FESAC Panel Report Addressing Policies and Practices Influencing the Dissemination of Research Results

FESAC Subcommittee
B. I. Cohen, chair, Lawrence Livermore National Laboratory
D. T. Anderson, U. Wisconsin-Madison
L. Berry, Oak Ridge National Laboratory
R. Betti, Princeton Plasma Physics Laboratory and U. Rochester
M. Brown, Swarthmore College
J. Finn, Los Alamos National Laboratory
C. Greenfield, General Atomics
A. Hubbard, Mass. Inst. of Tech.
S. Knowlton, Auburn University
R. Majeski, Princeton Plasma Physics Laboratory

M. Greenwald, *ex officio*, MIT

Executive Summary

The policies and practices controlling the dissemination of research results from research sponsored by the Department of Energy Fusion (DOE) Energy Sciences are described herein. Research results are shared with the public in the form of publications, conference presentations, technical reports, computer codes, and digital data. The determination that criteria for disseminating research results are met is made by the individual researcher and his/her research collaborators, and is subject to internal review to varying degrees at the universities and laboratories. Policies governing some aspects of dissemination are defined in DOE policies and orders. The institutions of the researchers define policies and practices effecting dissemination in many cases. Particular attention is given in policies and practice to intellectual property issues associated with copyrights, patents, royalties, and licensing. Formal policies related to dissemination and long-term retention of digital data are incomplete. In developing policy and guidance to the research community with respect to providing access to research results, some consideration should be given to the cost implications and long-term retention issues.

I. Introduction

The overarching question addressed in this assessment is what are the policies and practices that apply to the dissemination of and public access to results of research funded by Fusion Energy Sciences. The dissemination takes the form of publications, conference presentations, technical reports, computer codes, and digital data.

This report responds to the questions and issues listed in the February 25, 2011 letter from Dr. W. F. Brinkman to the Fusion Energy Sciences Advisory Committee (see FESAC charge in Appendix A), which were motivated by the America COMPETES Reauthorization Act of 2010. The questions and issues from the letter are reproduced here:

1. The criteria for dissemination and who makes this determination.

2. How access is provided and controlled.

Access could be provided through commercial or not-for-profit publishers or databases including archives, websites, and agency repositories.

3. Whether access is limited in any way.

For both written findings and digital data, the distribution could be limited by, for example, subscription fees, technological barriers, by request only, or limited to the members of a particular research group. Furthermore, access may be exclusive for a limited period of time.

4. Whether the access comes with any additional functionality.

For written material, this could be interoperable, cross-publisher searches or federated search and discovery tools; links to data or other supplementary material used in the research (particularly if this ensures reproducibility of the research result); or multimedia; etc.

For digital data, this could be the ability to reference the data as entered (or as part of a larger dataset), additional metadata or software interfaces for meaningful data mining by people outside the field, or interoperability with other data sets.

5. The version of the written material or data provided.

For example, for written findings, the Version of Record is usually considered to be the manuscript published and stewarded by the publisher; however, internal university or laboratory drafts may also be disseminated.

For digital research data, a distinction may be drawn between data sets that are statically preserved and those that are continually updated; whether the data are considered "raw" or "analyzed"; and whether the data that support a particular finding can be referenced, for example, by a persistent identifier.

6. Whether peer review is a condition of dissemination.

For written findings, a distinction could be drawn between external peer review, as usually happens with published articles, and an internal peer review as might happen within a Laboratory, university, or scientific collaboration for draft articles to be submitted for publication or conference proceedings.

Any comparable review process for digital data should be described in the report.

7. The institution, DOE user facility, or other body by which the policy is currently upheld.

Many Federal agencies, Laboratories, Universities, scientific collaborations, and user facilities have their own policies regarding the dissemination of research results including digital data. There may also be established practices that are not formally enforced by any institution but are broadly followed. For example, research communities may have dissemination practices that are followed, independent of agency/institutional requirements.

8. Whether, in addition to dissemination, long-term stewardship is accounted for by the existing policy or practice.

For digital data, the report could mention whether associated software for

accessing data is also available and maintained.

There are other issues related to the existing policies and practices regarding public dissemination of research results in the Fusion Energy Sciences community. Some of these are as follows:
1. How compulsory is dissemination of data? Who determines the obligation to disseminate the data?
2. What are the responsibilities and obligations of the originator of data and the providers/stewards of the archive in providing rules and structure for making the data useful/intelligible to others? What are the associated resource implications?
3. What protections are there against innocent or malicious abuse from those requesting access to and assistance with the data?
4. How do intellectual property rights, copyrights and patent issues affect sharing of research results and data?
5. With respect to digital data, what are the mechanics for defining common data formats, arranging for access (data servers inside or outside firewalls), and providing tools for and assistance in processing the data?

The response presented here to Dr. Brinkman describes the mechanics of how we share our research results in the various forms (scholarly publications, conference presentations, digital data, and codes); how public access occurs and what are the limitations; what policies control the process; who enforces the policy and how; what the obligations of the researchers with respect to dissemination of research results are; and what the formal and practical limitations affecting dissemination are. In addition, we identify some of the issues of which DOE should be mindful, e.g., responding to requests for access to research results and data can come at a cost for the respondent, the potential for innocent or malicious misuse of data, the potential for abusive requests for access, and the resource implications for providing meaningful access to digital data (which may require significant and continuing assistance by the originating researcher in processing the data).

Responses to the questions have been collected from a representative and significant sampling of research organizations funded by DOE Fusion Energy Sciences: Auburn University, UCLA, UC Irvine, Columbia University, General Atomics, Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory, Los Alamos National Laboratory, U. Maryland, Oak Ridge National Laboratory, Mass. Inst. of Technology and the Plasma Science Fusion Center, Princeton University and Princeton Plasma Physics Laboratory, University of Rochester and the Laboratory for Laser Energetics, Swarthmore College, U. Texas at Austin, and U. Wisconsin-Madison. The physicists contributing to this response represent the experimental and theoretical research communities in approximately equal numbers, and there is a depth of experience across the group in disseminating research results in all of the formats described here.

The panel report is organized as follows. Following this introduction, responses to Dr. Brinkman's questions are organized according to the four types of material being disseminated: publications (journal articles, books, book chapters, published proceedings, technical reports), conference presentations (viewgraphs and posters), digital data from

experiments and simulations, and codes. Within each category we describe the policies and the practices. A representative sampling of the formal policies and procedures that provide guidance to researchers are reproduced in the appendix, as is an example of the detailed responses from General Atomics to Dr. Brinkman's questions. DOE orders and regulations dictate much of the policy. Guidance for the researchers is to be found in documentation available from DOE and from the home institutions of the researchers in the case of universities and national laboratories. The DOE policies apply in various forms to all researchers working on grants or contracted research funded by DOE. Both the universities in their offices of sponsored projects or contracts and grants, and the national laboratories in their information management and intellectual property systems have significant human infrastructure to assure compliance with DOE rules and regulations on the dissemination of research results and to assist the researchers. In our description of the practices we include the mechanics and control of the dissemination. This report concludes with a discussion of some of the issues and challenges that are of special concern in the context of sharing research results, data, and codes.

II. Publications and Reports

With respect to publications and reports, most of the practices and policies for all researchers funded by DOE Fusion Energy Sciences (FES) are fundamentally the same independent of whether the researcher is part of a large research group at a university, national laboratory, or corporate laboratory, or is an individual principal investigator. The criteria for dissemination are as follows:

1. Sufficient scientific merit to warrant dissemination.
2. Satisfying the restrictions on data dissemination and publication that might be imposed by collaborative agreements with other domestic or foreign institutions.
3. Does not contain information that is potentially sensitive, export controlled or classified.
4. Does not contain potentially patentable information that is inadequately protected.
5. The material has been appropriately reviewed (see the following discussion).

Decisions on dissemination are made by the research staff originating the research, although other staff at the institution may become involved in some situations. The responsibility for upholding the policies of DOE with respect to dissemination of research results in whatever form is the responsibility of the originating researcher and his/her home institution. The originating researcher controls the version of record with respect to public dissemination.

Internal review of materials reporting research results at the originating institution and external review before publication constitute the primary means of controlling public dissemination of research results. The review process contributes a very significant element in the quality control of the research results to be disseminated. In practice and in policy, some form of peer review of the documents reporting research results is a

condition for disseminating research across our research community. There are many incentives for researchers in our community to publish research results in peer-reviewed journals and proceedings. These publications are the most important metric in determining whether a researcher is successful and whether the goals of the research are being met. A researcher's publication list is often used as a key input in the process for determining professional promotions and merit-based salary increases. Hence, dissemination of research results, particularly in prestigious peer-reviewed journals, is compulsory in practice.

It is generally true that the larger the research facility, the more internal peer review and screening of scholarly publications and reports occur before the materials are distributed outside the institution. For scholarly publications submitted to peer-reviewed journals and conference proceedings, there is a formal peer-review process run by the journal or editor(s) of the proceedings. Authors release their copyrights to the journal as a condition for publication, but for federally funded research, the U. S. government retains royalty-free, nonexclusive rights to the material. Peer review is the most important quality control on the dissemination of research results.

The internal screening and review of the research publication or report before it leaves the originating institution are fairly rigorous, but uneven in the fusion community. At the DOE laboratories (e.g., Los Alamos, Lawrence Livermore, Lawrence Berkeley, Oak Ridge, the Laboratory for Laser Energetics at U. Rochester, and Princeton Plasma Physics Laboratory), at General Atomics and at MIT's Alcator C-Mod National Facility, there is a formal process for reviewing and releasing all material leaving the institution. A publication or report will typically first be reviewed by a co-worker, collaborator or other peer with respect to its research content; but there is no standard for this practice. It is then reviewed with respect to export control issues and general content by a reviewer within the division or department, then for classification by an Authorized Derivative Classifier at some of the DOE laboratories (relevant and obligatory for inertial fusion but not applicable for magnetic fusion research which is a Designated Unclassified Subject Area), reviewed additionally by someone representing the office having oversight on classification and export control at some of the DOE laboratories, and reviewed specifically by an office responsible for intellectual property if the nature of the subject matter is scientific or technical. The rigor of each of these review steps is somewhat uneven, e.g., review for classification and export control issues at the national laboratories is the most rigorous. The policies for this are described in DOE orders to the laboratories, and the practices are described in documents maintained by the technical information departments at the laboratories (typically available on-line within the laboratory's internal web site). Some examples of the policies and practices documents are provide in the appendices. The policies defined by DOE at the corporate laboratories and universities are fundamentally the same as at the DOE laboratories. Although there tends to be less human infrastructure in place at the universities to screen the research publications than at the DOE laboratories, the basic elements in the review process are similar. However, a university researcher may have to execute most or all of the internal review and screening of his/her research with little or no assistance before dissemination.

We note that a manuscript or report may or may not receive a rigorous review for its validity, originality and topical importance at the institution of the originating author or authors. Even when there is an internal review, this review is often not equivalent with respect to the rigor of a peer review a manuscript receives at a peer-reviewed journal. A proper peer review of a publication will challenge the correctness, originality and importance of the paper's content, the clarity of the exposition, and whether the scholarship is up to standards. In the limiting case of an individual researcher working at a university in a small research group, a paper may not receive any rigorous review by another experienced researcher until the paper is submitted to a journal and reviewed by a referee.

As mentioned in the preceding, research at universities, DOE laboratories and corporate laboratories is reviewed with respect to its intellectual property content. For most inventions, the local institution may elect to retain the patent rights, but the U. S. government retains royalty-free, nonexclusive rights for government use. Computer software can be copyrighted and licensed for commercial value and is treated the same as the intellectual property leading to patents. The policies in this regard are set forth in DOE policies and orders (Table I has links to the internet to retrieve the relevant DOE documents). Both the universities and laboratories have offices and staff that assist with intellectual property issues. After there is a filing of a record of invention or a preliminary patent filing, and similarly for computer software, the research is then disseminated publicly. DOE patent licensing regulations are described in the following documentation available on-line: http://law.justia.com/cfr/title10/10-4.0.2.5.18.html

The following summarizes the steps associated with reviewing research results before public dissemination occurs:

1. The originating author (or authors) determine the readiness for reporting the research and the format in which the research results will be disseminated, e.g., journal publication, conference presentation, software, digital data, or patent disclosure.
2. The material is reviewed locally to assess (i) the validity and appropriateness of the scientific content, (ii) whether the material contains any information that is sensitive in any way, is classified, or is export controlled (classification may not pertain at a university), (iii) whether the material is subject to patent or licensing considerations (an invention, a process or software), (iv) whether authorship and acknowledgments are appropriate. Issues listed under (ii) and (iii) typically require additional review and processing.
3. The paper is submitted to a peer-reviewed journal or published proceedings or to an editor-reviewed book or published proceedings or submitted to an electronic preprint server which may be subject to some screening.

For research results published in journals, there is library and internet access to the publications. Typically only access is provided to the publication without any additional functionality except for coupling to search engines and certain scientific databases. Downloading a publication usually requires paying a fee unless access to the journal has been paid for already. Preprints (before publication) and reprints (after formal peer

review and publication in a journal) are often available free of charge from a web site maintained through the auspices of the home institution of the researcher (e.g., institutional library, departmental, research group and/or personal web site), and this is a common practice in the fusion research community. Sometimes pre-prints are also posted on the internet through http://www.arXiv.org.

There are no standards on how long publications and reports will continue to be available on institutional or personal web sites, but such sites are typically maintained for many years. The availability of the research in the databases and web sites of the journals is assumed to be longer term (unless the journal disappears). Publications and reports in the fusion community are also routinely submitted to the DOE Office of Scientific and Technical Information archive (OSTI, http://www.osti.gov/) and the Department of Commerce National Technical Information Service (NTIS) archives where materials are available for download subject to a nominal fee. The publications and reports produced at the DOE laboratories are routinely conveyed to OSTI and NTIS, where the material is assumed to remain available as long as these archives exist. The databases and archives in which publications and reports are stored and which are accessible to the public can be searched and cross-referenced with various search engines and specialized tools (http://crossref.org/). No other functionality is typically provided.

In general, the fusion research community receives and/or expects direction from DOE orders and policy with respect to record management and retention as it applies to our publications, reports, conference presentations, software, and digital data. In this regard, DOE is subject to the regulations defined for the National Archives and Records Administration (NARA), which is responsible for overseeing all Federal agencies record management and retention. The regulations can be found in NARA regulations Subchapter B- Records Management (http://www.archives.gov/about/regulations/subchapter/b.html). Given that the researchers are contractors to DOE and the NARA regulations are aimed at the Federal agencies and not directly at the contractors, the research community relies on DOE for the implementation strategy pursuant to the NARA regulations, which includes the definition of what records and files must be retained and the length of the retention time. From the researcher's perspective, the regulatory environment on records and retention lacks clarity. Policy and orders from DOE to our research community with respect to the totality of record management and retention are currently incomplete.

TABLE I. FINANCIAL ASSISTANCE INTELLECTUAL PROPERTY PROVISIONS

http://www.gc.energy.gov/financial_assistance_awards.htm

Type of Award	**Type of Project**	**Special Data Statute** e.g., EPACT)	**Type of Recipient**	**Set Number** (PDF)
Cooperative Agreement	Research, Development, or Demonstration (RD&D)	No	Domestic Small Business	CSB-1003
Cooperative Agreement	RD&D	Yes	Domestic Small Business	CDSB-1003
Cooperative Agreement	RD&D	No	Large Business, State or Local Government, and Foreign Entity	CLB-1003
Cooperative Agreement	RD&D	Yes	Large Business, State or Local Government, and Foreign Entity	CDLB-1003
Grant	RD&D	No	Domestic Small Business	GSB-1003
Grant	RD&D	Yes	Domestic Small Business	GDSB-1003
Grant	RD&D	No	Large Business, State or Local Government, and Foreign Entity	GLB-1003
Grant	RD&D	Yes	Large Business, State or Local Government, and Foreign Entity	GDLB-1003
Grant and Cooperative Agreement	RD&D	No	Nonprofit Organization	GNP-1003
Grant and Cooperative Agreement	Non RD&D	N/A	All types of recipients	NRD-1003

III. Conference Presentations

Conference presentations in the form of oral presentations and poster presentations involve viewgraphs that receive the same rigorous institutional review as described in the preceding for the publications and reports. However, the presentations are generally not peer-reviewed by the conference organization. For the major national and international conferences, one-page abstracts summarizing the content of the presentations are often posted on the internet or published in a book of abstracts. The abstracts are reviewed at the originating institution following the same procedure as for presentations and

publications. Some of the research groups and individual researchers post the presentations on an internet web site so that anyone may download the presentations. The originating researchers control the versions of the documents that are disseminated. There is no particular rule or policy dictating whether oral and poster presentations will be posted on the internet or how long availability will be maintained. At the national laboratories the library systems generally do not archive nor make available viewgraphs from oral or poster presentations, and these are not forwarded to OSTI or NTIS. In general, if someone wants a copy of an oral or poster presentation, one must make the request to the author, subject to the condition that the presentation has been reviewed and released for unlimited external distribution. Generally, researchers in the fusion community accommodate requests for copies of presentations. No additional functionality is provided to accompany dissemination of oral and poster presentations.

Most researchers in our research community are unaware of any DOE policy or orders that obligate DOE-funded researchers to respond to requests for copies of conference presentations. However, federal law such as the Freedom of Information Act may be applicable and obligates researchers to share any and all of their research (subject to classification, export control, or intellectual property issues). The fusion research community does not have much, if any, experience with requests for our research except from within our own research community, from research communities with overlapping interests such as such as computational fluid dynamics, space plasmas, pulsed power, etc., and from federal agencies, e.g., DOE and NSF.

IV. Digital Data from Experiments and Simulations

The practice of sharing digital data is an important and growing component within the plasma physics and fusion research community. Digital data is generated directly in experiments from experimental measurements, as a consequence of data processing associated with experimental data, and from computer simulations. Research collaborations involving more researchers from multiple institutions sharing access to digital data have defined procedures and rules for the process. In some cases, this can be quite formal and requires signing a memorandum of understanding. Examples of such are provided in Appendix C. By having a formal process with a well-defined memorandum of understanding, proper credit and priority for the scientific findings can be protected and controlled, and quality in processing the data and deriving inferences can be controlled by the research team. Digital data is typically shared by establishing password-protected computer accounts for all of the research collaborators on a common data server at one of the collaborating institutions or at a facility such as the National Energy Research Supercomputer Center. The research group decides on a specific data format and organizational structure for the data archive. Examples of common data formats used in our community are MDSplus, NETCDF, HDF5, and ASCI. The research collaborations and sharing of digital data are typically a result of self-organization by the researchers. Although DOE provides guidance on computer security practices and the aforementioned policies and orders controlling the release of any products of research, DOE program management does not generally direct our community to share our digital data.

Digital data is typically not peer reviewed in the sense that publications are peer reviewed. At a DOE laboratory, when sharing digital data with researchers from other institutions, the data originating from the DOE laboratory is subject to the same review and screening as described in the preceding for publications and presentations, and uses the same review criteria. However, the reviewers review a representative sample of the data and a text abstract describing the data furnished by the originating researcher. Digital data is far from the finished product that a publication or report represents. As such, there is considerable process involved in establishing the credibility of the digital data. The process may involve significant data analysis and/or validation by undertaking additional experiments and simulations. The process and tools are integral components of the research collaboration. Without including some of the specialized data analysis tools, access to the unprocessed digital data may be relatively useless. In practice, shared data often involves sharing of software with which to process and analyze the data, i.e., access to the data comes with additional functionality in our community. Experimental and simulation databases in our community grow in time as more experiments and simulations are performed. Once the data is logged into an archive, the raw data is preserved statically. Data that results from processing and analysis is sometimes preserved dynamically. When corrections to processed or analyzed data are made, there are typically new versions of the processed data; and records are created and kept to track changes.

There is no standard on how long data is maintained on the data server, except as agreed upon by the originating research team in most cases; and it is subject to resource considerations and certain practical considerations. The need for a DOE-wide policy on data management and retention was identified in a meeting of representatives of the major DOE data centers sponsored by DOE's Office of Scientific and Technical Information in 2004 (see meeting report in Appendix G). It is difficult to generalize in describing the long-term stewardship of digital data and the specialized software used to process the data. Our research community operates under the presumption that there is no guidance from DOE on how long to preserve raw data from experiments and simulations. However, there may exist DOE policy on data retention of which our community is just unaware. Some clarification from DOE is needed on data retention. Researchers in our community typically try to preserve the data for as long as possible, and the research teams define their own policies and practices on data retention. In so doing, researchers in our community are at the mercy of such factors as the lifetime of the storage media and/or the associated technology used to retrieve the data, the longevity of the humans who can locate the data and know how to make sense of the data, and the availability of resources to preserve and retrieve legacy data. The large computer centers such as the National Energy Research Supercomputer Center (NERSC) and the computer centers at the DOE national laboratories have policies and the resources to maintain data nearly indefinitely subject to the users deciding when to make deletions. When experimental data is not conveyed to long-term storage at the computer centers and when the experiments are no longer in operation, experimental digital data that is ten years old or less is generally easy to retrieve from the originating researcher(s) with limited effort,

while data that is ten to twenty years old is more problematic to retrieve; and data that is older than twenty years old is very difficult to recover.

An example is provided by the TFTR D-T (deuterium-tritium) experiments in the 1990s at the Princeton Plasma Physics Laboratory, which represent a considerable investment by the DOE and the U.S. fusion community. Processed (waveform) data from these experiments is still available, but only to researchers equipped to access the data format provided by an outdated VAX cluster (VMS operating system). An unfunded effort was also made to store raw data, available in principle on magnetic tapes. Some of these tapes have been subsequently damaged; data from the damaged tapes probably cannot be retrieved.

Sometimes there are requests for digital data from plasma or fusion researchers outside a specific research effort or collaboration. Taking into account the quality of the data, the merits and purposes of the request, how much assistance the requestor will need in order to extract something useful from the data, and other considerations, the originating researcher or research collaboration makes a decision as to whether to share the data and sometimes the data analysis tools as well. Some researchers in our community are very reluctant to share their data with other researchers and are very selective in this regard. There is some guidance from DOE that addresses sharing simulation data (see Appendix H), but this guidance does not address long-term retention issues; and many researchers in our community may be unaware of the existence of these guidelines. Furthermore, the FES research community has little or no experience in sharing digital data except with other plasma and fusion researchers, or with researchers in closely allied disciplines, and no experience with Freedom of Information Act requests.

V. Software

The plasma and fusion research community develops specialized software to undertake numerous calculations, computer simulations, and data processing associated with experiments and simulations. Researchers in our community share much of their software freely with one another. However, the degree to which software is shared varies widely. Classification, export control, and intellectual property considerations are applicable. Some codes such as the NIMROD nonlinear magnetohydrodynamics code are shared widely. Others, usually written by a single author or small team, may be shared only with collaborators. There seems to be no oversight of this. At the DOE national laboratories, the software is reviewed and screened with respect to the same classification, export control, appropriateness, and intellectual property issues as described in the preceding for disseminating publications before the software is shared. The amount of review that software originating from the universities undergoes varies and tends to be less rigorous than at the national laboratories. However, the consideration of intellectual property issues associated with computer software is particularly important, because the software may have potential commercial value. If this is the case, DOE may allow the originating institution to hold the copyright and license the software; but the U. S. government retains free access for government use. The DOE national

laboratories, the corporate laboratories funded by DOE, and the universities have offices specializing in intellectual property issues to assist the researchers if there is potential for commercially licensing original software.

At the DOE laboratories, there is a formal process for releasing software. The author of the software provides a CD with the software and documentation describing the software, links or references to a user manual and published examples of the use of the code (if they exist) and documents describing the dependencies on other software and software libraries. Most of the software originating in our research community is shared freely and is not commercially licensed, but there are significant examples of software that have been commercially licensed.

VI. Issues and Challenges

Access to the results of research does not come without some effort by the originators of the research. Responding to requests for access to research results and data can come at a significant cost for the respondents. If the effort involved in responding to requests for access increases above what might be described as incidental, the time involved subtracts from that which otherwise would be spent on conducting research or other duties. Researchers in our community have no guidance on how much time and effort to expend on sharing the results of our research with the public and thus must use their own judgment and discretion. There is also the possibility for innocent or malicious misuse of data, the potential for abusive requests for access, and the resource implications for providing meaningful access to digital data, which may require significant and continuing assistance by the originating researcher in the processing of the data by the requestor. In the absence of explicit policies, orders, or pre-existing guidance from DOE, the researcher responding to a request will seek guidance from management at his or her institution and the DOE program monitor if there is a question on whether to respond to a request. If the request is exploiting the Freedom of Information Act, but is deemed onerous, the researcher may find himself or herself in a challenging situation. Researchers in the climate modeling research community have suffered from excessive and perhaps abusive requests for access to their data and email. It is our understanding that under existing DOE policy, DOE does not protect its researchers from possible abuse under the auspices of FOI requests. We do not have solutions for the various awkward situations that may arise nor a policy to suggest. However, we wish to draw attention to some of the challenging issues that might be described as unintended consequences of trying to promote open access to all of the results of federally funded research. We encourage DOE to consider these issues and challenges with the goal of creating policy that nurtures the research process and facilitates appropriate sharing of the research results in a practical, constructive, and responsible manner. Providing informed feedback regarding these issues to the legislative process may be beneficial to the research community.

With respect to long-term stewardship of digital data not residing at the large computer centers, for example, data stored locally at the researcher's site, DOE should consider whether a policy should be instituted, a process defined, and adequate resources allocated

to preserve the data (see Appendix G). In general, the research community and the program managers in Fusion Energy Sciences would benefit from a clear, simplified, and accessible articulation of the definitions, rules and regulations governing record management and retention as it applies to our research. In the absence of a clear and easily accessed policy, a well-defined process, and appropriate resources, there is no guarantee of long-term stewardship of digital data.

VII. Acknowledgements

We gratefully acknowledge contributions to this document from scientists and staff at universities and laboratories in the research community sponsored by Fusion Energy Sciences. Special thanks go to James Van Dam, Chris Clayton, John Barnard, David Meyerhofer, Mike Mauel, Mary Nijhuis, and Bert Weis. This work was performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 and under DOE contracts and grants at the other contributing institutions.

Appendix A: FESAC charge from the Director of the Office of Science

Department of Energy
Office of Science Washington, DC 20585

Office of the Director

February 25, 2011

Dr. Martin Greenwald, Chair
Fusion Energy Sciences Advisory Committee
Plasma Science and Fusion Center Massachusetts Institute of Technology
NW17-107
175 Albany Street
Cambridge, MA 02138

Dear Dr. Greenwald:

The recently passed America COMPETES Reauthorization Act of 2010 highlights the importance of public access to research results, particularly in the forms of scholarly publications and digital data. A copy of the relevant section, Sec. 103, of the COMPETES Act is appended to this charge letter for your information.

As a first step in assessing the policies for researchers funded by the Office of Science, I am requesting your assistance. Please submit to me, no later than July 1, 2011 a report describing current policies and practices for disseminating research results in the fields relevant to the Fusion Energy Sciences program. For the purposes of this report, "dissemination" refers to the circulation of research results outside of the originating institutions or scientific collaborations; "research results" refers to both written research findings (scholarly papers, presentations, reports, etc.) and digital data; and "practices" refers to accepted practices within a scientific discipline. Policies from DOE and other federal and non-federal agencies, including foreign institutions and international scientific collaborations, should be considered within the scope of this report provided that these policies have notable impact on the dissemination of research results in your fields. Examples of relevant government policies include provisions in grants and contracts as well as overarching guidance as set forth in federal regulations and DOE orders.

Although your report should be sensitive to the differences between written findings and digital data (and, indeed, differences among each of these), you may find many of the same considerations useful in describing the existing policies, practices, and procedures:

1 See, for example, 10 CFR 605.20 (http://law.justia.com/us/cfr/titlelO/IO-4.0.1.3.13.html#10:4.0.1 .3.13 .0.59.20) and DOE 0 241.1B (https:llwww.directives.doe.gov/directives/current-d irectives/241.1-BOrder-b/vjew).

The criteria for dissemination and who makes this determination.
How access is provided and controlled.
Access could be provided through commercial or not-for-profit publishers or databases including archives, websites, and agency repositories.
Whether access is limited in any way.
For both written findings and digital data, the distribution could be limited by, for example, subscription fees, technological barriers, by request only, or limited to the members of a particular research group. Furthermore, access may be exclusive for a limited period of time.
Whether the access comes with any additional functionality.
For written material, this could be interoperable, cross-publisher searches or federated search and discovery tools; links to data or other supplementary material used in the research (particularly if this ensures reproducibility of the research result); or multimedia; etc.
For digital data, this could be the ability to reference the data as entered (or as part of a larger dataset), additional metadata or software interfaces for meaningful data mining by people outside the field, or interoperability with other data sets. *The version o f the written material or data provided.*
For example, for written findings, the Version of Record is usually considered to be the manuscript published and stewarded by the publisher; however, internal university or laboratory drafts may also be disseminated. For digital research data, a distinction may be drawn between data sets that are statically preserved and those that are continually updated; whether the data are considered "raw" or "analyzed"; and whether the data that support a particular finding can be referenced, for example, by a persistent identifier.
Whether peer review is a condition o f dissemination.
For written findings, a distinction could be drawn between external peer review, as usually happens with published articles, and an internal peer review as might happen within a Laboratory, university, or scientific collaboration for draft articles to be submitted for publication or conference proceedings.
Any comparable review process for digital data should be described in the report.
The institution, DOE user facility, or other body by which the policy is currently upheld. Many Federal agencies, Laboratories, Universities, scientific collaborations, and user facilities have their own policies regarding the dissemination of research results including digital data. There may also be established practices that are not formally enforced by any institution but are broadly followed. For example, research communities may have dissemination practices that are followed, independent of agency/institutional requirements.
Whether, in addition to dissemination, long-term stewardship is accounted for by the existing policy or practice. For digital data, the report could mention whether associated software for accessing data is also available and maintained.

In the case of digital data, these descriptions will likely depend on the type, size, and structure of the data sets under consideration. It would be useful, therefore, to include in your discussions, a brief survey of the kinds of data that are generated, the size of the data sets, and how they are stored.

As part of this report, I welcome the Committee's perspective on which dissemination models, if

any, successfully maximize the potential benefit of research results in a way that is sustainable within the research community. I also invite you to include any observations regarding opportunities where public access policies or practices could enhance the discovery potential of Office of Science research results.

Sincerely,

W. F. Brinkman Director, Office of Science

represented on the Committee, to identify and reduce regulatory, logistical, and fiscal barriers within the Federal government and State governments that inhibit United States manufacturing;

(4) facilitate the transfer of intellectual property and technology based on federally supported university research into commercialization and manufacturing;

(5) identify technological, market, or business challenges that may best be addressed by public-private partnerships, and are likely to attract both participation and primary funding from industry;

(6) encourage the formation of public-private partnerships to respond to those challenges for transition to United States manufacturing; and

(7) develop, and update every 5 years, a strategic plan to guide Federal programs and activities in support of advanced manufacturing research and development, which shall—

(A) specify and prioritize near-term and long-term research and development objectives, the anticipated time frame for achieving the objectives, and the metrics for use in assessing progress toward the objectives;

(B) specify the role of each Federal agency in carrying out or sponsoring research and development to meet the objectives of the strategic plan;

(C) describe how the Federal agencies and Federally Funded Research and Development Centers supporting advanced manufacturing research and development will foster the transfer of research and development results into new manufacturing technologies and United States based manufacturing of new products and processes for the benefit of society to ensure national, energy, and economic security;

(D) describe how Federal agencies and Federally Funded Research and Development Centers supporting advanced manufacturing research and development will strengthen all levels of manufacturing education and training programs to ensure an adequate, well-trained workforce;

(E) describe how the Federal agencies and Federally Funded Research and Development Centers supporting advanced manufacturing research and development will assist small- and medium-sized manufacturers in developing and implementing new products and processes; and

(F) take into consideration the recommendations of a wide range of stakeholders, including representatives from diverse manufacturing companies, academia, and other relevant organizations and institutions.

(c) REPORT.—Not later than 1 year after the date of enactment of this Act, the Director shall transmit the strategic plan developed under subsection (b)(7) to the Senate Committee on Commerce, Science, and Transportation, and the House of Representatives Committee on Science and Technology, and shall transmit subse- quent updates to those committees as appropriate.

SEC. 103. INTERAGENCY PUBLIC ACCESS COMMITTEE.

(a) ESTABLISHMENT.—The Director shall establish a working group under the National Science and Technology Council with the responsibility to coordinate Federal science agency research and policies related to the dissemination and long-term stewardship of the results of unclassified research, including digital data and peer-reviewed scholarly publications, supported wholly, or in part, by funding from the Federal science agencies.

(b) RESPONSIBILITIES.—The working group shall—

(1) identify the specific objectives and public interests that need to be addressed by any policies coordinated under (a);

(2) take into account inherent variability among Federal science agencies and scientific

disciplines in the nature of
research, types of data, and dissemination models;

(3) coordinate the development or designation of standards for research data, the structure of full text and metadata, navigation tools, and other applications to maximize interoperability across Federal science agencies, across science and engineering disciplines, and between research data and scholarly publications, taking into account existing consensus stand-
ards, including international standards;

(4) coordinate Federal science agency programs and activities that support research and education on tools and systems required to ensure preservation and stewardship of all forms of digital research data, including scholarly publications;

(5) work with international science and technology counter- parts to maximize interoperability between United States based unclassified research databases and international databases and repositories;

(6) solicit input and recommendations from, and collaborate with, non-Federal stakeholders, including the public, universities, nonprofit and for-profit publishers, libraries, federally funded and non federally funded research scientists, and other organizations and institutions with a stake in long term preservation and access to the results of federally funded research;

(7) establish priorities for coordinating the development of any Federal science agency policies related to public access to the results of federally funded research to maximize the benefits of such policies with respect to their potential economic or other impact on the science and engineering enterprise and the stakeholders thereof;

(8) take into consideration the distinction between scholarly publications and digital data;

(9) take into consideration the role that scientific publishers play in the peer review process in ensuring the integrity of the record of scientific research, including the investments and added value that they make; and

(10) examine Federal agency practices and procedures for providing research reports to the agencies charged with locating and preserving unclassified research. (c) PATENT OR COPYRIGHT LAW.—Nothing in this section shall
be construed to undermine any right under the provisions of title 17 or 35, United States Code.
(d) APPLICATION WITH EXISTING LAW.—Nothing defined in sec- tion (b) shall be construed to affect existing law with respect to Federal science agencies' policies related to public access.
(e) REPORT TO CONGRESS.—Not later than 1 year after the date of enactment of this Act, the Director shall transmit a report to Congress describing—

(1) the specific objectives and public interest identified under (b)(1);

(2) any priorities established under subsection (b)(7);

(3) the impact the policies described under (a) have had on the science and engineering enterprise and the stakeholders, including the financial impact on research budgets;

(4) the status of any Federal science agency policies related to public access to the results of federally funded research; and

(5) how any policies developed or being developed by Fed- eral science agencies, as described in subsection (a), incorporate input from the non-Federal stakeholders described in subsection (b)(6). (f) FEDERAL SCIENCE AGENCY DEFINED.—For the purposes of this section, the term ''Federal science agency'' means any Federal agency with an annual extramural research expenditure of over $100,000,000.

42 USC 6624. SEC. 104. FEDERAL SCIENTIFIC COLLECTIONS.
(a) MANAGEMENT OF SCIENTIFIC COLLECTIONS.—The Office of Science and Technology Policy shall develop policies for the management and use of Federal scientific

collections to improve the quality, organization, access, including online access, and long-term preservation of such collections for the benefit of the scientific
enterprise. In developing those policies the Office of Science and Technology Policy shall consult, as appropriate, with—

(1) Federal agencies with such collections; and

(2) representatives of other organizations, institutions, and other entities not a part of the Federal Government that have a stake in the preservation, maintenance, and accessibility of such collections, including State and local government agencies, institutions of higher education, museums, and other enti- ties engaged in the acquisition, holding, management, or use of scientific collections.

(b) CLEARINGHOUSE.—The Office of Science and Technology Policy, in consultation with relevant Federal agencies, shall ensure the development of an online clearinghouse for information on the contents of and access to Federal scientific collections.

(c) DISPOSAL OF COLLECTIONS.—The policies developed under subsection (a) shall—

(1) require that, before disposing of a scientific collection, a Federal agency shall—

(A) conduct a review of the research value of the collection; and

(B) consult with researchers who have used the collection, and other potentially interested parties, concerning—

(i) the collection's value for research purposes; and (ii) possible additional educational uses for the
collection; and

(2) include procedures for Federal agencies to transfer scientific collections they no longer need to researchers at institutions or other entities qualified to manage the collections. (d) COST PROJECTIONS.—The Office of Science and Technology
Policy, in consultation with relevant Federal agencies, shall develop a common set of methodologies to be used by Federal agencies for the assessment and projection of costs associated with the management and preservation of their scientific collections.

Appendix B: Department of Energy Policies and Orders – Examples Pertaining to Science and Technology Information

https://www.directives.doe.gov/directives/current-directives/241.1-BOrder-b/view

Appendix C.1: Alcator collaborative research agreement for data access

Version 2.2 September 3, 2003

Alcator Collaboration Policy Agreement

Preamble

The Alcator group welcomes collaborations with outside groups; they significantly enhance our program and increase our impact in the national and international arena. These collaborations benefit both parties and the fusion program as a whole.

It is intended that each collaborating group have a clear programmatic role appropriate to the scale of its effort and appropriate representation in Alcator executive committees. Programs will be mutually agreed between each collaborating party and the Alcator leadership. An outline of the program will normally be attached to this policy agreement except where the collaboration is of a brief or occasional nature.

We offer our collaborators full access to C-Mod data as it is collected and analyzed and, as far as feasible, access to the Alcator facility. With these privileges come responsibilities: to observe appropriate safety and security, to ensure that the data used are correct and are correctly interpreted, and to make certain that appropriate credit for providing measurements and analysis is given. The following agreement is intended to promote mutual understanding of these responsibilities.

Agreement

1. Visiting collaborators will observe all rules, regulations and requirements of MIT while on site, including but not limited to those associated with safety, security, health, hours of work and operations.
2. Collaborators will be allowed direct access to unpublished C-Mod data once this agreement has been signed by both parties. Computer accounts and other MIT privileges such as building access and office space will be reviewed periodically and will be renewed provided the agreement is being observed.
3. The results of all research from the C-Mod program, including that performed by collaborators, will be available to all members of the national and international C-Mod team.
4. Each organization in the collaboration will designate a principal contact who will have the responsibility for keeping his/her own staff informed on the relevant issues. The choice of this person will be agreed by both parties. The principal contact will have responsibility to ensure that the terms of this agreement are known and observed by the members of the collaborating group.
5. The visiting group's principal contact will be responsible for orientation and administrative support of his organization's personnel.
6. The principal contact will have authority to represent his organization in dayto-

day decisions respecting the collaboration.

7. We encourage presentation of the collaborative work in progress at C-Mod group meetings. Collaborators should plan to present a summary of their work at Alcator to the Alcator Physics meeting, at least once every six months.
8. No member of either organization will disseminate the work of the other to third parties outside of the C-Mod team without explicit approval of the principal contact of the other.
9. Presentations at conferences and workshops of papers that make substantial use of unpublished C-Mod results are coordinated by the Alcator program committee, and require the explicit prior approval of the committee chairman or his designee. For major conferences, a rehearsal presentation to the Alcator group will normally be expected. In other cases, an abstract should be submitted to the committee.
10. Manuscripts or other materials intended for publication, and proposals, that contain the work of the collaboration will be circulated among all contributing organizations, and appropriate authorship and/or acknowledgment based on data and analysis used will be mutually agreed prior to submission for publication.

Signed for Alcator:

Signed by Collaborator:

Alcator Principal Contact:

Collaborator’s Principal Contact:

Date:

Appendix C.2: DIII-D National Fusion Facility Data Usage and Publication Policy Agreement

[The following is made available via the DIII-D website and is "signed" electronically prior to somebody new getting a computer account.]

The DIII-D National Program is a multi-institutional collaboration. In addition, the DIII-D Program extends collaborations and outreach to national and international facilities and organizations to carry out the scientific research called for in the DIII-D Program objectives.

Collaborators are offered full access to DIII-D data as it is collected and analyzed. However, with these privileges come the responsibility for collaborators to ensure that the data used are correct and are correctly interpreted and to ensure that appropriate credit for providing measurements and analysis are given.

The following agreement is intended to help avoid misunderstandings over these responsibilities and to avoid potential loss of data access:

1. No collaborator will be given direct access to unpublished DIII-D data until this agreement has been signed.
2. Presentations at conferences and workshops of papers that make substantial use of unpublished DIII-D results will be coordinated and approved by the DIII-D Director or his designee. For major conferences, a rehearsal presentation to DIII-D peers is normally expected.
3. The DIII-D program has a technical review process for all papers. Papers by General Atomics (GA) researchers must be issued as a GA-A report. Collaborators' papers relating to programmatic results must also publish their papers as GA-A reports. Such papers will be published with DIII-D National Fusion Facility in the journal masthead. Researchers who develop specialized diagnostics or subsystems or carry out specialized analysis with their individual diagnostics or modeling codes, normally submit such papers under their institution's DOE guidelines with their institution in the journal masthead. These papers are also to proceed through DIII-D Program peer review, called "courtesy review" before submission to journals. The DIII-D Director is responsible to assure that publications are not unreasonably withheld or delayed.
4. GA's role in operating the DIII-D National Program for DOE includes the contractual responsibility for technical quality and to disseminate knowledge to the public and the national collaborators. This responsibility is met by posting GA-A reports, after they have cleared DOE review, on the DIII-D publications web site. For papers copyrighted at a collaborator's institution, normally the paper is posted on the collaborator's web site and linked to the GA DIII-D server in a timely manner, or, if the collaborator agrees, the paper will be posted on the DIII-D publications web site.
5. If a computer account is required on a General Atomics Fusion Group system, then the collaborator must contact the User Service Center and sign the

"Computer Usage Policies and Procedures and User Responsibilities" document before an account can be issued.

In addition, some general data analysis routines (e.g., GAPROFILES) make use of Atomic Data and Atomic Structure (ADAS) data. ADAS users must abide by rules established in the GA-ADAS Project Agreement executed on February 1, 1999. The primary rules for data users are:

1. Publications using results from ADAS should identify the ADAS database by name and include a reference stating that: "The originating developer of ADAS is the JET Joint Undertaking."
2. ADAS programs, subroutines or data sets may not be transferred to anyone outside the DIII-D Program without the express written permission of the owner of that program, subroutine or data set.

Appendix D: Practices and Policy at Los Alamos National Laboratory Controlling Dissemination of Research Results

From the laboratory web site on Publications Release and Accountability at Los Alamos:

> All scientific and technical information generated at the Laboratory and intended for public release must be reviewed and processed by Classification (SAFE-1) before publication or submission for publication. SAFE-1 is responsible for ensuring that publication release occurs in accordance with the requirements of DOE Order 241.1, Scientific and Technical Information Management and the P726, Conflict of Interest: Privileged Information. Publications must comply with the provisions of the privileged information (e.g., Official Use Only) policy and all other policies in the Administrative Manual related to conflict of interest or export control.
>
> The organizations of individual authors are responsible for both the technical content of the work reported and the professional propriety of reports, papers, and presentations. Individual groups or divisions may have their own release policies and procedures, and individual authors are responsible for fulfilling their own organization's requirements. To process a Laboratory publication or presentation, SAFE-1 and IRM-CAS require a submittal form with the name and signature of a person accepting responsibility for assuring that all organizational policies and procedures have been followed. For further detail, consult your local derivative classifier (DC) or call SAFE-1 at 7-5013.
>
> The DOE definition of this information category is sufficiently broad to encompass nearly all information generated at the Laboratory that is not purely administrative in nature. In addition, DOE requires that SAFE-1 process all scientific and technical information "products," even if distribution to the public is not intended. Examples of the latter material would include classified or otherwise controlled reports. The publication release process includes review for classified information, Unclassified Controlled Nuclear Information, export controlled information, other unclassified sensitive information and information of patent interest.

I. Manuscripts, abstracts for conferences, viewgraphs for seminars and conferences, etc. A derivative classifier (DC) at LANL reads the document and determines 'which information is essential for the public and what scientific and technical

information developed under government-sponsored research and development is to be widely disseminated to the extent possible.' It is the responsibility of the DC to protect 'certain national security (national defense and foreign relations) information against unauthorized disclosure.' If the material is covered under a DUSA (designated unclassified subject area), it does not need to be read by a DC. (There is a DUSA for MFE.) A copy of the manuscript is submitted to SAFE-1. If it is determined that the document contains no classified or otherwise sensitive information, it is asigned a LA-UR number (UR-'unlimited release') and can be disseminated widely, e.g. published in a journal or shown at a conference. Typically this process, including reading by the DC, takes 1-2 weeks.

II. Computer software

A process similar to that for manuscripts, etc is used, to protect against dissemination of classified material, export controlled material, etc. An abstract of the source code is required. A LA-CC number is assigned after the review and the author is notified if the code is available for unlimited release or if there are any restrictions on its use.

III. Data

A similar process is used. A sample of the data must be provided, assuring that there is no security banner. A representative sample of the data and description of the allowable fields must be provided.

MATERIAL FROM THE INTERNAL LANL WEB SITE
http://int.lanl.gov/security/protectinfo/class/ (for LANL staff only)

Classification

The Classification Group's (SAFE-1's) mission is to identify information that requires protection in the interest of national security while ensuring that the Laboratory's scientific and technical work is disseminated to the maximum extent possible. A major part of this mission is to process and account for the Laboratory's publications. Specifically, classification is the process of determining and identifying information to be protected, and is distinct from security, which is the mechanism for protecting identified classified or other sensitive information.

I. Classification Review Procedures

SAFE-1 is responsible for reviewing or delegating the review of all types of publications that flow out from the Laboratory. More information below.

II. Publication Release and Accountability

The Publications Office (7-5013) is responsible for ensuring that publication release occurs in accordance with the requirements of DOE Order 241.1-1A, Scientific and Technical Information Management, and the Laboratory Administrative Manual (AM721 Conflict of Interest (pdf). Publications must comply with the provisions of the privileged information (ex: Official Use Only) policy and all other policies in the Administrative Manual related to conflict of interest or export control. More information below.

III. Derivative Classifiers (DCs)

Derivative Classifiers (DCs) begin the process of limiting dissemination of classified information by identifying its importance to national security. More information below.

MORE DETAIL (from links to http://int.lanl.gov/security/protectinfo/class/)

Ia. Classification Review Procedures

The Classification Group reviews or delegates the review of all types of publications that flow out from the Laboratory. Please see Publications Release and Accountability at the Laboratory.

Classification Analysts and designated Derivative Classifiers (DCs) are also resources for reviewing your NEW project for classification issues. Be sure that any work conducted outside of security areas is indeed unclassified. If you need a review, contact your local DC or the Classification Group if you are located where DC assistance is not available.

The following links provide information, but check with your local DC or the Classification Group for details.

IIa. Publications Release and Accountability

All scientific and technical information generated at the Laboratory and intended for public release must be reviewed and processed by Classification (SAFE-1) before publication or submission for publication. SAFE-1 is responsible for ensuring that publication release occurs in accordance with the requirements of DOE Order 241.1, Scientific and Technical Information Management and the P726, Conflict of Interest: Privileged Information. Publications must comply with the provisions of the privileged information (e.g., Official Use Only) policy and all other policies in the Administrative Manual related to conflict of interest or export control.

The organizations of individual authors are responsible for both the technical content of the work reported and the professional propriety of reports, papers, and presentations. Individual groups or divisions may have their own release policies and procedures, and individual authors are responsible for fulfilling their own organization's requirements. To process a Laboratory publication or presentation, SAFE-1 and IRM-CAS require submittal form with the name and signature of a person accepting responsibility for

assuring that all organizational policies and procedures have been followed. For further detail, consult your local derivative classifier (DC) or call SAFE-1 at 7-5013.

The DOE definition of this information category is sufficiently broad to encompass nearly all information generated at the Laboratory that is not purely administrative in nature. In addition, DOE requires that SAFE-1 process all scientific and technical information "products," even if distribution to the public is not intended. Examples of the latter material would include classified or otherwise controlled reports. The publication release process includes review for classified information, Unclassified Controlled Nuclear Information, export controlled information, other unclassified sensitive information and information of patent interest.

LABORATORY-APPROVED DOCUMENT DESIGNATORS

LA (Los Alamos Series Report)

LA-UR (Los Alamos Unlimited Release)

LA-CP (Los Alamos Controlled Publication)

LALP (Los Alamos Laboratory Publication)

LA-SUB (Los Alamos Final Subcontract Report)

LA-CRADA (Final CRADA Report)

LA-CC (Los Alamos Computer Code)

IIIa. Derivative Classifiers (DCs)

Derivative Classifiers (DCs) begin the process of limiting dissemination of classified information by identifying its importance to national security. They recognize which information is essential for the public and what scientific and technical information developed under government-sponsored research and development is to be widely disseminated to the extent possible. They also realize that the interests of the United States and its citizens require the protection of certain national security (national defense and foreign relations) information against unauthorized disclosure.

Appendix E: Example of Responses to Questions on Practices and Policies for Dissemination of Research Results – General Atomics

Facility	**GA/DIII-D**
Responder(s)	C.M. Greenfield (T.S. Taylor, D.P. Schissel)
1. The criteria for dissemination and who makes this determination.	DIII-D research results are disseminated in several ways. In general, any distribution of results in written form or as a collection of data that can be used for modeling, etc., is controlled in accordance with the DIII-D DATA USAGE & PUBLICATION POLICY AGREEMENT (see Appendix A). For the purposes of this report, I use the term "collaborator" to refer to anyone who has signed the agreement. We have an open data policy toward collaborators – any collaborator has full access to DIII-D data. If data has been published, it can be freely used by the community with the expectation that it would be properly referenced when used. Any release of unpublished data, for presentation, publication, or external use, is done at the discretion of individual scientists or the group, but requires technical review and sign-off by the DOE onsite representative as specified by the Agreement.

Facility	GA/DIII-D
2. ***How access is provided and controlled.*** Access could be provided through commercial or not-for-profit publishers or databases including archives, websites, and agency repositories.	A. Data may be presented at workshops and conferences. B. Placement of the presentation on a publicly available website is considered equivalent to publication, and is reviewed in that light. C. Publication in refereed journals or conference proceedings requires internal technical review. D. Direct access to the data on DIII-D's computer systems is only allowed to signatories of the Data Usage and Publication Policy Agreement. Collaborators are not permitted to disseminate the data to anybody who is not a signatory of the Agreement. E. Addition of data to community databases (e.g. ITER or ITPA) is generally not done until after it has been published. Once published, it is not controlled, although there is a reasonable expectation that any publications arising from use of this data would include proper references to the paper where the results were published. F. Other releases of unpublished data would require full technical review and sign-off by the DOE on-site representative.
3. Whether access is limited in any way. For both written findings and digital data, the distribution could be limited by, for example, subscription fees, technological barriers, by request only, or limited to the members of a particular research group. Furthermore, access may be exclusive for a limited period of time.	• Publication of written findings usually requires us to transfer the copyright to the journal and can only be accessed by paid subscribers or at a library. Many journals allow free access via the web during a short period following the initial publication. DIII-D also maintains a public website (https://fusion.gat.com/pubs/) where the internal report ("GA-A") versions of most of these publications are posted free of charge. These are generally earlier versions of the journal articles (prior to copyright transfer). • Data in digital form is stored on DIII-D servers, and is available to any collaborator (open data policy).

Facility	GA/DIII-D
4. Whether the access comes with any additional functionality. For written material, this could be interoperable, cross-publisher searches or federated search and discovery tools; links to data or other supplementary material used in the research (particularly if this ensures reproducibility of the research result); or multimedia; etc. For digital data, this could be the ability to reference the data as entered (or as part of a larger dataset), additional metadata or software interfaces for meaningful data mining by people outside the field, or interoperability with other data sets.	This varies with the type and origin of the data. Specialized software interfaces are available to collaborators, and in many cases are necessary to make use of the data.
5. The version of the written material or data provided. ***A.*** For example, for written findings, the Version of Record is usually considered to be the manuscript published and stewarded by the publisher; however, internal university or laboratory drafts may also be disseminated. ***B.*** For digital research data, a distinction may be drawn between data sets that are statically preserved and those that are continually updated; whether the data are considered "raw" or "analyzed"; and whether the data that support a particular finding can be referenced, for example, by a persistent identifier.	A. The version of record is usually the published manuscript and corresponding internal company document ("GA-A"). Unpublished data may be documented in a "GA-C," which has limited distribution. B. This varies among different kinds of data. Raw data, as a rule, is static and cannot be changed. In some cases, corrections may be made to the data soon after acquisition, and care is taken that the corrected version is used for dissemination and reference. In most cases, data and code versions can be traced back through analysis results.

Facility	GA/DIII-D
6. ***Whether peer review is a condition of dissemination.*** For written findings, a distinction could be drawn between external peer review, as usually happens with published articles, and an internal peer review as might happen within a Laboratory, university, or scientific collaboration for draft articles to be submitted for publication or conference proceedings. Any comparable review process for digital data should be described in the report.	• Any abstract, presentation, or manuscript that is published in the proceedings of a meeting or journal or "web publication" must go through several levels of formal internal review and be approved by the DOE on-site representative. • Unpublished digital data is documented in a GA-C (not-for-distribution) report which goes through the same review process. • Authors are encouraged, but not required, to subject these to informal peer review. • Presentations presented at public meetings are reviewed by staff prior to presentation.
7. ***The institution, DOE user facility, or other body by which the policy is currently upheld.*** Many Federal agencies, Laboratories, Universities, scientific collaborations, and user facilities have their own policies regarding the dissemination of research results including digital data. There may also be established practices that are not formally enforced by any institution but are broadly followed. For example, research communities may have dissemination practices that are followed, independent of agency/institutional requirements.	The DIII D National Fusion Facility Data Usage & Publication Policy Agreement satisfies both our internal requirements and those of our sponsor, the DOE Office of Fusion Energy Sciences.
8. Whether, in addition to dissemination, long-term stewardship is accounted for by the existing policy or practice. For digital data, the report could mention whether associated software for accessing data is also available and maintained.	Although this is not covered by the data policy, it has been our practice to safeguard and maintain all DIII-D historical and current data, so that it can be retrieved and used by any signatory of the DIII D National Fusion Facility Data Usage & Publication Policy Agreement.
Comments	

Facility	GA/DIII-D
1. How compulsory is dissemination of data? Who determines the obligation to disseminate the data?	• Participants in the DIII-D program are strongly encouraged to present and publish the results of their research. The decision to proceed is usually made by the researcher(s), although in cases where the data is deemed of particular interest to a broader audience, a request may come from program management. • GA management considers presentation and publication of relevant data in salary and career advancement.
2. What are the responsibilities and obligations of the originator of data and the providers/stewards of the archive in providing rules and structure for making the data useful/intelligible to others? What are the associated resource implications?	• Standard scientific ethics govern issues of quality control. • Specialized tools for accessing the data in an intelligible fashion are usually the responsibility of the researcher producing the data (e.g. the diagnostician). • General tools and data access support is available from our Computer and Data Analysis Applications Groups. • The DIII-D program provides infrastructure for making the data available (centralized storage, servers, etc.)
3. What protections are there against abuse from those requesting access to and assistance with the data?	• The data use policy requires review of the manuscript or presentation before dissemination. • We rely on professional courtesy and conduct for scientists not to misuse fellow scientist data or analysis.

Facility	GA/DIII-D
4. How do intellectual property rights, copyrights and patent issues affect sharing of research results and data?	• Intellectual property rights , copyrights, and patent issues do not affect the sharing of the data. • The review process identifies items that would fall under new or patentable material, and that is handled accordingly. The material is reviewed by DIII-D management, GA legal, and the DOE on-site representative. Since the research is funded by DOE, cases where the review is held up for a patent clearance are very rare. • When published manuscripts require a transfer of copyright, usage of the final version of the paper is restricted by the journal.
5. With respect to digital data, what are the mechanics for defining common data formats, arranging for access (data servers inside or outside firewalls), and providing tools for and assistance in processing the data?	• Tools are provided by the researchers and/or our Computer and Data Analysis Applications Groups, and are generally tailored to the needs to local researchers. We can and have coordinated with other groups to make our data more broadly available. An example is the ITPA profile databases, which were designed with our input. Our data is formatted and exported to those databases (usually after publication) using tools developed locally.

Appendix F: Examples of Policies and Practices Regarding Dissemination, Licensing and Commercialization of Research Results -- University of Texas, Austin

Reference: http://www.otc.utexas.edu/index.jsp

1. **The University of Texas System Rules and Regulations of the Board of Regents**
Rule: 90101 Rules for Intellectual Property: Purpose, Scope, Authority
http://www.utsystem.edu/bor/rules.htm - A10

2. Protect Your Intellectual Property Rights
http://www.otc.utexas.edu/publications/ProtectYourIPRights.jsp

3. The 8 Steps of Technology Commercialization
http://www.otc.utexas.edu/publications/8StepsOfTechCommercialization.jsp

4. Commercialization for Inventors
http://www.otc.utexas.edu/InventorComm.jsp

5. General Technology Licensing
http://www.otc.utexas.edu/IndustryComm.jsp

6. Inventor Forms
http://www.otc.utexas.edu/InventorForms.jsp

7. Administrative Policy on Disclosure, Distribution and Licensing of Software
http://www.utsystem.edu/ogc/INTELLECTUALPROPERTY/swadmpol.htm

8. Software Licensing and Copyrights
http://www.otc.utexas.edu/SoftwareAndCopyrights.jsp

9. OpenSource Toolkit
http://www.otc.utexas.edu/publications.jsp

Appendix G: THE STATE OF DATA MANAGEMENT INTHE DOE RESEARCH AND DEVELOPMENT COMPLEX -- Report of the Meeting “DOE Data Centers: Preparing for the Future”

http://www.osti.gov/publications/2007/datameetingreport.pdf

Appendix H: OFES Data Sharing Expectations for Large-Scale Computational Projects

Introduction and Background

Large scale computations—including projects supported by the Office of Science's Scientific Discovery through Advanced Computing (SciDAC) program—have increasingly assumed a vital role in addressing the scientific challenges facing the U.S. Fusion Energy Sciences Program as equal and complementary partners with the more traditional approaches of analytic theory and experiment. High performance simulation codes developed under the auspices of SciDAC and the OFES base program have provided us with new and significant insights into the physical mechanisms responsible for turbulent transport from the core and edge of magnetically confined plasmas, the interaction of electromagnetic waves with plasmas, and the mechanisms responsible for the macroscopic stability of present and next generation fusion devices. In the coming era of burning plasmas and ITER, large scale integrated simulations will be essential for the design, operation, and interpretation of results from future experiments.

With such responsibility comes the challenging task of verifying and validating the predictions of these codes. Considering the difficulty and significant computational expense of replicating results from massively parallel codes as well as the inherent limitations of analytic theory as a comprehensive verification approach for the highly nonlinear problems addressed by these simulations, cross-benchmarking among different codes is an indispensable and often-used verification tool which also leverages the OFES investment in the development of codes based on different technical approaches. The success of this approach relies heavily on the availability and sharing of simulation data and other supporting materials in a timely fashion and at no more than incremental cost by investigators engaged in large-scale simulation research.

Data Sharing Guidelines

For the reasons stated above, the OFES expects the timely sharing of simulation data among OFES-funded researchers engaging in large-scale computational research. At the same time, OFES recognizes the right of individual scientists and research groups to get fair credit for their work by establishing initial periods of *exclusive use*, defined as the period between the generation of the simulation data and publication and/or presentation of research results based on these data. During the period of exclusive use, PIs are not required to share their data with others.

In more detail:

- OFES strongly encourages the PIs of its large-scale computational projects to ensure that: (i) their most important findings be published in a timely way in peer-reviewed journals; and (ii) their results be presented at major conferences and workshops that are widely attended by the members of our community.
- Following publication in a peer reviewed journal, or after a year following a major conference presentation—whichever comes first—simulation data should be available for sharing upon request
- Users of shared data should consult with the donating author and try to reach a consensus on any technical issues pertaining to these data and the simulations that generated them before publishing or presenting results based on these data

- Users of shared data are expected to give proper credit to the researchers that generated them in any publication or presentation that makes use of these data
- Providing data and assisting in their interpretation can be time and resource consuming, especially for smaller research groups. Those requesting data should be aware of this fact and be reasonable in their requests, including being prepared to possibly share in the associated costs and labor
- Research teams involved in large-scale simulations should designate a point-of-contact person responsible for data sharing issues
- In the event that data sharing disputes or issues emerge, OFES will work to help resolve associated problems in consultation with the investigators