American Chemical Society (ACS) and PubChem

May 23, 2005

Introduction

Some of the press coverage, mailing list, and blog communications on this subject suggest that there is some confusion, misunderstanding, and in places, polemics, surrounding ACS/Chemical Abstracts Service (CAS) concerns with NIH’s PubChem project. The following open statement, we hope, will answer most, if not all, of the questions that have been forwarded to us, or are found on the Web. CAS is a division of the ACS and in this statement all references to CAS and ACS should be taken to refer to the same organization and public position.

Summary

The ACS believes strongly that the federal government should not seek to become a taxpayer-supported, competitive scientific publisher. By collecting, organizing, and disseminating small molecule information whose creation it has not funded and which duplicates CAS services, NIH has started, rather ominously, down the path to unfettered scientific publishing. ACS has been a primary and secondary scientific publisher, and served the scientific community, since the late 1800’s. ACS’ position is straightforward. We have asked NIH to “refocus” PubChem, not discontinue it, but refocus it, on the stated mission: linking and communicating data created by the Molecular Libraries Screening Center initiative which NIH funds. If NIH would agree to simply focus PubChem on its stated purpose, the controversy would end immediately.

We believe that the federal government has a legitimate interest in seeing research it funds published. To the extent that NIH pursues that goal, whether through their RoadMap for Medical Research initiative or otherwise, ACS has not objected. ACS cooperates with NIH in a range of ways and has long been a staunch supporter of federal funding for basic research. Therefore, we begin with the fact that ACS has not asked, and does not ask that NIH “shut down” PubChem.

As things stand, in PubChem, NIH has created a mini- replica of the CAS RegistrySM, and a replica poised to expand. That replica will, over time, pose an insurmountable threat to CAS’ survival for the very reason that it is a taxpayer-supported resource. The fact that the data collected into PubChem is “public domain” is completely irrelevant. Assembling information and publishing it in a variety of forms is what the private sector does. We believe that taxpayers should not fund the entry of NIH into the information industry more broadly than is necessary to disseminate the information whose creation it funds.

Today, CAS competes against (and in some cases cooperates with) major information conglomerates as well as private sector web services. CAS understands that its services must meet customer demands; CAS must innovate, price to market, and provide world-class service. CAS has welcomed competition and sought to fulfill its mission and renew itself by providing world-class products and services.

Frequently Asked Questions

Are PubChem and the CAS Registry complementary or competitive?

NIH has duplicated both the platform and content of the CAS Registry. The platform replicates the searching features of the Registry and information such as the structure, name and synonyms, properties, and identifier. These are exactly the fields contained in the Registry. Regarding content, PubChem already replicates data on nearly 1 million substances, with millions more planned. ACS has demonstrated this to NIH as well as representatives of the press. Side-by-side comparisons of substance records from the PubChem and CAS Registry demonstrate that they contain the same core content. If a scientist obtains this data from PubChem, there is no reason to purchase it from the CAS Registry. We, therefore, believe that
statements that PubChem is “complementary” to the CAS Registry, rather than competitive, are
wrong.

PubChem has fewer than a million substance records. How can it be a threat to the CAS Registry?

For now, PubChem offers a smaller range of substances. However, as millions more substances are added, it will more and more mirror the Registry service. Clearly, ACS needs to respond to the likely impact of PubChem if it is allowed to develop as planned. It would be irresponsible to wait until PubChem has five or ten million substance records to raise an alarm.

PubChem’s records come from publicly available, mostly government, sources. Why does CAS object to that?

For now, the main issue is not whether publishing the records violates copyright. Instead, the issue is whether an arm of the government should involve itself in general information publishing by replicating an existing private service. By aggregating existing sources, creating a platform for future aggregation, and by announcing its intention to grow PubChem “without limits,” NIH has done just that. It is no solace to us—or excuse for NIH—to say that the foundation for PubChem, for the moment, is an aggregation of “publicly available databases.” The purpose of the foundation is clear, NIH’s intent is clear, and the future direction has been charted. It is that future that we question.

If PubChem is critical to the NIH RoadMap, then shouldn’t ACS simply step aside for the good of the health of U.S. and world citizens?

We don’t accept the notion that a Registry replica is needed for the NIH RoadMap. This assertion is made by NIH, but never with any proof or explanation. To put it simply, if the goal is to distribute bio-assay data, why doesn’t NIH simply set up a “switching station” to link scientists to that data when it is created? In fact, CAS has offered to help in that process, so long as the objective is reasonably limited to that goal.

NIH says that PubChem will open new markets for CAS and is happy to link to CAS services...

CAS believes that it is able, on its own, to reach research markets with its services. SciFinder Scholar™, for example, which offers the Registry to universities around the world, is now installed in well over 1,000 universities. It has enjoyed a steady rate of growth, and is the most accepted chemical information research tool in the world. STN® on the WebSM is available around the world as well, and features the Registry.

CAS knows of no scientific markets that can only be reached by NIH services and which will then be available to CAS through the new PubChem. In fact, it is obvious that damage to CAS’ existing market will far outstrip any incremental business generated by PubChem and “links to CAS.” Far from being innocent and “complementary,” PubChem is already drawing use by CAS customers, which is only rational behavior given that its costs are being underwritten by the U.S. taxpayers.

Isn’t a “free” service, such as PubChem, good for science and the scientific community? We understand that NIH has criticized CAS’ “Business model.” Is that model outmoded?

PubChem may be free to the user, but it is taxpayer subsidized. CAS has been in the publishing business for nearly one hundred years and has served world science during that time. It now seems strange to be told that there is a new public policy that our services should be provided free by the federal government. One might ask if NIH ought to use its funding to manufacture pharmaceuticals and provide them free to the public. Are free pharmaceuticals more directly in the interests of citizens than free information? If the Registry were a pharmaceutical, this controversy would be concluded rapidly. The fact is that the Registry is an information service,
with significant annual revenues. Building and distributing that service employs more than 1,000 staff at CAS. The federal government ought not to simply decide to replicate it and provide that service at the taxpayer’s expense, no matter how altruistic this may sound on the surface.

CAS’ business model involves private sector development of a valuable database and sale of that database worldwide. It does not seem proper that NIH should have the authority to declare that model of free enterprise outmoded. This would appear to be a statement of national policy reserved for Congress.

Is ACS overstating the potential harm to CAS from PubChem? After all, NIH only has a small number of staff and a small budget assigned to this project.

Despite our requests, NIH has never responded specifically to a request for information detailing the budget and number of staff working on PubChem. The fact is, NIH can draw on vast resources, whether or not they are “cost accounted” to this project. They have consultants, related organizations, related projects, etc., all of which can be drawn upon to build PubChem. But, in the end, whether twenty or two hundred government employees build PubChem, it is still an inappropriate use of government staff and resources.

As to the damages…

The CAS Registry is the centerpiece of CAS services. It has been developed over nearly forty years by CAS scientists. If a competing, taxpayer-supported service becomes available, a significant portion of CAS revenues will be eliminated, restricting CAS’ ability to invest in future database building efforts as well as technology developments needed to stay competitive. In the long run, CAS will be unable to hire and retain staff. We believe it is fair to say that a competing and taxpayer-supported Registry would eliminate CAS as a viable publishing organization. If that were to happen, the valuable research tools used by thousands of scientists around the world would disappear.

What steps has ACS taken to resolve this issue?

ACS officials met with the NIH and expressed concerns about PubChem several times in letters, meetings, and phone calls dating back to 2004. When it was clear that NIH intended to pursue PubChem, with no acknowledgment of the risks to CAS, ACS contacted interested Members of Congress and other policymakers.

ACS will continue to seek a resolution that focuses NIH on distribution of information it funds, and allows CAS and its staff and resources to continue to focus on the distribution of chemistry-related information, especially the CAS Registry, as CAS has for nearly 100 years.

Did CAS receive funding from the U. S. Government to build the CAS Registry?

Between thirty and forty years ago, CAS received funding from the NSF to support initial technical development of the CAS Registry system. In accordance with agreements, CAS subsequently published extensively about this technology platform, putting that information in the public domain. The investments CAS has made over the past forty years in building the Registry since that time dwarf the amounts provided by the Federal Government. Furthermore, it is clear that the U.S. Government encouraged CAS to develop the Registry and generations of scientists at CAS have fulfilled that trust by developing and enhancing the system since that time. The fact that NSF turned to CAS decades ago to create the Registry in no way justifies NIH in replicating it today.

Where can I get more information on this issue?

Media representatives may call 614-447-3847 in Columbus or 202-872-4400 in Washington, D.C. For questions about CAS products or services, call 800-753-4227 (North America) or 614-447-3700 (worldwide).