to the item’s splash page in SciTech Connect. The full text can be stored in many different locations, such as SciTech Connect’s database, a DOE site’s servers, an Open Access journal, or a public repository. This makes it difficult for a user to export full text.

16.) There is no information in the database or its help files that explains how often users who set up an alert on a search will receive an email.

17.) The database contains multimedia items that you can download. The full record display metadata does not describe the file type or if the item is video or audio (Screenshot 22).
18.) Multimedia items display empty word clouds and “more like this” suggestions on their full record display (Screenshot 23). Other non-text items, datasets and software, do not include these features on the page because they don’t apply.
19.) When a user initiates a new search using the search form in the header, the database will keep all of the filters and advanced search fields from the previous search (Screenshot 24 and 25).
A search that yields no results does not give any helpful information to users (Screenshot 26).
Heuristic Analysis

Heuristics are broad guidelines for interface design. For this analysis activity, I used Jakob Nielsen’s 10 usability heuristics for user interface design to evaluate SciTech Connect (Nielsen, 1995).

1.) Visibility of system status. The database makes good use of breadcrumbs with a few exceptions. There is some inconsistency with the browse feature’s breadcrumbs that was identified in cognitive walkthrough issue #3. In addition, on the full record view, the breadcrumbs do not display what number the record is in the search.

For records that don’t have an electronic item attached, the database gives more information depending on the record type. Software records display a link to a software request form. Documents like technical reports and journal articles display links to search Google Scholar and WorldCat on the title. If the accepted manuscript to a journal article is in an embargo period, the database displays when it will become available. However, there is no information for
users when the database hides the accepted manuscript because the publisher made their version Open Access.

2.) **Match between system and the real world.** In general, the labels and guidance in the database match commonly used terms. They match how a person would think about navigating a physical collection. There were only a few labels that users might find confusing: identifier number, semantic search, and term search. These labels do not have a common usage outside of the database.

3.) **User control and freedom.** Users have four options for controlling their location in the database. They can use the forward and back buttons in their browser while maintaining all of their search selections, select a link in their breadcrumbs, click the database logo on the top-left of the page to start over, or start a new search in the basic search bar that remains in the header throughout the application. There is an issue with the last option. If a user initiates a new search in the basic search bar from anywhere except the homepage, the database will maintain all filters or advanced search field selections the user made in the previous search.

4.) **Consistency and standards.** For the most part, the database provides consistency with labels and behaviors with a few issues. As noted in cognitive walkthrough issue #14, the labels and behaviors for full text and DOI’s to journal article records are inconsistent. The database calls setting up a regular email with new records for a search an “alert” in the account management feature of the database, but when a user saves a search and is given the opportunity to set an alert from the search results page, the database does not use that label. Finally, there are issues when selecting different terms in the browse features as noted in cognitive walkthrough issue #3.
SciTech Connect’s basic functionally for searching, filtering, and viewing records matches closely with that of popular databases like Web of Science and library catalogs. However, there are two behaviors with filtering that do not match other databases. First, the database automatically filters on full text, instead of searching for everything and letting the user set the filters. Second, the database only allows users to filter on the top authors and subjects in the search.

5.) Error prevention. In the text search forms, the character “%” breaks the application and routes users to a browser message “Bad Request”. Also, the date forms do not provide a date picker making it easy to enter in data the application does not recognize.

6.) Recognition rather than recall. Only a few features in the database might require users to remember information in order to use them. The in-document search tool is not easy to locate on a full record display and the links to save a search, save a record, and export metadata are small and easy to overlook. Setting up email alerts and saving searches and records requires several steps including setting up a user account.

7.) Flexibility and efficiency of use. Power users have several options to change how they view data and interact with the database. They can switch from the regular search results view to a “Detail View” where they view the results as a grid and can decide which metadata fields display in the grid as well as how many results to view per page. If a user has an account and is logged in, they can set their preference for the grid view that automatically loads when searching. Also, if a user prefers, they can use the browse feature to find their results instead of the traditional search.

8.) Aesthetic and minimalist design. The database keeps to a minimal and appropriate amount of information and options for users, excluding the basic search. Here users are asked to
decide the format of information for their search and well as to selection either a semantic search or a term search.

9.) Help users recognize, diagnose, and recover from errors. The database provides no information or suggestions to users whose search retrieved zero results. Also, some features, like the citation builder, do not work consistently across browsers and browser versions. Instead of seeing an error message or warnings about supported browsers, users see a blank pop-up.

10.) Help and documentation. The database provides minimal documentation, a FAQ page, a video about using the browse feature, and a video about making and using accounts. The contact us page includes for multiple ways to contact the database owners and helpful information about when their business hours are and what their response time is for inquiries. Some users would benefit from more information and instruction about advanced searching.

Comparisons to Similar Applications

I selected two similar public databases to compare to SciTech Connect. I selected PubMed Central (http://ncbi.nlm.nih.gov/pmc) because it is also run by a federal agency, NIH, who uses the application to comply with the same federal public access mandates that DOE has. I selected eScholarship (http://escholarship.org) because they have a subset collection for Lawrence Berkeley National Laboratory who also deposits those same papers into SciTech Connect. However, eScholarship is run by an academic institution, the University of California, and has a wider collection scope and range of features. Below are features from the databases that SciTech Connect does not have and could benefit from.

1.) Both databases provide statistics for the entire database and eScholarship provides statistics for individual records (Screenshot 27, 28, and 29).
Screenshot 27

The image shows a screenshot of the SCITECH CONNECT website. The page displays statistical information indicating that 3.8 MILLION Articles are archived in PMC. The content is provided by various partners, with specific numbers for Full Participation Journals, NIH Portfolio Journals, and Selective Deposit Journals: 1816, 322, and 3621 respectively.
A USABILITY STUDY OF SCITECH CONNECT

Screenshot 28
2.) The in-document search for eScholarship is embedded in the full record display and loads with the user’s search terms highlighted (Screenshot 30). SciTech Connect users have to select a tab to see the in-document search form and then enter new keywords to search the record’s full text (Screenshot 31).
Screenshot 30

Screenshot 31

3.) PubMed Central displays a user’s recent search history on the advanced search form, the search results page, and the full record display (Screenshot 32 and 33).
4.) Both databases have a simpler search form and return all results without filtering anything the user did not specify (Screenshots 34, 35, and 36). SciTech Connect’s basic search form asks users to filter by format and decide if they want a semantic or a term search. By default, SciTech Connect returns only full text items and filters out citations, multimedia, datasets, and software instead of returning everything and letting the user decide how they want to filter the results.
5.) PubMed Central has search suggestion dropdowns as users start to type in the search forms (Screenshot 37). It also has viewable indexes in the advanced search form (Screenshot 38).
6.) Both databases provide social network sharing options (Screenshot 39 and 40).
7.) Both databases provide information when a search yields no results and suggests new search terms (Screenshot 41 and 42). PubMed Central suggests new search terms when the search returns low results (Screenshot 42).

**Screenshot 41**

![Screenshot 41](image1)

**Screenshot 42**

![Screenshot 42](image2)

8.) PubMed Central provides information about public access on its homepage (Screenshot 43).
User Study

Twelve participants were recruited and completed a user test session between May 20\textsuperscript{th} and July 1\textsuperscript{st}, 2016. Test sessions took up to an hour to complete and were conducted in private areas of public library sites in Albuquerque and Los Alamos, NM. Participants were asked to complete 20 tasks in SciTech Connect while verbalizing their thoughts and actions and to fill out a ten-question survey about their experiences based on the System Usability Scale (System Usability Scale (SUS), n.d.). A copy of the protocol can be found in Appendix B: User Test Protocol. Participants likely had average to advanced computer and literacy skills because of how they were recruited, either through social media via Twitter and Facebook promoted posts or by being approached at Albuquerque Public Library’s Main Library site.
Results

Participants had the most difficulty successfully completing three tasks: “3. Search for everything in the database about renewable energy,” “4. Narrow these results down to items by the author R. Meadows,” and “19. Go back to the record. Does it have a link to the journal publisher’s website?” For each of these tasks, five or more participants failed to complete it successfully. There were four tasks that participants had moderate difficulty completing: “2. Find something that was recently added to the database,” “6. Find a video about clean energy,” “10. Find all records created by people at Argonne National Laboratory in the last 5 years,” and “16. Find a journal article about coal power published in the last year.” For each of these tasks, three participants failed to complete it successfully (Chart 1).

Chart 1: Task Success – The number of participants who completed each task successfully.

<table>
<thead>
<tr>
<th>Task</th>
<th>Participants Completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What are your first impressions of the database? What kind of content does it contain?</td>
<td>12 / 12</td>
</tr>
<tr>
<td>2. Find something that was recently added to the database.</td>
<td>9 / 12</td>
</tr>
<tr>
<td>3. Search for everything in the database about renewable energy.</td>
<td>4 / 12</td>
</tr>
<tr>
<td>4. Narrow these results down to items by the author R. Meadows.</td>
<td>7 / 12</td>
</tr>
<tr>
<td>5. Select one of the records that has a full text pdf. See if the document contains the terms “wind turbines”.</td>
<td>10 / 12</td>
</tr>
<tr>
<td>6. Find a video about clean energy.</td>
<td>9 / 12</td>
</tr>
<tr>
<td>7. Send the record for the video to my email: <a href="mailto:huphoff2@illinois.edu">huphoff2@illinois.edu</a></td>
<td>12 / 12</td>
</tr>
<tr>
<td>8. Find a scientific dataset about solar radiation.</td>
<td>10 / 12</td>
</tr>
<tr>
<td>9. Find all other records by the first author of this scientific dataset.</td>
<td>11 / 12</td>
</tr>
<tr>
<td>10. Find all records created by people at Argonne National Laboratory in the last 5 years.</td>
<td>9 / 12</td>
</tr>
</tbody>
</table>
11. Sort these records by date with the most recent displaying first. 12 / 12
12. Go to the last page of the search results. 12 / 12
13. Export the entire search results into a spreadsheet. 12 / 12
14. Find the technical report assigned the number PNNL-24460. 11 / 12
15. Find the citation for this report, the description you would copy and put in the references of a school paper. 12 / 12
16. Find a journal article about coal power published in the last year. 9 / 12
17. What journal published this article? 10 / 12
18. Access the full text for this article. 11 / 12
19. Go back to the record. Does it have a link to the journal publisher’s website? 6 / 12
20. If you had questions about this database or wanted help using it, what would you do? 12 / 12

Much of the trouble participants had with tasks were linked to one of the seven major issues identified in the user study, or an issue that at least half of the study participants identified or experienced (Chart 2).

*Chart 2: Major Issues – Issues that at least half of the study participants identified or experienced and the number of participants that identified or experienced it.*

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>When initiating a new search from the basic search form in the header, the database keeps the search filters and advanced search form input from the previous search. The participant did not notice and it interfered with their search task. 10 / 12</td>
</tr>
<tr>
<td>2.</td>
<td>Participant used an incorrect format for the date field in either the browse feature or the advanced search form or said that they would prefer a date picker. 10 / 12</td>
</tr>
</tbody>
</table>
3. Participant tried to use the author filter expecting to find the author they were looking for instead of a selection of the top cited authors.

4. Participant did not notice the in-document search feature on the full record view when asked to see if the full text contains the terms “wind turbines.”

5. Participant commented that they couldn’t tell what kind of multimedia they found in the search. Participant was looking for that information in either the metadata, the facets, or the icon next to the link to access the item.

6. Participant had difficulty finding the link to the journal’s website because they did not know what the DOI was supposed to be or they did not notice the link off in the corner of the record.

7. Participant said they expected the basic search to return everything and not default to full text.

\[ \text{Screenshot 44: Basic Search Form} \]

The basic search form followers’ users throughout the application by remaining in the header and instructs users to “Start new search” here (Screenshot 44). Yet, if users enter search terms in the form and select “find” the database will keep any advanced search field values and search filters used in the previous search. Six participants had difficulty completing a task because they did not notice that old values were interfering with their new search. Once observed, one participant suggested adding an option to edit your previous search while making the basic search form initiate a new search.

Six participants expected the basic search form to return everything and not limit to full text. Once participants noticed and started to understand the extra selections on the basic search
form, they expressed other difficulties with the form. Three participants were not sure what citation versus full text meant. One suggested that citations meant journal articles. However, citation here means that it is a record without a link to the item and that users have to request the item or locate it elsewhere. Two participants believed that by clicking “dataset” or one of the format selections above the form that is would take you to another webpage with information about what a “dataset” means. One participant believed that by clicking “dataset” or one of the format selections above the form that it would immediately initiate a search for all datasets in the database. Three participants said they didn’t know what semantic versus term search meant.

Screenshot 45: Date Field on the Advanced Search Form

Ten participants used an incorrect format for the date field in either the browse feature or the advanced search form or said that they would prefer a date picker (Screenshot 45). One participant noticed that the form allows you to select an end date that is before the start date. One participant did not like that they had to select a full date and would like the option of only selecting a year.
The advanced search form for the database is a popup with 15 rows of possible selections. Therefore, users cannot see the whole form on many monitors. Two participants commented that they did not like having to scroll down to use the pop-up. One participant commented that they would prefer the advanced search to be a regular webpage. One participant accidently closed the pop-up by clicking off of the form while they were still entering search values.

Two participants said they disliked that you could not change the format selections, full text, citations, multimedia, datasets, software, and everything, from the advanced search form. Users have the choice of changing the format in the basic search form before they select to open the advanced search form pop-up or after they click search and see the results. Also, on the form, there are drop-down selections for type and site. One participant said they didn’t like that you couldn’t select more than one site for your search, especially since some sites, like Argonne National Laboratory, had more than one selection for sub-sites of the main laboratory.

There was some confusion on labels in the advanced search form which are not explained anywhere in the database. Five participants said they did not know what update date versus publication date meant. One participant was unsure of what research organization meant and how that differed from sponsoring organization. Another asked what identifier number meant.
In the search results list, users can filter by author. Only a selection of the top cited authors in the search are displayed with the number of times they are cited in parenthesis next to their name (Screenshot 46). Ten participants tried to use the author filter expecting to find the author they were looking for listed there instead of a selection of the top cited authors. One participant wanted the option to view the authors listed in the facet alphabetically as well as by number of citations.

The author facet is one of only two facets in the database. The other is subject. Three participants wanted the option to limit the search results by year. One participant wanted the option to limit the search results by type, a predefined field in the advanced search form for technical report, journal article, patent, etc.

Participants made other suggestions to the layout and functionality of the search results list. Two participants wanted to see the full date and not just the year on records’ brief display. One participant expected the author name on the brief display to be a hyperlink to a search on everything by that author, which is how author names behave in the full record display. One
participant did not understand why a publisher was displayed for some records and wanted to see what type of record it was from the brief display. Two participants thought that it was required to have an account to use the “save your search” tools which allow users to export their search results in Excel, CSV, and XML formats. One participant wanted the option to export their search as a pdf document.

*Screenshot 47: Page Scroll Tool*

On the search results page, users see a default of ten search results and buttons to click to the previous page, to the next page, and to display a slider to move between search result pages. One participant asked for a “go” label on the slider tool because they did not know what “>>” meant. Another participant expected two extra buttons: “<<” to move to the first page of results and “>>” to move to the last page of results (Screenshot 47).

*Screenshot 48: In-Document Search Tool on Full Record Display*

On the full record display webpage, there is an in-document search tool which is accessed by clicking the tab on the upper right (Screenshot 48). Eight participants did not notice the in-document search feature on the full record view when asked to see if the full text contains the terms “wind turbines.” One participant who did notice and use the in-document search
feature clicked the browser’s back button expecting to be taken back to the record information page, but instead went back to the search results.

*Screenshot 49: Full Record Display for a Multimedia Record*

When searching for a video in the database, eight participants commented that they could not tell what type of multimedia records they were finding in their search. They were looking for that information either in the metadata, the filters, or the icon next to the “view multimedia” hyperlink. Information about the file format of the item is not listed anywhere on the full record view (Screenshot 49). Three participants did not know what multimedia meant.

As discussed in issue 14 identified in the cognitive walkthrough, full text links and DOI links to non-journal articles, older journal articles, and newer journal articles all behave differently and the labeling is inconsistent. When participants were asked to find a journal article and then locate a link to the publisher’s website for that article, six did not complete the task successfully. Seven participants had trouble locating the link to the journal’s website for the article, either commenting that they didn’t know what the DOI was, clicking the DOI and being re-routed back to the records page in SciTech Connect, or not noticing the DOI link in the top
right-hand corner of the webpage. Two participants expected the link to be listed in the metadata on the left-hand side of the webpage like all of the other record information.

Study participants had a few other suggestions for the full record display. Two participants expected an email icon when they were looking for a tool to email the record. Two participants wanted a comment field for the email record tool in case they were sharing the email with someone else. For the citation tool, because it was a pop-up, three participants used a keyboard shortcut to highlight the citation and highlighted everything on the webpage, including all of the text outside of the pop-up.

When asked how they would contact the database administrators, all participants found the contact us link the header. Two participants noticed the pop-ups throughout the site asking for feedback. Three participants expected a contact information to also be in the application’s footer. One participant commented that they liked that information about business hours and average turnaround times for inquiries were listed in the contact us page.

The user study did not collect much information about the browse feature. Only three participants used browse instead of search for one of their tasks. One of these participants expected that clicking the subject link on the homepage would immediately initiate a search for that subject instead of taking them to an intermediate browse step with more options before seeing records.

After all tasks were completed, participants were asked to fill out a ten question survey about their experience. Each question asked for a number between 1 and 5, with 1 meaning strongly disagree and 5 meaning strongly agree. The average score for each question was above or below 3, neutral, as desired for each question except for question 1, “I think that I would like
to use this system frequently” (Chart 3). Because potential users were not specifically targeted for this study, that question does not solicit a red flag about the database.

Question 4, “I think that I would need the support of a technical person to be able to use this system,” generated the best response from participants with an average score of 1.67 towards the disagree end of the scale. Three questions have room for improvement: “5. I found the various functions in this system were well integrated,” “7. I would imagine that most people would learn to use this system very quickly,” and “9. I felt very confident using the system.” The answers to these questions were .5 or closer to neutral (Chart 3). Addressing the issues identified and suggestions made during the user study could improve those scores.

*Chart 3: Final Survey Average Scores – Participants were asked to give a number between 1 and 5, with 1 meaning strongly disagree and 5 meaning strongly agree. The average scores are rounded to two decimal places.*

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. I think that I would like to use this system frequently</strong></td>
<td>2.92</td>
</tr>
<tr>
<td><strong>2. I found the system unnecessarily complex</strong></td>
<td>2.09</td>
</tr>
<tr>
<td><strong>3. I thought the system was easy to use</strong></td>
<td>3.75</td>
</tr>
<tr>
<td><strong>4. I think that I would need the support of a technical person to be able to use this system</strong></td>
<td>1.73</td>
</tr>
<tr>
<td><strong>5. I found the various functions in this system were well integrated</strong></td>
<td>3.33</td>
</tr>
<tr>
<td><strong>6. I thought there was too much inconsistency in this system</strong></td>
<td>2.25</td>
</tr>
<tr>
<td><strong>7. I would imagine that most people would learn to use this system very quickly</strong></td>
<td>3.45</td>
</tr>
<tr>
<td><strong>8. I found the system very cumbersome to use</strong></td>
<td>2.33</td>
</tr>
<tr>
<td><strong>9. I felt very confident using the system</strong></td>
<td>3.5</td>
</tr>
<tr>
<td><strong>10. I needed to learn a lot of things before I could get going with this system</strong></td>
<td>2.17</td>
</tr>
</tbody>
</table>
Redesign Recommendations

The following redesign recommendations for SciTech Connect are intended to address the issues identified and the suggestions made in the cognitive walkthrough, heuristic analysis, and comparison to other databases activities as well as the user study. The redesign moves away from using pop-ups and tabs for important features like the advanced search form and the in-document search tool. It attempts to normalize the behavior of the search to be more like similar applications which would respond to some of the issues identified in the user study. Finally, the redesign is an effort to make labeling and behavior more consistent throughout the application and understood by users.

Basic Search Form

Mockup 1: Basic Search Form
Homepage

Mockup 2: Homepage

![SciTech Connect homepage mockup](image-url)
Advanced Search Form

Mockup 3: Advanced Search Form Webpage

1. Move the advanced search form from a pop-up to a webpage.
2. Add a link, “Search Help,” that takes users to search information on the FAQ page.
3. Add a field for format.
4. Add help icon next to confusing fields that explain their definition to users when selected.
5. Replace the text fields for dates with date pickers. Ensure that users cannot select an end date earlier than a start date.
6. Add recent search history that initiates the search when selected.
Mockup 4: Advanced Search Form’s Author Select Tool

1. Button to select people picker.
2. Replace the "Add" hyperlink that displays after an author is selected with a "Remove" hyperlink that removes the author if selected.
A USABILITY STUDY OF SCITECH CONNECT

Search Results List

Mockup 5: Search Results List
Mockup 6: Author Facet Pop-Up

1. Add link “Show more” to Author/Creator and Subject facets that opens up a modal window with all results.
2. Allow users to click through all results.
3. Allow users to view the list by numerical sort or by alphabetic sort.

4. Performance Analysis of Transportation Models Simulating Solar Radiation on Inclined Surfaces
   - Xia, Y. (409)
   - Wagner, P. (409)
   - Antoni, J. (407)
   - Annoi, A. (405)

Transportation models have been widely used in the solar energy industry to simulate solar radiation on inclined...
Mockup 7: Search Results List for Zero Results

1. No results found for your search.
2. Did you mean solar radiation?
3. Try using fewer keywords to start, then refine your search using the links on the left.

1. Notify users when their search yields no results.
2. Give users search suggestions based on misspellings and term variations when their search yields no results.
3. Offer users tips for modifying their search when their search yields no results.
A USABILITY STUDY OF SCITECH CONNECT

Full Record Display

Mockup 8: Full Record Display


Solar radiation can be computed using radiative transfer models, such as the Rapid Radiation Transfer Model (RRTM) and its general circulation model applications, and used for various energy applications. Due to the complexity of computing radiation fields in aerosol and cloudy atmospheres, simulating solar radiation can be extremely time-consuming, but many approximations—e.g., the two-stream approach and the delta-M truncation scheme—can be utilized. To provide a new fast option for computing solar radiation, we developed the Fast All-sky Radiation Model for Solar applications (FARMS) by parameterizing the simulated diffuse horizontal irradiance and direct normal irradiance for cloudy conditions from the RRTM runs using a 16-stream discrete ordinates radiative transfer model. The solar irradiance at the surface was simulated by combining the cloud irradiance parameterizations with a fast clear-sky model, RESTZ. To understand the accuracy and efficiency of the newly developed fast model, we analyzed FARMS runs using cloud optical and microphysical properties retrieved using GOES data from 2008-2012. The global horizontal irradiance for cloudy conditions was simulated using FARMS and RRTM for global circulation modeling with a two-stream approximation and compared to measurements taken from the U.S. Department of Energy’s Atmospheric Radiation Measurement Climate Research Facility Southern Great Plains site. Our more <

Authors: Xie, Yu; Sengupta, Manjit

Publication Date: 2018-06-01

OSTI Identifier: 1390052

Report Number(s): NREL/CP-5000-66596

DOI: 10.1002/2018E000048

Word Cloud
accuracy agency all data science energy atmosphere available clear cloud government laboratory model models national solar stream surface time transfer two used using

More Records Like This
1. Fast Radiative Transfer Model for Solar Resource Assessment
Xie, Yu; Sengupta, Manjit (Jun. 2016)
2. Application of Improved Radiation Modeling to General Circulation
Michael J. Isom (Apr. 2011)

Some links on this page may open in a new browser tab.

1. Add the number of the record in the search.
2. Add previous/next buttons to navigate through the records in the search results.
3. Remove the DOI to the record in SciTech Connect from the full text options.
4. Add the DOI to the record in SciTech Connect to the metadata.
5. Add an icon indicating what type of file is available for the record.
6. Add an email icon next to the "Send to Email" link.
7. Add social media sharing icons.
8. Move the in-document search toolbar hidden in a tab to the sidebar.
9. Add metrics for the record and an option to show more metrics.
10. Move the more records like this display hidden in a tab to below the wordcloud.
Mockup 9: Email Record Form

1. Button to open the email record feature.
2. Add an optional note field that sends the text to the email along with the record.
Mockup 10: Citation Builder

1. Button to open the citation builder feature.
2. Add a "Copy Citation" button.
Mockup 11: In-Document Search

1. When users click search for the in-document search tool, they are taken to a new webpage with their results.
2. Add "Back to Record" button. Ensure that the browser's back button also takes users to the full record display.
Mockup 12: Metrics Report

- **Fast All-Sky Radiation Model for Solar Applications (FARMS): A Brief Overview of Mechanisms, Performance, and Applications: Preprint**

<table>
<thead>
<tr>
<th>Year</th>
<th>Hits</th>
<th>Downloads</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>110</td>
<td>45</td>
</tr>
<tr>
<td>2015</td>
<td>130</td>
<td>21</td>
</tr>
<tr>
<td>2014</td>
<td>1023</td>
<td>58</td>
</tr>
</tbody>
</table>

1. When users select "Show More" under the metrics portion of the full record display, they are taken to a new page with more metrics details.
2. Users can view metrics by year or broken down further by month.
3. Add a "Back to Record" button to return to the full record display. Also ensure that the browser’s back button returns the user to the full record display.
4. Allow users to export the metrics report in Excel, CSV, or XML.
Links to Items

*Mockup 13: Full Text Display for Journal Articles*

1. Normalize the full text display for all journal article records.
2. For full text that is an official accepted manuscript, display a pdf icon and the label "Accepted Manuscript."
3. For links to the publisher’s site where users have to pay-to-access, display a link icon and the doi under the label "Publisher’s Version of Record."
4. For full text that is not an official accepted manuscript, display a pdf icon and the label "View Full Text" under the same label that accepted manuscripts have, "Free Publicly Accessible Full Text."
5. If the accepted manuscript is embargoed, display the date when it will become available.
6. If the publisher’s version is open access, display an open access icon next to the link so users know they don’t have to pay-to-access.
Conclusions and Future Study

SciTech Connect, a public database with unique research funded by the DOE, has changed and grown to adapt to a scientific community and federal policies that err towards open research. In this environment, it is essential that the database interface’s design and function is
analyzed and updated to be usable for its public audience. This study found a few major issues. The full text and DOI links and displays are inconsistent, the date inputs allow user error, and the basic search form does not behave like similar applications and caused difficulty for participants in the user study. This study also offers many suggestions to address those major issues and improve other aspects of the database in the redesign recommendation mockups based on feedback from participants in the user study and expert analysis of the database’s interface. There are two areas of usability not addressed in this particular study but would be useful to further improve the database: mobile use of the database on phones and tablets and accessibility of the database for users with disabilities. While touched on in this study, a closer look at users’ understanding of labels and icons throughout the database could also bring to light needed improvements.
References


Subject: Increasing Access to the Results of Federally Funded Scientific Research.

Retrieved from

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


10.5062/F4513W6K
Appendix A

Federal Public Access Plans Chart
<table>
<thead>
<tr>
<th>Agency</th>
<th>Plan</th>
<th>Date Plan Goes Into Effect</th>
<th>How Authors Submit Scholarly Publications</th>
<th>Scholarly Publications Repository</th>
<th>Plans for Scientific Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agency for Healthcare Research and Quality (AHRQ)</td>
<td>AHRQ Public Access to Federally Funded Research <a href="http://www.ahrq.gov/funding/policies/publicaccess/index.html">http://www.ahrq.gov/funding/policies/publicaccess/index.html</a></td>
<td>February 1, 2015</td>
<td>Authors submit accepted manuscript or the Publisher version if they have the rights to redistribute it to PubMed Central through the National Institutes of Health Manuscript Submission System (<a href="https://nihms.nih.gov/db/sub.cgi">NIHMS</a>) no later than 12 months after publication date.</td>
<td>PubMed Central (<a href="http://www.ncbi.nlm.nih.gov/pmc/">http://www.ncbi.nlm.nih.gov/pmc/</a>)</td>
<td>Researchers must submit a data management plan for sharing data with funding proposals or explain why data will not be shared. Generally expected to have timely release and sharing of data no later than acceptance for publication of main findings from final dataset. Normal exclusions apply (proprietary, PPI, classified, cost). Grantees expected to write updates on public access to publications and data in reports using My NCBI. The agency will be working to find a commercial archive to preserve and provide public access to submitted datasets.</td>
</tr>
</tbody>
</table>
### Centers for Disease Control and Prevention (CDC)

<table>
<thead>
<tr>
<th>Plan Title</th>
<th>Date</th>
<th>Action Description</th>
<th>Plan Link</th>
</tr>
</thead>
</table>

### Department of Defense (DOD)

<table>
<thead>
<tr>
<th>Plan Title</th>
<th>Date</th>
<th>Action Description</th>
<th>Plan Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan to Establish Public Access to the Results of Federally Funded Research</td>
<td>Q4 FY2016</td>
<td>Authors submit accepted manuscript or the Publisher version if they have the rights to redistribute it to DTIC's submission system upon acceptance by publisher.</td>
<td><a href="http://www.dtic.mil/dtic/pdf/dod_public_access_plan_feb2015.pdf">http://www.dtic.mil/dtic/pdf/dod_public_access_plan_feb2015.pdf</a></td>
</tr>
</tbody>
</table>

### Department of Education Institute for Educational Studies (DOE/IES)

<table>
<thead>
<tr>
<th>Plan Title</th>
<th>Date</th>
<th>Action Description</th>
<th>Plan Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>IES Policy Regarding Public Access to Research</td>
<td>October 1, 2012</td>
<td>Authors submit accepted manuscript to ERIC upon acceptance by Publisher.</td>
<td><a href="http://eric.ed.gov/submit/">http://eric.ed.gov/submit/</a></td>
</tr>
</tbody>
</table>

Researchers must submit a data management plan with funding proposals. Must use at a minimum the common core metadata schema in use by the federal government when documenting metadata. Allows for embargo period on data sharing till after publication of results. Normal exclusions apply (proprietary, PPI, classified, cost).

Researchers must submit a data management plan with funding proposals that includes their plans for the preservation and public access to data. Grantees are expected to deposit their data into appropriate repositories. DTIC will be creating a catalog of data locations that grantees will have to submit metadata to for the publicly available datasets. Normal exclusions apply (proprietary, PPI, classified, cost).

Researchers must submit a data management plan with funding proposals that describe the method of data sharing, types of data to be shared, and documentation that will be
| Department of Energy (DOE) | DOE Public Access Plan (http://www.energy.gov/sites/prod/files/2014/08/f18/DOE_Public_Access%20Plan_FINAL.pdf) | October 1, 2014 | Authors submit accepted manuscript or link to accepted manuscript along with metadata including the DOI to the Publisher's version to OSTI through their institution's OSTI program if one exists, or through e-link if one does not (https://www.osti.gov/elink/) not later than 12 months after publication date. *Update to this policy on PAGES FAQ webpage: "If your article has been published in an Open Access journal or as an Open Access article in a hybrid journal, you may submit the article in lieu of the accepted manuscript version of the article. The article should clearly include markings indicating that it is an Open Access article (typically with a PAGES (http://www.osti.gov/pages/) and SciTech Connect (http://www.osti.gov/scitech/) Researchers must submit a data management plan with funding proposals that follows templates available here: http://science.energy.gov/funding-opportunities/digital-data-management/. Plans should address how data will be shared and preserved, with special attention to how results could be validated. Data should be made available at the time of publication of results, with the publication indicating how the data can be accessed. Normal exclusions apply (proprietary, PPI, classified, cost). | created to promote responsible use of data. It is generally expected that data will be shared by the time the results are published. Exclusions for protecting privacy and rights of human subjects. |
| Food and Drug Administration (FDA) | Plan to Increase Access to Results of FDA-Funded Scientific Research ([http://www.fda.gov/downloads/ScienceResearch/AboutScienceResearchatFDA/UCM435418.pdf](http://www.fda.gov/downloads/ScienceResearch/AboutScienceResearchatFDA/UCM435418.pdf)) | October 1, 2015 | Authors submit accepted manuscript or the Publisher version if they have the rights to redistribute it and a set of standard article metadata to PubMed Central through the National Institutes of Health Manuscript Submission System (NIHMS [https://nihms.nih.gov/db/sub.cgi](https://nihms.nih.gov/db/sub.cgi)) no later than 12 months after publication date. | PubMed Central ([http://www.ncbi.nlm.nih.gov/pmc/](http://www.ncbi.nlm.nih.gov/pmc/)) | Agency will develop data management plan requirements to meet OMB M-13-13 and ensuring data is shared publically at time the results are published. |
| National Aeronautics and Space Administration (NASA) | NASA Plan: Increasing Access to the Results of Scientific Research ([http://science.nasa.gov/media/medialibrary/2014/12/05/NASA_Plan_for_incr](http://science.nasa.gov/media/medialibrary/2014/12/05/NASA_Plan_for_incr)) | November 21, 2014 | Authors submit accepted manuscript or the Publisher version if they have the rights to redistribute it to PubMed Central and metadata either to their institution's NASA Center | PubMed Central ([http://www.ncbi.nlm.nih.gov/pmc/](http://www.ncbi.nlm.nih.gov/pmc/)) | Researchers must submit a data management plan with funding proposals describing whether and how data will be shared and preserved or explaining why it won't be with emphasis on how |
National Institute of Standards and Technology (NIST)

Plan for Providing Public Access to the Results of Federally Funded Research

October 1, 2015

Authors submit accepted manuscript or the Publisher version if they have the rights to redistribute it to PubMed Central through the National Institutes of Health Manuscript Submission System (NIHMS https://nihms.nih.gov/db/sub.cgi) no later than 12 months after publication date.

PubMed Central (http://www.ncbi.nlm.nih.gov/pmc/)

Researchers must submit a data management plan with funding proposals describing whether and how data will be shared and preserved or explaining why it will not be with emphasis on how results could be validated. Generally, data should be made publicly available at time results are published or within a reasonable time period after publication. Normal exclusions apply (proprietary, PPI, classified, cost). Creators will eventually need to submit the metadata to their shared datasets to NIST's Enterprise Data Inventory (EDI), currently in development. This will become available in a Common Access Platform (CAP) that is also in development.
<p>| National Institutes of Health (NIH) | Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research (<a href="http://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf">http://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf</a>) | April 1, 2008 | Authors submits accepted manuscript or the Publisher version if they have the rights to redistribute it to PubMed Central through the National Institutes of Health Manuscript Submission System (NIHMS <a href="https://nihms.nih.gov/db/sub.cgi">https://nihms.nih.gov/db/sub.cgi</a>) upon acceptance by Publisher. | PubMed Central (<a href="http://www.ncbi.nlm.nih.gov/pmc/">http://www.ncbi.nlm.nih.gov/pmc/</a>) | Researchers must submit detailed data management plan with varying requirements by sub-funding agency. Generally, data should be made publicly available by the time results are published. NIH is working to expand its data requirements. Exclusions for protecting the rights and privacy of human subjects apply. NIH hosts a variety of data repositories that researchers can choose from and allows them to find other repositories and means to share their data. |
| National Oceanic and Atmospheric Administration (NOAA) | NOAA Plan for Increasing Public Access to Research Results (<a href="http://docs.lib.noaa.gov/noaa_documents/NOAA_Research_Council/NOAA_PARR_Plan_v5.04.pdf">http://docs.lib.noaa.gov/noaa_documents/NOAA_Research_Council/NOAA_PARR_Plan_v5.04.pdf</a>) | Q1 - Q2 FY2016 | Authors submit accepted manuscript along with metadata, including the DOI to the Publisher's version, to NOAA Institutional Repository submission system currently in development upon acceptance by Publisher. | NOAA Institutional Repository (a sub-set of CDC Stacks in development) | Digital data refers specifically to environmental data. Researchers must submit detailed data management plan with emphasis on metadata and documentation to make the dataset independently understood. Some funding programs have different requirements and specify the repository that data must go into. Generally, data is expected to be shared within 2 years of creation. The agency has several data repositories. |
| National Science Foundation (NSF) | Today's Data, Tomorrow's Discovers: Increasing Access to the Results of Research Funded by the National Science Foundation (<a href="http://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf">http://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf</a>) | January 1, 2016 | Authors submit accepted manuscript, link to accepted manuscript, or Publisher's version if they have the rights to redistribute it, along with metadata or the DOI to the Publisher's version, to the Department of Energy's PAGES system (<a href="https://www.osti.gov/elink/">https://www.osti.gov/elink/</a>) no later than 12 months after publication date. | PAGES system <a href="http://www.osti.gov/pages">http://www.osti.gov/pages</a> using a different interface NSF-PAR or Public Access Repository (<a href="http://par.nsf.gov">http://par.nsf.gov</a>) | They are continuing with their current data management plan requirements for funding proposals. Different requirements for each sub-funding agency and different data repositories are available for some sub-funding agencies. Normal exclusions apply (proprietary, PPI, classified, cost). |
| Office of the Assistant Secretary for Preparedness and Response (ASPR) | Office of the Assistant Secretary for Preparedness and Response Public Access to Federally Funded Research: Publications and Data (<a href="http://www.phe.gov/Preparedness/planning/science/Documents/AccessPlan.pdf">http://www.phe.gov/Preparedness/planning/science/Documents/AccessPlan.pdf</a>) | October 1, 2014 | Authors submit accepted manuscript to PubMed Central through the National Institutes of Health Manuscript Submission System (NIHMS <a href="https://nihms.nih.gov/db/sub.cgi">https://nihms.nih.gov/db/sub.cgi</a>) no later than 12 months after publication date. | PubMed Central (<a href="http://www.ncbi.nlm.nih.gov/pmc/">http://www.ncbi.nlm.nih.gov/pmc/</a>) | Researchers must submit a data management plan with funding proposals describing how they will provide for the long-term preservation of, and access to, data. It is generally expected to share data within 30 months from creation or upon publication of research based on that data. Normal exclusions apply (proprietary, PPI, classified, cost). |</p>
<table>
<thead>
<tr>
<th>Smithsonian Institution (SI)</th>
<th>Plan for Increased Public Access to Results of Federally Funded Research (<a href="http://public.media.smithsonianmag.com//file_upload_plugin/1f143b54-a9f9-a9f9-a9f9-a9f9-a9f9-a9f9">http://public.media.smithsonianmag.com//file_upload_plugin/1f143b54-a9f9-a9f9-a9f9-a9f9-a9f9-a9f9</a>)</th>
<th>October 1, 2015</th>
<th>Authors submit accepted manuscript, link to accepted manuscript, or Publisher’s version if they have the rights to redistribute it to Smithsonian Research Online or other Smithsonian approved</th>
<th>Smithsonian Research Online (<a href="http://research.si.edu/">http://research.si.edu/</a>) or Smithsonian-approved</th>
<th>Researchers must submit a data management plan with funding proposals. Researchers should include links to underlying data when submitting their publications to Smithsonian Research Online.</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States Department of Agriculture (USDA)</td>
<td>Implementation Plan to Increase Public Access to Results of USDA-funded Scientific Research (<a href="http://www.usda.gov/documents/USDA-Public-Access-Implementation-Plan.pdf">http://www.usda.gov/documents/USDA-Public-Access-Implementation-Plan.pdf</a>)</td>
<td>January 1, 2016</td>
<td>Authors submit accepted manuscript or the Publisher’s version if they have the rights to redistribute it to PubMed Central through the National Institutes of Health Manuscript Submission System (<a href="https://nihms.nih.gov/db/sub.cgi">NIHMS</a>) no later than 12 months after publication date.</td>
<td>PubAg (<a href="http://pubag.nal.usda.gov/pubag/home.xhtml">http://pubag.nal.usda.gov/pubag/home.xhtml</a>)</td>
<td>Establishes plan to create data policy that addresses requirements of OMB M-13-13.</td>
</tr>
<tr>
<td>United States Department of Veteran's Affairs (VA)</td>
<td>Policy and Implementation Plan for Public Access to Scientific Publications and Digital Data from Research Funded by the Department of Veterans Affairs</td>
<td>December 31, 2015</td>
<td>Authors submit accepted manuscript to PubMed Central through the National Institutes of Health Manuscript Submission System (<a href="https://nihms.nih.gov/db/sub.cgi">NIHMS</a>) no later than 12 months after publication date.</td>
<td>PubMed Central (<a href="http://www.ncbi.nlm.nih.gov/pmc/">http://www.ncbi.nlm.nih.gov/pmc/</a>)</td>
<td>Researchers must submit a data management plan with funding proposals which specifically describes how and to what extent data underlying all publications will be made available to the public. Expected that all data underlying all publications will be...</td>
</tr>
<tr>
<td>U.S. Geological Survey (USGS)</td>
<td>Public Access to Results of Federally Funded Research at the U.S. Geological Survey: Scholarly Publications and Digital Data (<a href="https://www2.usgs.gov/quality_integrity/open_access/downloads/USGS-PublicAccessPlan-APPROVED-v1.03.pdf">https://www2.usgs.gov/quality_integrity/open_access/downloads/USGS-PublicAccessPlan-APPROVED-v1.03.pdf</a>)</td>
<td>October 1, 2016</td>
<td>Authors submit accepted manuscript to USGS Publications Warehouse upon completion of peer-review.</td>
<td>USGS Publications Warehouse (<a href="https://pubs.er.usgs.gov">https://pubs.er.usgs.gov</a>)</td>
<td>All supporting digital research data for funding proposals should be made available free-of-charge for public access simultaneously with or prior to the associated publication, unless the agency agrees that a demonstrated circumstance restricts the data from being made publicly available; accompanied by a data management plan; and documented with metadata using the FGDC Content Standard for Digital Geospatial Metadata standard or the ISO Suite of Standards.</td>
</tr>
</tbody>
</table>
Appendix B

User Test Protocol
You are being asked to participate in a research study. Researchers are required to provide a consent form such as this one to tell you about the research, to explain that taking part is voluntary, to describe the risks and benefits of participation, and to help you to make an informed decision. You should feel free to ask the researchers any questions you may have.

**Contact Information**

Principal Investigator Name and Title: Professor Michael Twidale, PhD  
Department and Institution: Graduate School of Library and Information Science  
Address and Contact Information: 501 E Daniel, Champaign IL 61820, (217) 265-0510,  
twidale@illinois.edu

Co-Investigator Name and Title: Heidi Uphoff, Certificate of Advanced Study Candidate  
Department and Institution: Graduate School of Library and Information Science  
Address and Contact Information: 501 E Daniel, Champaign IL 61820, (217) 549-4310,  
huphoff2@illinois.edu

If you feel you have not been treated according to the descriptions in the form, or if you have any questions about your rights as a research subject, including questions, concerns, complaints, or to offer input, you may call the Office for the Protection of Research Subjects (OPRS) at 217-333-2670 or e-mail OPRS at irb@illinois.edu.

**Study Overview, Purpose, and Benefits**

This is a usability study of a database, SciTech Connect, which provides public access to the Department of Energy’s research, documents, and data. The purpose of this study is to identify usability issues in the database and make recommendations to the database owners to improve the application for users.

You may learn about a resource useful to your work. To the best of our knowledge, the things you will be doing have no more risk of harm than you would experience in everyday life.

**Procedures**

This session may take up to one hour to complete. You will be asked to perform a series of tasks in the database.

During this session, I would like you to think aloud as you work to complete the tasks. I will not be able to offer any suggestions or hints, but from time to time, I may ask you to clarify what you have said or ask you for information on what you were looking for or what you expect to have happen.
There are no right or wrong answers. We are testing the database together.

After you have completed all of the tasks, I will ask you to fill out a final survey about your experience using the database.

**Confidentiality / Privacy**

Results of this study will be published by co-investigator Heidi Uphoff in her final report for her Certificate of Advanced Study graduate degree. No personal information that could reveal your identity will be published.

However, laws and university rules might require us to tell certain people about you, such as representatives of the university committee and office that reviews and approves research studies and other federal, state, and university representatives responsible for ethical, regulatory, or financial oversight of research.

**Consent**

Your participation in this research is voluntary. If you decide to participate, you are free to withdraw at any time.

I have read (or someone has read to me) the above information. I have been given an opportunity to ask questions and my questions have been answered to my satisfaction. I agree to participate in this research. I am over 18 years of age.

______________________________  ____________________
Signature                                        Date

______________________________
Printed Name

______________________________  ____________________
Signature of Person Obtaining Consent                     Date

______________________________
Printed Name of Person Obtaining Consent
Each task is followed by (Y N) to record if the participant was successful in the task.

1. What are your first impressions of the database? What kind of content does it contain? (Y N)

2. Find something that was recently added to the database. (Y N)

3. Search for everything in the database about renewable energy. (Y N)

4. Narrow these results down to items by the author R. Meadows. (Y N)
5. Select one of the records that has a full text pdf. See if the document contains the terms “wind turbines”. (Y N)

6. Find a video about clean energy. (Y N)

7. Send the record for the video to my email: huphoff2@illinois.edu  (Y N)

8. Find a scientific dataset about solar radiation. (Y N)
9. Find all other records by the first author of this scientific dataset. (Y N)

10. Find all records created by people at Argonne National Laboratory in the last 5 years. (Y N)

11. Sort these records by date with the most recent displaying first. (Y N)

12. Go to the last page of the search results. (Y N)
13. Export the entire search results into a spreadsheet. (Y N)

14. Find the technical report assigned the number PNNL-24460. (Y N)

15. Find the citation for this report, the description you would copy and put in the references of a school paper. (Y N)

16. Find a journal article about coal power published in the last year. (Y N)
17. What journal published this article? (Y N)

18. Access the full text for this article. (Y N)

19. Go back to the record. Does it have a link to the journal publisher’s website? (Y N)

20. If you had questions about this database or wanted help using it, what would you do? (Y N)
A USABILITY STUDY OF SCITECH CONNECT

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I think that I would like to use this system frequently</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. I found the system unnecessarily complex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I thought the system was easy to use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. I think that I would need the support of a technical person to be able to use this system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. I found the various functions in this system were well integrated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. I thought there was too much inconsistency in this system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. I would imagine that most people would learn to use this system very quickly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. I found the system very cumbersome to use</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. I felt very confident using the system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. I needed to learn a lot of things before I could get going with this system</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>