The journey from George Hurst's work to the iPhone took decades, and transformed the way we live.

ON THE RIGHT TRACK

Trace people, not papers; a new way to measure the impact of science on society.

BY NEIL SAVAGE

In 1970, George Samuel Hurst, a physicist at the University of Kentucky, came up with a device to quickly graph data by touching an electrically conductive screen, thus inventing the first pressure-sensitive touchscreen. In 1996, Wayne Westerman, pursuing a doctorate in electrical engineering at the University of Delaware, began working on an opaque touch screen that was sensitive to multiple points of contact, allowing users to input information by dragging their finger across it. Westerman co-founded a company based on that technology, which Apple bought in 2005. In 2007, Apple married a display screen, based partly on the patents it had acquired in the purchase, to a cellular phone and Internet-capable software, and the iPhone was born.

"That movement of people and ideas contributed to a fundamental change in how we do just about everything, whether dating, shopping or communication or consumption of media," says Jason Owen-Smith, a sociologist at the University of Michigan and executive director of the university's Institute for Research on Innovation and Science (IRIS).

Many people —including sociologists, economists, and funders of research — would like to clarify how scientific research leads to life-changing, practical innovations. A better understanding of this process might lead to better use of limited funds to spur useful developments. "We have a responsibility to the public, to Congress, to spend our money as wisely as possible," says Michael Lauer, deputy director of the Office of Extramural Research at the US National Institutes of Health (NIH).

But it can be difficult to measure how, for instance, a scientist's invention of a lab instrument might one day contribute to the rise of dating apps. The journey from Hurst's work to the iPhone took decades, involved dozens of researchers from different disciplines, and incorporated other technologies that also had

winding paths, such as computer chips and software algorithms. Predicting the ultimate outcome from the beginning would have taken a feat of technological clairvoyance. "There's nothing apparent in the smartphone that suggests that one might want to start by looking at the University of Kentucky, circa 1970," Owen-Smith says.

Researchers and funding agencies have long tried to track the near-term impact of science by looking at patterns of citations for journal papers or patents, in part because it's one of the most readily accessible methods for seeing how knowledge is transferred. An agency such as the NIH, for example, will look at papers resulting from research it funded, and see if they're cited by many subsequent papers or in patents, to get a measure of their impact. In the same way, patents that are cited in other patents indicate they might have impact in the private sector.

Although bibliometrics are widely used, their limitations have prompted several international groups to explore new sources of information to glean insights on the movement of ideas to inventions.

Robert Tijssen, the chair of science and innovation studies at Leiden University in the Netherlands, says while tracking patent and literature citations has some value, they are an imperfect measure of impact. That's because bibliometrics and patent citations represent only a small portion of all the knowledge created, shared, expanded and drawn upon. For instance, knowledge about laboratory techniques is typically left out of journal articles, but is shared among colleagues on research teams, and spreads when team members move to another university or take a job in industry. Likewise, some of the most consequential technical developments — such as the world wide web — have emerged without patents. Many discoveries in mathematics and economics are also never patented.

"Documents don't do science. Documents don't do innovation," says Julia Lane, an economist at New York University's Wagner School of Public Policy, who has been involved in efforts to move beyond bibliometrics. "If you're interested in the creation, transmission and adoption of knowledge, you're going to have to trace people." For a more comprehensive picture of how research interacts with industry to create new knowledge and products, funding agencies and academics like Lane are developing more robust measures of science impact.

> "IF YOU'RE INTERESTED IN THE CREATION, TRANSMISSION AND ADOPTION OF KNOWLEDGE, YOU'RE GOING TO HAVE TO TRACE PEOPLE."
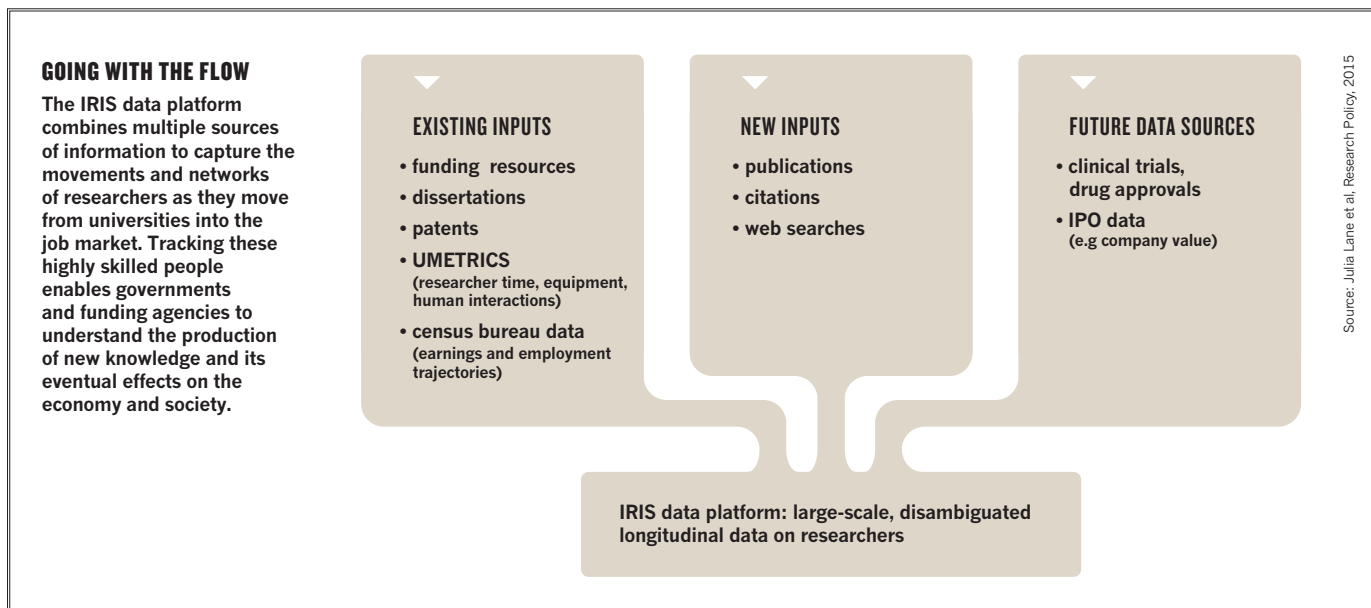
Lane, for instance, started a project that developed into the Institute for Research on Innovation and Science (IRIS), an initiative that traces people as they move through academia and industry. The IRIS project gathers basic administrative information from some 30 universities about everyone whose salary was paid at least in part by a sponsored research project. A network of collaborators on projects within and across campuses is then combined with data from the US Census Bureau's Longitudinal Employer-Household Dynamics programme, which tracks how groups of people move through the workforce, including industries and location.

All these data allows IRIS analysts to identify, on a large scale, which companies hire people formerly employed on campus research projects. They then look at the hiring firms for measures of their research productivity, such as the number of patents they file. Using the data, Owen-Smith says, IRIS — which hopes to expand to 40 more universities — is developing statistical models that describe the flow of people and ideas, from research grant to industry innovation.

As government budgets for science are squeezed, funding agencies are also focused on how to measure returns from their investments. The Federal Reporter database in the US uses data, some dating back to 2004, to track the impact of research funded by the NIH, National Science Foundation, Department of Agriculture and Department of Defense. The NIH defines impact by whether the research it funds is producing new devices, therapies, public health strategies, or new guidelines for standards of care. They also look to see whether their funds are providing material that scientists might use in future discoveries, from new reagents and cell lines to shareable genomic data.

As well as standard bibliometrics, Federal Reporter uses text mining algorithms to develop a more sophisticated understanding of a paper's contribution to its field. The database, which Lane helped develop, also looks for results that have been replicated, suggesting the original research is of high quality. "You can construct a story as to how one of these transformative discoveries came about," Lauer says. For instance, using patent and publication data, plus information gathered by the US Food and Drug Administration in the course of reviewing drug applications, the NIH has tracked how research they funded contributed

GOING WITH THE FLOW

The IRIS data platform combines multiple sources of information to capture the movements and networks of researchers as they move from universities into the job market. Tracking these highly skilled people enables governments and funding agencies to understand the production of new knowledge and its eventual effects on the economy and society.

EXISTING INPUTS
- funding resources
- dissertations
- patents
- UMETRICS (researcher time, equipment, human interactions)
- census bureau data (earnings and employment trajectories)

NEW INPUTS
- publications
- citations
- web searches

FUTURE DATA SOURCES
- clinical trials, drug approvals
- IPO data (e.g company value)

IRIS data platform: large-scale, disambiguated longitudinal data on researchers

Source: Julia Lane et al, Research Policy, 2015

BIBLIOMETRICS MEET BIG DATA

Robert Tijssen and a colleague at Leiden University's Centre for Science and Technology Studies, Jos Winnink, are searching for early indicators of research impact by combining bibliometrics with today's big data capabilities. They have taken 15 million publications and 15 million patents from the early 1990s and looked for pieces of research that had a pronounced spike in citations in patents and other publications in their first couple of years. Ten years later, about a third of those papers showed signs of impact, such as being mentioned in highly cited patents, influential review articles, or in conversations about who may be in line for a Nobel prize. It will take more data and more time, however, to see if they're on to something, says Tijssen, who is also appointed at Stellenbosch University in South Africa. ∎

to the development of PCSK9 inhibitors. These treat patients who have high cholesterol and are more effective than statins at reducing the rate of heart problems. "If you were to do this many, many times over, you might be able to tease out patterns," he says.

For a more qualitative type of information, analysts turn to case studies that talk about how university research led to a spinoff company or a new drug. Such case studies are often available on university websites, and they were required in the UK government's audit of university research, the Research Excellence Framework, in 2014. The review asked institutions to write narratives outlining the impact of their work. Analysts from King's College London pored over the almost 6,700 case studies submitted, using text mining to identify topics the authors discussed, such as 'commercialization' or "spinoff." They identified early signs of impact, such as research cited in clinical guidelines, which suggested the work might be influencing

medical practice. But they could not tell if new practices were being adopted.

Another potential source of information on how science interacts with industry is so-called alternative metrics, or altmetrics, which track research through social media. Figuring out the social impact of science is difficult, Tijssen says; using altmetrics to see what people are talking about is the best researchers can do at the moment to figure out the social impact of particular pieces of science.

As well as tracking the impact research has on products and knowledge, Tijssen says analysts could put more effort into capturing the long-term effects on the people affected by the research. That's a challenging task, however. "Impacts get more difficult to detect over time," he says. "Attributing straightforward causality becomes the domain of historians rather than economists or business analysts."

The amount and type of data tracking the link between science and its impacts is growing,

and sophisticated algorithms to sift through it are being developed. But Tijssen says the lack of theoretical underpinning for these exercises remains a problem. "We haven't managed to bring these sources of information into an impact model that allows us to draw out conclusions from each of these separate sources."

Without such a model, Tijssen says, it's difficult to know what other variables beside the research itself might be important to any eventual outcomes, and to figure out the best way to analyse the data. And with no sense of what factors are significant, gleaning whether there has been an impact becomes a complicated exercise.

Lane believes the process of combining imperfect data sources together in large volumes will tell researchers a lot about innovation. But it will still require considerable effort to figure out how best to measure impact. "Innovation is a very complex, non-linear process," she says. "We've taken the first step, but there's a lot more to be done." ∎