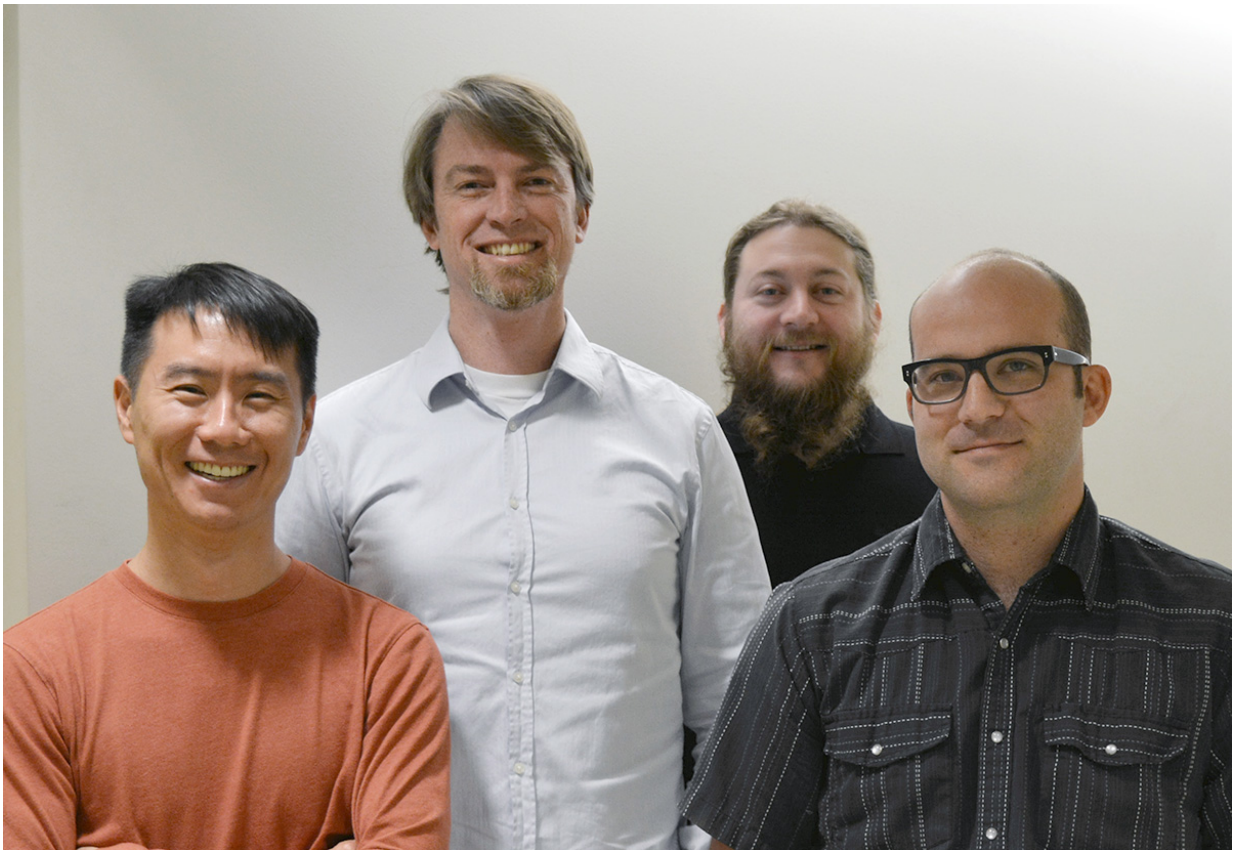


New project helps researchers build a biomedical knowledgebase

April 14 2016



The Scripps Research team working on the bioinformatics challenge includes (left to right) Andrew Su, Benjamin Good, Sebastian Burgstaller-Muehlbacher and Tim Putman. Credit: The Scripps Research Institute.

April 14, 2016 - Imagine attempting to bake a cake—except you have to

go to different stores for flour and milk, drive across town to get eggs and call a friend to borrow a cake pan.

This is the kind of disjointed scenario many scientists face when they attempt to gather [data](#) scattered across small databases and hard-to-search PDF files.

"It's not that the data doesn't exist," said Andrew Su, associate professor at The Scripps Research Institute (TSRI). "The data just isn't stored in a way that scientists can easily access."

"Open data is vital for progress and research," added TSRI Assistant Professor of Molecular and Experimental Medicine Ben Good. "We need to break down those barriers."

To solve this problem, Su, Good and their colleagues at TSRI have integrated biomedical data into Wikidata, a public, editable database where researchers can easily link genes, proteins and more. Their work was announced in two recent papers in the journal *Database*.

A Better Way to Research

Technological breakthroughs in the last 10 years have led to rapid increases in the volume and rate of biomedical research, which in turn has led to a rapid growth in biomedical knowledge. However, this knowledge is currently fragmented across countless resources—from online databases to supplementary data files to individual facts in individual papers.

"As a research community, we spend a lot of time searching for good resources and trying to link them together," said TSRI Research Associate Tim Putman, who was first author of one of the studies. "It's cringeworthy."

Even when databases are open to the public, current knowledge isn't always organized in a uniform way, Putman explained.

Rather than leave each research group to tackle data integration individually, Wikidata offers a new model for organizing all this information. Built on the same principles as Wikipedia, Wikidata enables anyone to add new information to an open community database.

While other Wikidata editors have added information on millions of items as diverse as works of art to U.S. cities, the TSRI team has focused on adding information on biomedical concepts.

TSRI Research Associate Sebastian Burgstaller-Muehlbacher, first author on one study, added data on all human and mouse genes, all human diseases and all drugs approved by the U.S. Food and Drug Administration.

Putman then extended Wikidata with a focus on microbial genomes. With all this information collected in one system, researchers can more easily spot connections between diseases, pathogens and biological processes. As an example, Putman used the model to show that other microorganisms in the body can influence chlamydia infections.

As a proof of concept, Putman led the development of a genome browser based on Wikidata. Rather than having to develop one browser for every sequenced genome, this genome browser allows users to browse any genome that has been loaded into Wikidata.

"You can zoom in on a gene, click on it and the sequence will pop up," said Good. The genome browser will then link back to the original Wikidata entry.

In the end, the researchers plan to have a comprehensive, uniform

database that is easy to search and open to anyone who wants to add data and link related concepts.

"We think this data should all be open," said Su. "This just makes intuitive sense."

More information: Tim E. Putman et al. Centralizing content and distributing labor: a community model for curating the very long tail of microbial genomes, *Database* (2016). [DOI: 10.1093/database/baw028](https://doi.org/10.1093/database/baw028)

Provided by The Scripps Research Institute

Citation: New project helps researchers build a biomedical knowledgebase (2016, April 14) retrieved 26 April 2024 from <https://phys.org/news/2016-04-biomedical-knowledgebase.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>
--