ISCB Public Policy Statement on Open Access to Scientific and Technical Research Literature

Richard H. Lathrop, Chair, ISCB Public Affairs Committee
Burkhard Rost, ISCB President
For the ISCB Membership, Executive Committee, Board of Directors, and Public Affairs Committee
Current signatories are at: http://www.iscb.org/iscb-policy-statements/literature_open_access/signatories

Preamble

The International Society for Computational Biology (ISCB) is dedicated to advancing human knowledge at the intersection of computation and life sciences. On behalf of the ISCB members, this public policy statement expresses strong support for open access, reuse, integration, and distillation of the publicly-funded archival scientific and technical research literature, and for the infrastructure to achieve that goal.

Knowledge is the fruit of the research endeavor, and the archival scientific and technical research literature is its practical expression and means of communication. Shared knowledge multiplies in utility because every new scientific discovery is built upon previous scientific knowledge. Access to knowledge is access to the power to solve new problems and make informed decisions. Free, open, public, online access to the archival scientific and technical research literature will empower citizens and scientists to solve more problems and make better, more informed decisions. Attribution to the original authors will maintain consistency and accountability within the knowledgebase. Computational reuse, integration, and distillation of that literature will produce new and as yet unforeseen knowledge.

We strongly encourage open software, data, and databases, issues which are not addressed here. A prior ISCB public policy statement on sharing software provides very clear support for Open Source/Open Access (http://www.iscb.org/iscb-policy-statements/software_sharing). We support open database access, standards, and interoperability. We also recognize that databases are complex dynamic entities, with ongoing roles and needs that cannot be treated properly within this statement. In contrast, the publicly-funded archival research literature, addressed here, is the static historical record of publicly-funded research outcomes.

ISCB supports many of the principles set forth in other open access policies and statements, including the 'Budapest Open Access Initiative,' the 'Bethesda Declaration on Open Access Publishing,' the Bulletin of the World Health Organization 'Equitable Access to Scientific and Technical Information for Health,' the U.S. National Academies of Sciences report on 'Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences,' the Organisation for Economic Co-Operation and Development 'Principles and Guidelines for Access to Research Data from Public Funding,' and the 'Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities.'

The public policy statement put forward here builds upon these principles to elucidate in more detail the public policy position of ISCB and its members on this important issue in scientific dissemination.

Public Policy Statement

> **The International Society for Computational Biology strongly advocates free, open, public, online: (i) access by person or machine to the publicly-funded archival scientific and technical research literature; and (ii) computational reuse, integration, and distillation of that literature into higher-order knowledge elements.**

Supporting Statements

(1) The possibilities latent in the digital information age make it essential to achieve open access, and computational reuse, integration, and distillation, of the publicly-funded archival research literature.

    (a) Immediate access is preferable, and when access is at an interval following publication, that interval should not exceed one year.

    (b) At a minimum, every scientific journal should offer an open access option to every published research paper, as does every official or affiliated journal of the ISCB.

    (c) Copyright licenses explicitly should permit computational reuse, integration, and distillation, using standard existing language that eliminates the need for manual or legal review.

    (d) The format of the available article should be easy to parse by both human and machine (e.g., HTML). Ideally, a plain text version should be available as well (e.g., TXT), to facilitate computational reuse and integration (e.g., computational text mining for knowledge extraction).

    (e) Computational reuse, integration, and distillation should give attribution to the original authors.

---

(2) Existing open access models show high impact, scientific benefit, feasibility, and acceptability.

    (a) The public benefit from open access to the world's online information via the publicly-funded Internet provides a good model of expected impact.

    (b) The scientific fertilization from open access to genomic information via the publicly-funded Human Genome Project provides a good model of expected scientific benefit.

    (c) Open access policies by the U.S. National Institutes of Health, the Howard Hughes Medical Institute, and the Wellcome Trust provide good models of feasibility and acceptability.

    (d) The Creative Commons Attribution License and the Science Commons Open Access Data Mark provide good models of legal mechanisms for computational reuse, integration, and distillation.

(3) Open literature access, reuse, integration, and distillation will enable a whole new generation of innovative computational tools and processes. The literature will be endowed with enriched commentary and usability. It will be connected seamlessly, by proper semantic links, to relevant websites, data, databases, and algorithms. Creating a web of knowledge around publications is an important consequence of semantic enrichment of the research literature. Such tools already are being built by publishers, researchers, entrepreneurs, and others. The further development of these tools should be supported aggressively. Removing all barriers to literature open access, reuse, integration, and distillation is critical to achieving such a knowledge-level transformation.

(4) Supplementary data and methods should be openly available online, in sufficient detail to replicate the reported research results and facilitate reuse. Such material should be deposited in appropriate public repositories, in compliance with accepted community standards, and in accord with the existing ISCB public policy statement on sharing software. It should allow for application of other computational methods to the data and application of other data to the computational methods.

(5) Publishing high-quality peer-reviewed scientific literature incurs costs. We recognize that cost recovery is a serious issue that must be addressed carefully if open access is to be a mandated policy.

(6) Open access policy details—which version, where stored, how annotated and organized, what incentives, etc.—must be considered carefully. However, it has now become essential to put forward a broad policy mandate for public access to, and computational reuse, integration, and distillation of, the publicly-funded archival scientific and technical research literature.

(7) This statement is intentionally neutral about any specific funding policy. Many implementations all may achieve the same essential goal. Acceptable funding policies should:

    (a) Remove barriers to open access and subsequent computational reuse, integration, and distillation.

    (b) Encourage public, private, and philanthropic funding organizations to establish policies that mandate free, open, public, online access to, and computational reuse, integration, and distillation of, the research results funded from their public, private, or philanthropic support.

    (c) Promote the body of publicly-funded archival research literature as a public investment that bears interest, and not as an ongoing access cost to the public.

    (d) Establish copyright licenses in standard terms that permit literature access, reuse, and integration.

    (e) Specify a format that is easy to parse by both human and machine (e.g., HTML); and, ideally, also provide a plain text version (e.g., TXT) to facilitate computational reuse and integration.

    (f) Recognize the need to fund activities of peer review, copy editing, and publishing.

    (g) Provide fairness to several groups, including the developing world and its health concerns, unfunded or under-funded researchers, and others.

    (h) Provide fair interim support or compensation, if and where needed, to facilitate making transitions and adaptations to new models for publishing and sustaining essential revenue.

    (i) Be consistent with government laws, patent requirements, other existing regulations, and research dissemination through viable commercial mechanisms.

(8) The expected cost of complete open access to the publicly-funded archival research literature is only a very small percentage of the entire publicly-funded international research endeavor. Nevertheless, it is undesirable to divert funding from current research and thus risk underfunding basic science. New funding should be made available for open access policy implementations.

Conclusion

Currently, scientific advancement is limited by article availability, access costs, copyright restrictions, document formats, bulk download limits, etc. All such barriers should be removed.

The publicly-funded archival scientific and technical research literature represents a substantial investment by the public, governments, foundations, non-profit institutions, publishers, individuals, and others. We in the ISCB are committed to the continuous enhancement and

leveraging of society's knowledge resources. One of our primary missions is the computational integration of individual pieces of knowledge from the research literature and databases, in ways that provide powerful new ideas and insights for next-stage research, to the benefit of the scientific community and society in general.

To achieve these public benefits, we strongly advocate free, open, public, online access to the publicly-funded archival scientific and technical research literature, and the computational reuse, integration, and distillation of that literature into higher-order knowledge elements.

Appendices

Appendix A. Example Scenario

An automated malaria website might access location-specific information from thousands of publicly-funded malaria research articles daily, and then integrate that information into a free online interactive world map. Such a map might be annotated with up-to-date information about disease occurrences, drug resistance profiles, current best control practices, etc., as distilled from the research literature extracted for and attached to each local region. A hypothetical user might be a public health official in the developing world responsible for controlling a sudden malaria outbreak in a remote area. Such a website should encounter no barriers while performing this free, useful, and potentially essential public service.

Example Discussion

A search for 'malaria' in the U.S. NIH/NLM PubMed literature database yielded more than 55,000 hits (July, 2010; http://www.ncbi.nlm.nih.gov/pubmed/). The publicly-funded portion of these 55,000 'malaria' hits should be freely available in bulk to this hypothetical malaria website, using technologies well suited to bulk tasks, for purposes of: (i) the initial bulk literature download; (ii) regular updates; and (iii) intermittent bulk repeat downloads to reinitialize an improved knowledge base. The relevant copyright permissions should permit computationally recombining the publicly-funded portion of these 55,000 texts into whatever final form is most useful and informative to the user.

This example scenario illustrates an important public health benefit that could be achieved immediately: the opportunity to pursue useful knowledge-based innovations, by computational reuse, integration, and distillation of the publicly-funded archival research literature, across many areas in biology and medicine.

Appendix B. Documents Mentioned in the Statement Text[1]

(1) *Text of ISCB public policy statement on sharing software*
http://www.iscb.org/iscb-policy-statements/software_sharing

(2) *Text of the 'Budapest Open Access Initiative.'*
http://www.soros.org/openaccess/read.shtml
http://en.wikipedia.org/wiki/Budapest_Open_Access_Initiative

(3) *Text of the 'Bethesda Declaration on Open Access Publishing.'*
http://www.earlham.edu/~peters/fos/bethesda.htm

(4) *Text of the Bulletin of the World Health Organization 'Equitable Access to Scientific and Technical Information for Health.'*
http://www.scielosp.org/scielo.php?script=sci_arttext&pid=S0042-96862003001000003

(5) *Text of the U.S. National Academies of Sciences report on 'Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences.'*
http://www.nap.edu/openbook.php?isbn=0309088593

(6) *Text of the OECD (Organisation for Economic Co-Operation and Development) 'Principles and Guidelines for Access to Research Data from Public Funding.'*
http://www.oecd.Org/dataoecd/9/61/38500813.pdf

(7) *Text of the 'Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities.'*
http://www.eprints.org/events/berlin3/outcomes.html http://oa.mpg.de/openaccess-berlin/berlindeclaration.html
http://en.wikipedia.org/wiki/Berlin_Declaration_on_Open_Access_to_Knowledge_in_the_Sciences_and_ Humanities/

(8) *Human Genome Project.*
http://genome.gov/ http://genome.energy.gov/

(9) *Text of Open Access Policy from the U.S. National Institutes of Health.*
http://publicaccess.nih.gov/policy.htm http://publicaccess.nih.gov/

---

[1]These URLs were correct when this statement was written, but are by their nature ephemeral and not archival.

(10) *Text of Open Access Policy from the Howard Hughes Medical Institute.* http://www.hhmi.org/about/research/sc320.pdf
http://www.hhmi.org/about/research/Q A_papp.pdf

(11) *Text of the Wellcome Trust's 'Position Statement in Support of Open and Unrestricted Access to Published Research.'*
http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/Policy/index.htm

(12) *Text of the Creative Commons Attribution License.*
http://creativecommons.Org/licenses/by/3.0/

(13) *Text of the Science Commons Open Access Data Mark.*
http://www.sciencecommons.org/projects/publishing/open-access-data-protocol/

Appendix C. General Background Material, Other Statements and Materials[2]

(1) *Academic publishing— Wikipedia, the free encyclopedia.*
http://en.wikipedia.org/wiki/Academic_publishing

(2) *Open access (publishing) — Wikipedia, the free encyclopedia.*
http://en.wikipedia.org/wiki/Open_access_(publishing)

(3) *ROARMAP (Registry of Open Access Repository Material Archiving Policies) as recommended by the Berlin Declaration.*
http: //www.eprints.org/openaccess/policysignup/

(4) *Text of Public Library of Science 'Open Letter to Scientific Publishers' (signed by -34,000 scholars worldwide).*
http://www.plos.org/about/letter.php

(5) *Text of Research Councils of the UK 'Access to Research Outputs.'*
http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/documents/2006statement.pdf
http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/news/oareport.pdf
http://www.rcuk.ac.uk/research/outputs/default.htm

(6) *Text of European Research Advisory Board Final Report 'Scientific Publication: Policy On Open Access.'*
http://ec.europa.eu/research/eurab/pdf/eurab_scipub_report_recomm_dec06_en.pdf

(7) *Open Science Directory.*
http://www.opensciencedirectory.net/

(8) *Peter Suber's 'Open Access Overview.'*
http://www.earlham.edu/~peters/fos/overview.htm
http://www.earlham.edu/~peters/fos/brief.htm

(9) *Scholarly Kitchen on the Open Access Financial Model*
http://scholarlykitchen.sspnet.org/2010/09/14/why-the-open-access-financial-model-will-continue-to-transmogrify/

---

[2]These URLs were correct when this statement was written, but are by their nature ephemeral and not archival.