evgeny klochikhin **view from the top**

Researchers' skills have yet to catch up with Big Data

Big Data is a fascination of our era. Hardware and software applications are now capable of analysing huge data sets in reasonable time, and are increasingly dependent on the stable flow of information. The Big Data market is projected to almost triple in the next five years, and data firms are emerging and growing rapidly.

The list of possible applications is almost infinite. In science, data experts are working on cyber-infrastructure and integrated data platforms, finding better ways to treat disease and improving diagnostics through enhanced biomedical imaging and bioengineering. In policy, tools for collecting and analysing administrative data are increasingly important sources of evidence for improving governance and implementation.

Yet new data are not always easily available in a structured and comprehensible form. Extraction and processing are difficult and can present significant methodological challenges. Twitter feeds or websites, for example, have varied formats and provide extensive textual and visual information that has until now seldom been used in policy analysis or academic research.

This difficulty in gathering and acquiring information is one of four main challenges in the use of Big Data. Another is organisation: when the data are extracted, how can they be arranged in an accessible format and prepared for further analysis? This leads on to the third difficulty: analysis. What do you do with unconventional data types such as raw text? Finally, communication is a challenge. What is the best way to deliver results in a clear visual format that can be grasped by decision-makers and public officials who lack the time to go through huge databases and tables of unlinked information?

To tackle these problems, we need data scientists who can develop and apply statistical and technical approaches stretching far beyond the traditional methodologies used in political science, sociology and economics. For example, data gathering and acquisition requires skills such as web scraping and handling application programming interfaces. An understanding of database architecture, integration techniques and matching is imperative for order and analysis.

This is about much more than just computing and statistical skills. Data science should be a genuinely interdisciplinary field, and its specialists have to understand the root of, and the solutions to, the problems. Few possess the right combination of skills. A report commissioned by the statistical-software company SAS revealed that 57 per cent of UK firms find it hard or very hard to recruit Big Data talent.

The UK strategy for data capability, published in October, recognises that data-related analytics are driven by three core skills: "data management (data storage and linking); data analysis (including numeracy, statistics and coding knowledge); and business and policy insight (subject specialism, context awareness and entrepreneurial spirit)".

Improved training and better organisation could help promote all three skills. The challenges of Big Data call for innovations in curricula and teacher development. Modelling and programming are vital tools but not enough on their own. We need problem-oriented, or integrated, curricula. The latest research suggests that teachers are more effective as facilitators of learning than as direct deliverers of knowledge. We can prepare students for the world of Big Data by making them aware of the complexities, and focusing on multidisciplinary and interactive learning outcomes.

However, an overemphasis on multidisciplinarity in education can be dangerous. It is impossible to embed all skills in one individual. So good organisation is an essential aspect of data capability. A flat structure—in which computer scientists, business and policy analysts and subject specialists all have an equal say in identifying the major objectives and directions of development—is, perhaps, the most appropriate form of organisation for Big Data tasks.

Today, IT personnel are often perceived as a supporting unit, with various computer tasks distributed erratically among different divisions and groups. Raising computer literacy among all employees and engaging computer specialists in a wide variety of content-related and client-facing tasks is an important step toward promoting a more integrated and productive approach to IT services. In science, such a policy could involve investing in multidisciplinary labs where, instead of temporary staff supporting individual (monodisciplinary) projects and often duplicating the work of their colleagues in other departments, researchers from various backgrounds would come together and share the workload.

Something to add? Email comment@ResearchResearch.com

'We need approaches beyond those common in political science, sociology and economics.'

Evgeny Klochikhin is a researcher at the American Institutes for Research in Washington DC.