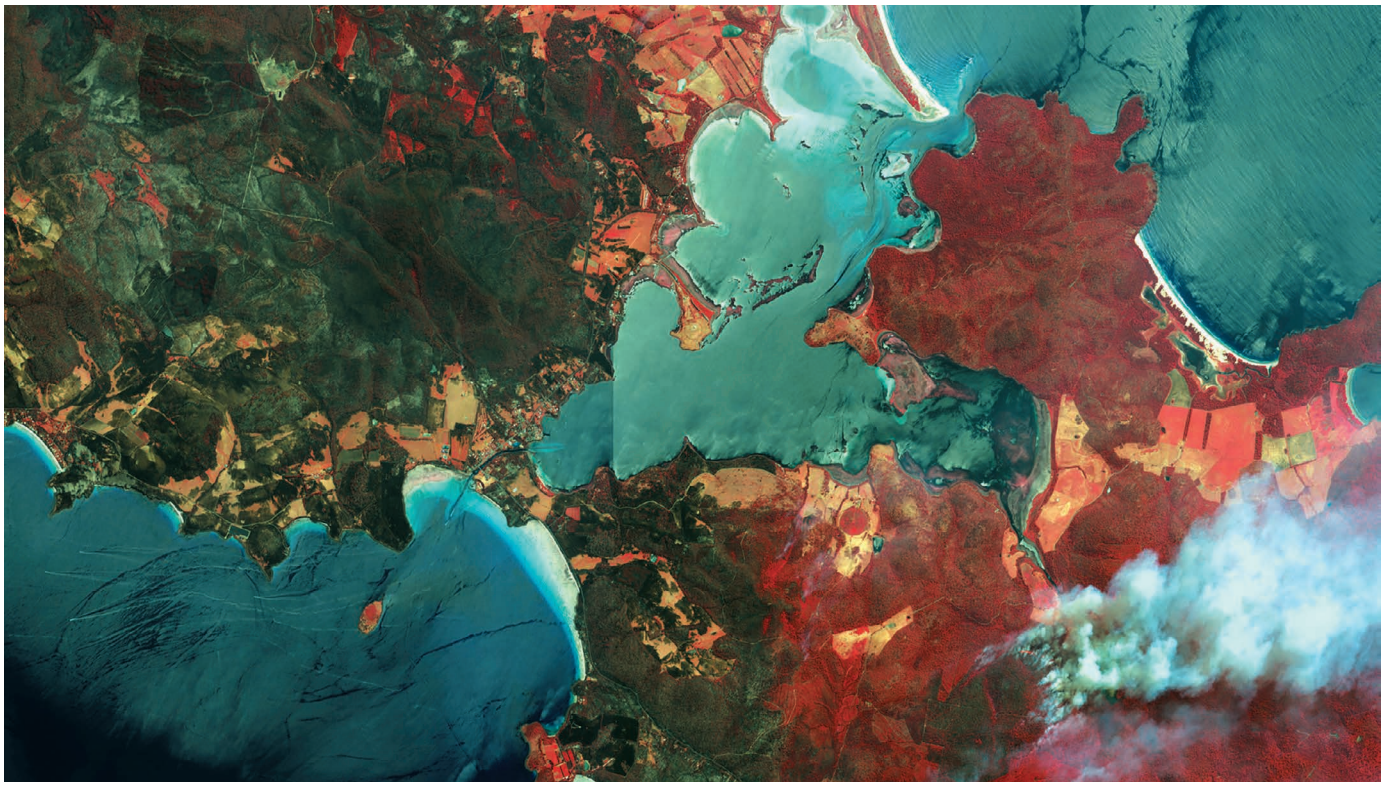
CAREERS

PUBLISHING Men expected to outnumber women as authors for next 80 years **p.747**

CONDUCT Petitioners urge US science academy to drop harassers **p.747**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

DIGITALGLOBE/GETTY



Satellite images such as this one, depicting wildfires in Tasmania, Australia, can be used to track and visualize threats to the natural environment.

IMAGERY

Crunch time for data

Satellite and cloud-computing services are opening worlds of opportunity for researchers.

BY GABRIEL POPKIN

Samapriya Roy remembers when it would take him up to an hour to download a single 1-gigabyte image taken by the Landsat Earth-imaging satellites. That was in the late 2000s, when he was analysing satellite imagery as part of his undergraduate studies at Visvesvaraya National Institute of Technology in Maharashtra state, India. And the computer analysis of a picture could take even longer. Sometimes Roy would start the analysis at night and it would still be running the next morning.

Things are very different nowadays. Roy, who is a PhD student at Indiana University in Bloomington, uses a Google platform to store his data and run his algorithms and is able to crunch tens of thousands of images in minutes; all he needs is a web browser. "It brings everyone to a level playing field," he says. In addition to data from US government sources such as Landsat, he uses sharp, detailed images from three commercial satellite companies — two of which didn't exist when he was an undergraduate — to research coastal land loss in Louisiana and the Amazon region of Brazil.

In the past few years, technology and satellite companies' offerings to scientists have increased dramatically. Thousands of researchers now use high-resolution data from commercial satellites for their work. Thousands more use cloud-computing resources provided by big Internet companies to crunch data sets that would overwhelm most university computing clusters. Researchers use the new capabilities to track and visualize forest and coral-reef loss; monitor farm crops to boost yields; and predict glacier melt and disease outbreaks. Often, they are analysing much larger areas than has ever been possible — sometimes even encompassing the entire globe. Such studies are landing in leading journals and grabbing media attention.

Commercial data and cloud computing are not panaceas for all research questions. NASA and the European Space Agency carefully calibrate the spectral quality of their imagers and test them with particular types of scientific analysis in mind, whereas the aim of many commercial satellites is to take good-quality, high-resolution pictures for governments and private customers. And no company can compete with Landsat's free, publicly available, 46-year archive of images of Earth's surface. For commercial data, scientists must often request images of specific regions taken at specific times, and agree not to publish raw data. Some ▶

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

companies reserve cloud-computing assets for researchers with aligned interests such as artificial intelligence or geospatial-data analysis. And although companies publicly make some funding and other resources available for scientists, getting access to commercial data and resources often requires personal connections. Still, by choosing the right data sources and partners, scientists can explore new approaches to research problems.

CONNIE ZHOU/GOOGLE/ZUMAPRESS/EYEVINE

Google's data centres support its cloud-based Earth Engine platform, which processes geospatial data.

MAPPING POVERTY

Joshua Blumenstock, an information scientist at the University of California, Berkeley (UCB), is always on the hunt for data he can use to map wealth and poverty, especially in countries that do not conduct regular censuses. "If you're trying to design policy or do anything to improve living conditions, you generally need data to figure out where to go, to figure out who to help, even to figure out if the things you're doing are making a difference."

In a 2015 study, he used records from mobile-phone companies to map Rwanda's wealth distribution (J. Blumenstock *et al. Science* **350**, 1073–1076; 2015). But to track wealth distribution worldwide, patching together data-sharing agreements with hundreds of these companies would have been impractical. Another potential information source — high-resolution commercial satellite imagery — could have cost him upwards of US$10,000 for data from just one country.

Blumenstock then learnt that Facebook had bought commercial satellite images for a programme it launched in 2014 to connect the global population to the Internet. After chats with a Facebook researcher on the project, he and the social-networking giant hammered out an agreement. Facebook would fund one of his graduate students to use the company's technology to study how economic data from public surveys correlated with the visual characteristics of buildings represented in the satellite data. Facebook, in turn, could potentially gain a sharper view of the socio-economic characteristics of rural areas, whose residents are least likely to have Internet connections. (Facebook declined to comment.)

The arrangement presented some challenges, however. Facebook demanded a non-disclosure agreement before sharing data. (Blumenstock does not have access to personal Facebook user data, only to satellite and other aggregated data.) And UCB industry-partnership specialists scrutinized the agreement to ensure that it wouldn't compromise academic integrity. Privacy concerns are likely to loom larger from now on. In the wake of allegations in March that a UK consultancy had deployed Facebook user data for US political purposes, universities and companies might be examining their agreements more closely.

Facebook's command of machine learning and cloud computing was also the main draw for Robert Chen, a geographer at Columbia University in New York City, who collaborates with the company to study global population distribution. Data crunching that would have once taken years was completed in weeks, enabling Chen and his colleagues to produce high-resolution population maps of rural areas in 18 countries around the world (see go.nature.com/2s1dgq4). "Facebook can process 14.5 billion images in a couple of weeks," he says. The social-media firm's main goal for the project is to provide global Internet access (and reach more potential users). Chen aims to apply the maps to humanitarian assistance, conservation and development planning.

Other hi-tech goliaths are making resources available to researchers. Microsoft's AI for Earth, which launched in late 2017, has enabled more than 60 research groups from more than 20 countries to analyse remote-sensing data sets from Esri, a mapping and geospatial-analysis company in Redlands, California, using Microsoft's artificial intelligence (AI) algorithms and computing power. Microsoft's chief environmental scientist, Lucas Joppa, says that AI can supercharge remote-sensing research by ferreting out previously hidden patterns in data. For example, a team including Milind Tambe, a computer scientist at the University of Southern California in Los Angeles, has used Microsoft algorithms to predict wildlife-poaching activity in Africa from drone imagery (see go.nature.com/2s2z5ta).

> ***"Around 20 companies worldwide now offer or plan to offer Earth-observing capabilities."***

Researchers apply online for initial access to the program. If Joppa and his colleagues find a project promising, they collaborate and share expertise and in-kind resources, such as computing time, to help the research advance.

Amazon Web Services, the cloud-computing branch of the e-commerce giant Amazon, started hosting the Landsat archive in early 2015. In September 2016, the company launched its Earth on AWS programme, through which it hosts around 15 data sets, including imagery, weather data from the US National Oceanic and Atmospheric Administration, and air-quality data from the non-profit organization OpenAQ in Washington DC. Although anyone can pay to analyse the data using Amazon's computers, scientists can apply for donations of computing time; applications must include a description of the research problem and plans for dissemination of the results.

Google now hosts more than 600 public satellite, weather, population and other Earth and environmental data sets through its Earth Engine platform. More than 70,000 users — most of them researchers — have created free accounts on the platform, says Rebecca Moore, Earth Engine's director of engineering.

The first global study done on the platform yielded a blockbuster paper on maps of forest change based on Landsat data; it has racked up nearly 3,000 citations in less than 5 years (M. C. Hansen *et al. Science* **342**, 850–853; 2013). Google's infrastructure jump-started the project in 2013 by turning what would have been 15 years' worth of data crunching on one computer into a job that took just a few days, says Matthew Hansen, a geographer at the University of Maryland in College Park who led the study.

The platform has since supported global studies of surface water, fish stocks, urban agriculture and transport networks, as well as smaller-scale studies. For Daniel Weiss, an epidemiologist at the University of Oxford, UK, who used Earth Engine to map travel time from any point on the globe to the nearest city (see go.nature.com/2ibwhbm), the platform

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

efficiently crunched a computationally expensive algorithm, saving months of work. The map itself is now a public resource on Earth Engine, and Weiss and his team are using it to produce better forecasts of malaria outbreaks.

MORE THAN PRETTY PICTURES

The growing fleet of satellite companies is serving up an increasingly diverse menu of data and images. Around 20 companies worldwide now offer or plan to offer Earth-observing capabilities. These firms, which have conventionally served military and private-sector clients in finance, agriculture and other arenas, are increasing their overtures to scientists.

In 2017, satellite company DigitalGlobe in Westminster, Colorado, provided scientists with high-resolution images worth around $6 million through its DigitalGlobe Foundation, according to the foundation's president, Kumar Navulur. For some researchers, the company's super-sharp satellite-borne cameras have enabled previously difficult or impossible studies. Sarah Parcak, for example, an archaeologist at the University of Alabama at Birmingham, has used DigitalGlobe imagery to discover hidden sites in Egypt and elsewhere, and to track looting incidents.

Satellogic, a company in Buenos Aires founded in 2010, has promised to make hyperspectral data — information-rich imagery derived from light in dozens of wavelength bands — available to any scientist who wants them. No public satellite currently collects such data, which many scientists prize for its usefulness in applications such as detecting drought stress in plants and exploring for minerals. The company says that it has shared hyperspectral data with around two dozen researchers; Roy says he got access to some data for his Louisiana research after an e-mail exchange.

The satellite company Planet, based in San Francisco, California, images the globe daily, the side of each pixel in an image representing between 3 and 5 metres on the ground. The company makes data available to scientists through its research and education programme, which offers free data for up to 10,000 square kilometres a month to scientists who apply.

Institutions can also take out subscriptions for larger data volumes. Planet has provided imagery to more than 1,600 researchers from more than 70 countries, according to Joseph Mascaro, the company's director of academic programmes. The company's frequent images enabled Andreas Kääb, a geoscientist at the University of Oslo, to track melting glaciers in near-real time in Tibet, which showed that weather and climate change caused the glaciers to suddenly collapse (A. Kääb *et al. Nature Geosci.* **11**, 114–120; 2018). In 2016, he had warned the Chinese government of an impending avalanche in Tibet on the basis of signals he had detected in Planet's images.

Kääb's research has benefited not just from the imagery itself but also from access to company staff, he says. "We typically write to Joe [Mascaro] and he connects us to someone from the team," Kääb says. "I feel to some extent I am part of the game, part of the process."

Using commercial data can have downsides. Companies such as DigitalGlobe and Satellogic typically take pictures that paying customers request, so scientists might find that no data are available for their area or time of interest. Government restrictions can also limit data availability. Mascaro and Navulur are prohibited by US law from sharing extremely high-resolution imagery of certain countries such as Israel, and cannot share data with anyone in Iran or North Korea. Blumenstock once found that Planet imagery he wanted for a project in Afghanistan was unavailable owing to an unspecified reason. Identifying individual people or vehicles is impossible, Navulur says; this alleviates some privacy concerns, although pictures can be sharp enough to make out houses and other structures. (Of course, for large areas of the world, so is Google Maps' public imagery.)

KNOW YOUR NEEDS

Use of commercial images can also be restricted. Scientists are free to share or publish most government data or data they have collected themselves. But they are typically limited to publishing only the results of studies of commercial data, and at most a limited number of illustrative images.

Many researchers are moving towards a hybrid approach, combining public and commercial data, and running analyses locally or in the cloud, depending on need. Weiss still uses his tried-and-tested ArcGIS software from Esri for studies of small regions, and jumps to Earth Engine for global analyses.

The new offerings herald a shift from an era when scientists had to spend much of their time gathering and preparing data to one in which they're thinking about how to use them. "Data isn't an issue any more," says Roy. "The next generation is going to be about what kinds of questions are we going to be able to ask?" ■

Gabriel Popkin *is a freelance writer in Mount Rainier, Maryland.*

CORRECTION

The Careers feature 'Behind the scenes' (*Nature* **556**, 525–527; 2018) erred in its description of the winning photo. Callie Veelenturf was measuring pH, conductivity and temperature after the turtle had laid her eggs, not taking samples beforehand.

PUBLISHING

Unequal authorship

An analysis of more than 10 million scientific and medical studies published between 2002 and 2018 suggests that male authors will continue to outnumber female authors for at least the rest of this century. The findings, published in *PLoS Biology*, examined articles in science, technology, engineering, mathematics and medicine (L. Holman *et al. PLoS Biol.* **16**, e2004956; 2018). Female authors were particularly scarce in the fields of physics, computer science, maths and surgery. For example, women accounted for just 13% of prestigious last-author spots in physics studies. That percentage has crept upwards by about 0.1% a year since 2002, suggesting that if current trends hold, authorship in physics studies could reach equality in roughly 260 years. The gender disparity in authorship was especially pronounced in papers from Japan, Germany and Switzerland, whereas the most gender-equitable countries were in South America, Africa and elsewhere in Europe. "Without novel interventions, these fields are likely to remain gender-biased for many decades," says lead author Luke Holman, an evolutionary biologist at the University of Melbourne in Australia. "Despite recent gains, we still have far to go."

CONDUCT

Drop harassers

An online petition is calling for the US National Academy of Sciences (NAS) to revoke the membership of scientists who commit sexual harassment or assault. The move brings fresh attention to a troubling issue. As of 23 May, the petition, created by neuroscientist BethAnn McLaughlin of the Vanderbilt Kennedy Center in Nashville, Tennessee, had gained 2,289 signatures. "I'm impressed by the response," McLaughlin says. "It speaks to the fact that time's up in academia." Comments on the petition point to other non-scientific organizations that have revoked membership for misconduct. In February, the US National Science Foundation (NSF) announced that it would require universities and institutions to disclose the identities of NSF-funded researchers who have been disciplined for harassment. The NAS takes harassment and assault "very seriously", says Jennifer Walsh, a spokesperson for the National Academies of Sciences, Engineering, and Medicine. A 22 May statement by the presidents of each of the three academies says that a dialogue has started about the standards of professional conduct for membership.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

CORRECTION

The Careers Feature 'Crunch time for data' (*Nature* **557**, 745–747; 2018) erroneously stated that an image from Planet was unavailable owing to a security concern. In fact, the reason for its unavailability was not specified. Also, DigitalGlobe is headquartered in Westminster, Colorado, not in Boulder.

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.