

Visualizing big science projects

Katy Börner¹, Filipi Nascimento Silva and Staša Milojević

Abstract | The number, size and complexity of ‘big science’ projects are growing — as are the size, complexity and value of the data sets and software services they produce. In this context, big data gives a new way to analyse, understand, manage and communicate the inner workings of collaborations that often involve thousands of experts, thousands of scholarly publications, hundreds of new instruments and petabytes of data. We compare the evolving geospatial and topical impact of big science projects in physics, astronomy and biomedical sciences. A total of 13,893 publications and 1,139 grants by 21,945 authors cited more than 333,722 times are analysed and visualized to help characterize the distinct phases of big science projects, document increasing internationalization and densification of collaboration networks, and reveal the increase in interdisciplinary impact over time. All data sets and visual analytics workflows are freely available on GitHub in support of future big science studies.

‘Big science’ today is international, interdisciplinary and inter-institutional. Big science projects are anchored around expensive, large and complex instruments, they can run for several decades and they involve thousands of experts. Big science projects make breakthroughs not only in basic research but also in innovation that impacts economy and solves challenging societal needs. As more science fields move towards the big science model of knowledge creation, the lessons learned from previous successful endeavours become essential. This is because big science projects are not just larger and more expensive than other projects but they require specific organizational and management structures. Different knowledge production processes also bring new research roles, changes in the division of labour and adjustment in formal and informal scholarly communication. One way to communicate these aspects of big science, on which this Perspective focuses, is to use various visualizations. Visualizations in this Perspective — and interactive online ones — show that big science projects go through phases with different input needs, expected outputs and impacts. As big science projects mature, their collaborations densify and internationalize; at the same time, scholarly impact increases in terms of citation counts and interdisciplinary reach.

Big science as a phenomenon can be traced all the way back to fifteenth-century cartography and astronomy^{1–3} or to eighteenth-century natural history expeditions^{2,4}. Nineteenth-century extensive archival projects (the *Corpus Inscriptionum Latinarum* and the *Carte du Ciel*) had many characteristics of present-day big science in terms of funding (state backing by Prussia and France), workforce and timescale (requiring more than a lifetime of effort), and were associated with the initial coinage of the term ‘big science’ (or, originally, *Gorswissenschaft*) by classical philologist and Prussian Academy of Sciences member Theodor Mommsen⁵. The better known and more immediate precursors of what became known as big science are the establishment of the University of California cyclotron by Ernest Lawrence in the 1930s for energy research⁶ and the World War II Manhattan Project⁷. The term ‘big science’, however, was introduced in the 1960s by Alvin M. Weinberg^{8,9} and Derek J. De Solla Price¹ to describe post-World War II developments in physics that built large and very expensive instruments (reactors and accelerators), accompanied by the growth in scientific team sizes working on nuclear-related research⁷. Making advances in nuclear and, later, particle physics became part of the competition among superpowers, with the expectation that breakthroughs would

lead to both scientific and technological superiority^{10,11}. In addition, big science has been propelled into the general public’s awareness by the founding of the National Aeronautics and Space Administration (NASA) and its active and publicly visible space programme². Although most of the early focus regarding big science was on physics, as early as 1965, Weinberg¹² proposed that biomedical science and biomedical technology were ready to enter the ‘big biology’ era. This entry was made only in the 1990s with the Human Genome Project (HGP), the first big science project in biology¹³. The expansion of the big science mode of knowledge production to other areas of science, such as big biology, brought with it new organizational and collaborative forms, such as ‘networked’ science enabled by information and communication technologies¹⁴ and some debates as to whether such coordinated efforts can be called big science^{15,16}.

Big science accentuated the central role instruments play in the development of science as “engines of discovery”¹⁷. Historically, instruments such as the telescope, the microscope and the air pump opened new vistas and led to scientific revolution, fundamentally changing the nature of scholarship^{18–21}. The quest for increased sensitivity and accuracy of instruments led to their constant evolution, making these ever more expensive tools^{19,22} obsolete fairly quickly¹⁹. This process has been described²³ as ‘tinkering’, in which ‘lineages of technology’ are adapted and combined, leading to networks, or ‘genealogies’ of technologies. However, the power of instruments, such as a scanning tunnelling microscope, can be realized only when they engage a community of researchers in what has been called ‘an instrumental community’, eventually leading to the formation of new scientific fields, such as nanotechnology²⁴. Furthermore, the relationship between science and technology is complex and interdependent, with science also contributing to technology development^{25–27}.

Early scientists, such as Galileo Galilei and Isaac Newton, engaged in instrument building as well as theoretical and experimental work^{28,29}. While not without precedent, instrument building