index

Speed-sifting data for next big thing

Systems to make data standardized and accessible are key for screening new material possibilities. **By Neil Savage**

Materials scientists are increasingly turning to machine learning and other computational techniques to discover new materials. From corrosion resistant aeroplane components and better batteries to new drugs or novel catalysts, big data can help to find them.

"The problem is that the number of possible materials is infinite," says Matthias Scheffler, a computational materials scientist at the Fritz-Haber Institute in Berlin, Germany. "With high-throughput screening, you can screen thousands of systems, and a thousand is nothing compared to infinite."

Along with physicist Claudia Draxl, of Humboldt University Berlin, Scheffler launched the Novel Materials Discovery Laboratory (NOMAD) at Fritz-Haber, a data repository for a wide variety of information about chemical compounds. Draxl and Scheffler are now expanding on that with a consortium called FAIRmat, which is set to receive German federal government funding of €3 million (US$3.5 million) per year for five years, to build infrastructure that standardizes data produced by many researchers so that others can use them.

The acronym comes from the principle that data should be 'findable, accessible, interoperable and reusable' by any researcher, but Scheffler also views the AIR as standing for artificial-intelligence-ready. He'd like the reams of data that chemists produce in the course of their research, much of which he says they never publish, to be collected in compatible data formats and shared so scientists can build new machine-learning models with them.

NOMAD collaborates with projects around the world, including the Materials Genome Initiative, founded in 2011 by the US National Institute of Standards and Technology (NIST) in Gaithersburg, Maryland, with similar goals.

The following year Shanghai University in China founded the Materials Genome Institute. It has launched joint projects with several other Chinese universities and held seminars drawing researchers from the United States, Japan, Singapore, France, Spain, Russia and Australia, among others. The efforts are already beginning to pay off.

For instance, Zachary Ulissi, a chemical engineer at Carnegie Mellon University in Pittsburgh, Pennsylvania, and Ted Sargent, an electrical engineer at the University of Toronto, Canada, wanted to find a way to recycle carbon dioxide into industrially useful products, rather than dump it into the atmosphere. CO_2 can be converted to ethylene using a copper catalyst to speed up the chemical reaction that removes oxygen and adds hydrogen to the molecules, but Ulissi and Sargent wanted to find a better catalyst to make the process more efficient and affordable.

MATERIALS SCIENCE AND ENGINEERING DEPARTMENT/CARNEGIE MELLON UNIVERSITY

Zachary Ulissi (right) explores how surface chirality affects chemical reactions.

Ulissi retrieved data about 244 crystals that contained copper from the Materials Research Project, part of the Materials Genome Initiative. The crystals had a total of 12,229 surfaces, and 228,969 sites on those surfaces where CO_2 molecules would stick.

He needed to run complex calculations on each one to find the most promising candidates. "Any one of those calculations takes something like two or three days, and for any given composition or crystal structure, I might have to do a hundred to a thousand atoms," he says.

To lighten the workload, he performed calculations on only a subset of the sample, then used those results to train a machine-learning algorithm. The computer suggested copper–aluminium (CuAl) alloys would be optimal, so Sargent synthesized 17 types of CuAl crystals to perform further tests and learn more about their properties. The tests showed the alloys were significantly more efficient in the ethylene reaction than pure copper catalysts.

The approach of collecting vast quantities of chemical and structural data and making it available to AI algorithms could lead to an array of new materials, from more efficient solar-cell components to more resilient smartphone screens. "Basically every new commercial product relies on a new material," Scheffler says. "So finding these materials is really the purpose AI can help with."

Neil Savage is a freelance science writer based in Lowell, Massachusetts

Correction

This Index article erroneously stated that Zachary Ulissi synthesized 17 types of CuAI crystal. In fact, the synthesis was done by his colleague Ted Sargent. Also, Sargent is an electrical engineer, not a computer scientist.