

Oxford, Harvard scientists lead data-sharing effort: New standards allow disparate data sets to integrate

January 29 2012

Led by researchers at University of Oxford and the Harvard Stem Cell Institute (HSCI) at Harvard University, more than 50 collaborators at over 30 scientific organizations around the globe have agreed on a common standard that will make possible the consistent description of enormous and radically different databases compiled in fields ranging from genetics to stem cell science, to environmental studies.

The new standard provides a way for scientists in widely disparate fields to co-ordinate each other's findings by allowing behind-the-scenes combination of the mountains of data produced by modern, technology driven science.

"We are now working together to provide the means to manage enormous quantities of otherwise incompatible data, ranging from the biomedical to the environmental," says Susanna-Assunta Sansone, Ph.D, Team Leader of the project at the University of Oxford's Oxford e-Research Centre.

This standard-compliant data sharing effort and the establishment of its on-line presence, the ISA Commons – www.isacommons.org, is described in a Commentary published today in the journal *Nature Genetics*. The commentary is signed by all the [collaborators](#).

"An example of how this works at the Harvard Stem Cell Institute is that

we can now find a relationship between experiments involving normal blood stem cells in fish and cancers in children", says Winston Hide, director of HSCI's new Center for Stem Cell Bioinformatics, and an associate Professor of Bioinformatics at the Harvard School of Public Health.

ISA Commons is also being used at Harvard Medical School (HMS) by the [HMS LINCS \(Library of Integrated Network-based Cellular Signatures\) project](#), led by Professors Peter Sorger and Timothy Mitchison.

It was necessary to establish common data standards, say the commentary's authors, because of the tsunami of data and technologies washing over the sciences. "There are hundreds of new technologies coming along but also many ways to describe the information produced" said Sansone, noting that "we can take a jigsaw puzzle of different sciences and now fit the many pieces together to form a complete picture".

"One of the things that I find most empowering about this effort is that now small research groups can begin to store laboratory data using this framework, complying with community standards, without their own dedicated bioinformatics support. It is a bit like Facebook allowing everyone to create their own website pages - suddenly you don't need to be an expert in computing to get your data out to the rest of the world", says Dr. Jules Griffin, of the University of Cambridge.

"What we like about it is its unifying nature across different bioscience fields and institutions", says Dr. Christoph Steinbeck, European Molecular Biology Laboratory, The European Bioinformatics Institute.

And "it also has the potential to work for large centers too", says Scott Edmunds, editor of the journal published by open-access publisher

BioMedCentral and BGI Shenzhen (previously known as the Beijing Genomics Institute) the world's largest genomics institute, "We are working with this framework to help harmonizing and presenting may large-data types as possible in a common standardized and usable form, publishing it in the associated *GigaScience* journal."

Provided by Harvard University

Citation: Oxford, Harvard scientists lead data-sharing effort: New standards allow disparate data sets to integrate (2012, January 29) retrieved 9 April 2024 from <https://phys.org/news/2012-01-oxford-harvard-scientists-data-sharing-effort.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--