

[Assigned ID #]

[Assigned Entry date]

**Name/Email**

Morteza Gharib, Vice [Provost](#)

**Affiliation/Organization**

California Institute of Technology

**City, State**

Pasadena, CA 91125

Caltech is a PhD university employing 922 principal investigators whose research funding comes largely from 6-10 different federal agencies. In addition, Caltech is committed to education and recognizes its profound obligation toward public dissemination of its research results ideally unfettered by the demands of commercial profit so that learning and discovery, two major pillars of the enterprise, will thrive. The global network provides the means to ensure maximum access for uptake of new knowledge via electronic distribution of publicly funded research results. Therefore, Caltech urges and supports action to require prompt public access to results of all government funded research.

In response to the request for information from your office released on November 3, 2011 on the topic of public access to peer-reviewed scholarly publications resulting from federally funded research we offer the following comments.

922 includes professorial faculty, research faculty and postdoctoral scholars.  
The Dept. of Defense is counted as one agency.

*Comment 1:*

First, new markets frequently result from innovations as outcomes of research. Since the work of the economist, Dr. Edwin Mansfield in the mid to late 20<sup>th</sup> century, the synergistic relationship between research, industrial innovation, and commerce or social return has been well-established. Now the global network has created a communication revolution in which immediacy and reusability are significant variables in productivity or social return on research investment. Therefore more current economic research has delved into the different methods or business models for distributing peer-reviewed research papers, the primary vehicle for communicating reliable research results.

Dr. John Houghton of the Centre of Strategic Economic Studies at Victoria University has conducted a number of such studies for the U.K., as well as for Australia, Denmark, and even the U.S., showing that unfettered access to the results of publicly funded research have a significant positive economic impact for a country as a whole. See:

<http://www.cfses.com/projects/knowledge-access.htm> . In the U.K these results are taken seriously enough that government policy is actively shifting to ensure that the public (all individuals, all learning, all research, and all commercial entities regardless of size or means) can actively benefit on their own terms from government funded research. See: UK Department for Business Innovation and Skills report, <http://www.bis.gov.uk/assets/biscore/innovation/docs/i/11-1387-innovation-and-research-strategy-for-growth.pdf>

The U.S. should be concerned that small companies and entrepreneurs experience significant barriers to federally funded research results. Eight new companies are launched each year, based on Caltech intellectual property, yet the staffs of these small enterprises cannot readily access research papers from their work place as purpose and needs dictate. This circumstance results in inefficient use of time and unreasonable hurdles to staying abreast of developments since information seeking and use cannot be seamlessly integrated into the work.

Opening up access to the research papers will level the playing field to allow new approaches, new companies to provide value-add services. With accessible, reusable, digital access to publicly funded research papers, new businesses can compete by offering computational tools for data mining, subject or linguistic analysis, bibliometrics, indexing, alert services, and more. The more ways that information from research results can be utilized, the more productive will be society. The demand for unfettered access to research papers in order to create value-add services using network based technology is illustrated in the following 14 services (and there are more) that have sprung up in just the last few years: Academia.edu, Epernicus | Network, Google, Scholar Citations, iamResearcher, JournalFire, Laboratree, Mendeley, Microsoft Academic Search, Nature Network, PeerEvaluation, ResearcherID, ResearchGate, Researchr.

The federal government has a number of successful models in which a basic level of formatted information is made publicly available (census data, patents, Securities and Exchange Commission information), that and other third parties (e.g. Economagic, Derwent, Intellectual Property Network, Morningstar Document Research) reuse and augment according to diverse business models. In this way the government is not responsible for generating all possible uses and presentations. Instead the private sector performs those services and competes for clients.

**Secondly, optimal scientific research productivity requires immediate and full public access to all government funded research results** so that curious and creative minds, no matter where they are located, have opportunity to build on that work in whatever way is useful and productive. Caltech research applies first principles to problem areas of medicine, energy, and the environment, among many fields, - ultimately all critical areas of social need. Specific examples include: drug delivery systems, bio-inspired fluid flow systems, design of prostheses and cornea implants, heart pumps, signal processing, and communication in regard to intercellular signaling, mechanics of cell scattering leading to metastasis using a digital volume correlation, insect wood digesting enzymes, multiscale models for large-scale engineering on the scale of earthquakes, green IT, photovoltaics, multi-junction semiconductor composites for water purification, cheap catalysts for solar power, artificial photosynthesis, and more. These interdisciplinary research problems involve a broad range of federal agencies are involved. For the scientific research community to be optimally productive, access to all the research results from all the agencies is necessary.

*Comment 2:*

The current copyright framework allows licensing as needed to be flexible in meeting the needs of the creator, the funder, and those entities that provide services. This flexibility is exemplified in the Creative Commons Attribution (CC-BY) licenses that can be imbedded with XML tagging into documents. CC-BY is a copyright license that grants permission to the public to reproduce,

distribute, perform, display or adapt the licensed materials for any purpose so long as the user gives attribution to the author. See: <https://creativecommons.org/licenses/by/3.0/>

Thus, the rights travel with the object, thereby encouraging use and greater, faster impact because the rights are immediately clear. There is no delay for special permissions for uses that further learning and discovery as is the predominant intent of authors of research papers. The efficacy of the CC licenses vis-à-vis the copyright law has been upheld in the courts, and these licenses are enforceable under current law. See [http://wiki.creativecommons.org/Case\\_Law/](http://wiki.creativecommons.org/Case_Law/) It is entirely appropriate that publicly funded government agencies assert at the outset certain rights to research papers on behalf of the public. The NIH 2008 policy operates on that basis and has been successful. The one shortcoming that needs correcting is the restriction on further public use.

Copyright was intended in the U.S. Constitution to “To promote the Progress of Science and useful Arts...” It is time to “shift from a model that uses copyright to control reuse of content to one that uses copyright to encourage republication, preservation, and translation.” (Carroll MW , 2011 Why Full Open Access Matters. *PLoS Biol* 9(11): <http://www.plosbiology.org/e1001210>, [doi:10.1371/journal.pbio.1001210](https://doi.org/10.1371/journal.pbio.1001210)) . In short an IP model that maximizes the dissemination and uptake of a researcher’s output is in the best interest of the research authors, the originators of the content, and the public. In contrast to the author of a book, who may anticipate royalties on sales, those of research papers seek to make their work available to others. Their compensation comes in the form of recognition for their contribution to the field.

*Comment 3:*

There will need to be flexibility. We already see that a centralized repository such as PubMedCentral can work and yet there is also a role and purpose for institutional repositories. The main issue is that the administration of those reference repositories is committed to archiving, preservation and unfettered dissemination.

A centralized repository whose content can be downloaded may enable many analytic or discovery tools specific to a research field or project. WormBase at Caltech is a biological database that extracts information from scholarly publications and puts it into computable form. Some of the steps in this process are automatable using machine learning methods applied to article texts., for example, identifying articles worth reading by the professional curators. While curators can obtain through site license or purchase articles with key information, it is not practical to obtain all biomedical articles and then analyze their text to decide if it is worth human effort. A centralized repository of all scholarly articles that can be accessed electronically is crucial to allow efficient use of human effort.

No entity (private, for-profit, public, non-profit) should be hindered from harvesting and re-using scholarly output with attribution. Universities should be able to position their output for maximum distribution and measurable impact.

*Comment 4:*

Publishers are not cultural memory institutions over the long haul. They are businesses focused on current commerce. Universities and their libraries have performed the archival and

preservation role enabling discovery and access to the nearly infinite long tail of information for hundreds of years and will continue to do so. The shift to private, corporate control of the archives in electronic databases puts future continuity at risk.

Businesses do not naturally collaborate. They compete and are designed as proprietary silos. This model is suited to competitive services but it is not productive when unique results of publicly funded research are barricaded behind a publisher wall.

Partnerships between publishers and academic libraries and government are possible by teasing apart the social roles consistent with the funding models. Valuable publisher services must and can be remunerated on a price for service basis, but not at the cost of public access. Libraries are funded through their institution to provide the basic archival persistent infrastructure needed for learning and scholarship over the long haul while the government establishes the rules to ensure public benefit for publicly funded research.

*Comment 5:*

Standards are going to evolve over time and will need to be incorporated into repository activities as appropriate at the time. A basic specification of Dublin Core, OAI-PMH (Open Access Initiative Protocol for Metadata Harvesting), Data Cite Metadata Schema is needed at this time. In addition the OAI-ORE (Open Access Initiative Object Reuse and Exchange) must be included to facilitate computational tools and reuse. Licensing conditions are another concern. CC-BY licensing of content greatly facilitates interoperability and the creation of discovery services.

Nevertheless, publicly funded research results must first be made publicly accessible. Then clear national standards for interoperability can be formally designed and promulgated to ensure that research and business can optimally benefit,

*Comment 6:*

A basic consistent, mandatory policy (research results must be publicly accessible) across all federal agencies is needed so that all stakeholders have a single set of rules to follow for federally funded research.

The government can leverage infrastructure already developed within the NIH and build on tools already in development e.g. SWORD (Simple Web-Service Offering Repository Deposit).

Benefits will accrue in other spheres such as grant reporting, bibliography generation, return on investment measurements, and patent office review, reducing duplicative effort and documentation across the public sector. On a smaller scale, this is already happening at Caltech where the local digital repository dynamically delivers accurate descriptive metadata to researchers' web pages saving time in re-keying and editing. The government stands to gain from an order of magnitude improvement in output re-use.

*Comment 7:*

Before considering extending a public access requirement to other formats, first establish it firmly for peer-reviewed research articles. The rest will logically follow.

All federally funded projects need to include a statement regarding required public access to the certified results in whatever genre they are finalized for earliest distribution.

*Comment 8:*

The ultimate goal must be no embargo for optimal return on investment through research and business productivity. Internal studies at Caltech show that one third of researchers' reading or use activity involves the most current year's papers. Therefore, the greatest benefit to research productivity will be achieved by making papers immediately accessible.

A transition period starting with an embargo of six months for a defined period of time is appropriate. A twelve-month embargo is much too long and thwarts timely uptake of new information.