

A New Structure of Life Sciences Organizations Driven by Big Data

Jingyuan Xing

School of Philosophy, Beijing Normal University, Beijing, China

Email: xingjingyuan@mail.bnu.edu.cn

How to cite this paper: Xing, J. Y. (2024). A New Structure of Life Sciences Organizations Driven by Big Data. *Open Journal of Social Sciences*, 12, 138-147. <https://doi.org/10.4236/jss.2024.124010>

Received: February 16, 2024

Accepted: April 16, 2024

Published: April 19, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Driven by high technologies such as big data, a revolution has taken place in the field of life sciences and a paradigm shift. The paradigm of science determines the organizational model of science. In the past, most of the studies on the organization model of science took physics as the research object, and there were few studies on the organization model of life science. In order to explore the changes in the organizational structure of life sciences, this paper sorts out the revolutionary changes of life sciences with the participation of data methods, and tries to summarize the new characteristics of the organizational structure of life sciences from the aspects of organizational personnel, organizational information, organizational technology, organizational structure, organizational culture, etc. The results show that the organizational structure of life sciences shows three characteristics: platform, ecological and sharing. First of all, based on the development of the Internet, from the construction of national databases, various biobanks, and sequencing platforms, to the development of biomedical programs and software on mobile phones, life sciences have begun to develop towards platform-based science. Second, the platformization and popularization of scientific research have expanded the participation of scientific research from scientists to industry, citizens, community organizations, etc. The diversity of scientific research participants and the convergence of exchanges between them make life science organizations appear as ecological systems that flow within the system and have an impact on the whole. Thirdly, the basic tone of open sharing of biological data information after the genome revolution is based on the platform-based, ecological, and open and inclusive scientific organization and cultural atmosphere, and the life sciences are developing towards shared science.

Keywords

Life Sciences, Organizational Structure, Big Data, Platformization

1. Introduction

At present, driven by high and new technologies such as big data, artificial intelligence, and cloud computing, science in many fields has undergone a revolution and paradigm shift, and science has entered the “fourth paradigm” driven by data. The paradigm of science determines the organizational model of science. In previous studies, most of the studies on the organization model of science took physics as the main research object, and the “department system” was commensurate with the organization model of modern science. However, the era of the typical dominance of physics is over, and the way physics is studied is not representative of many other natural sciences, such as the life sciences. In fact, the development of the life sciences (especially biomedicine) at the end of the 19th century has shown a completely different mode of research from physics. At present, most research on the life sciences reveals revolutions, but there are no studies that explore the occurrence of revolutions from the perspective of organizational structure.

In order to fill the gap in this part of the theoretical research, this study attempts to analyze the changes and characteristics of the organizational structure of life sciences driven by big data from the perspective of organizational model. From some cutting-edge research of the life science revolution, it can be known that recently, in the participation of big data methods, the process of life science revolution research can be briefly summarized as: data collection, data processing, data analysis and data storage. In this process, life data is based on a variety of platforms, and the participants in scientific research are not limited to professional scientists. (A lot of data processing requires forces outside of academia.) Also, in the whole research process, not only are the researchers “heterogeneous”, but the scientific knowledge applied is across multiple sciences, involving organizations and departments, and also across the scope of academia.

From the perspective of organizational structure, this study will examine the changes in life sciences driven by big data from the aspects of organizational personnel, organizational information, organizational technology, organizational structure, organizational culture, etc., and condense the characteristics of changes in the organizational structure of life sciences, including development trends.

2. Platformization of Life Sciences Organizations

2.1. Databases and Various Big Platforms

The platform model was originally a business operation model. The platform party does not directly provide products or services, and makes the buying and selling double issuance for trading through the platform provided, the standards and trading rules of the platform. Typical Internet trading platforms such as JD.com, Taobao, etc. The platform-based development of life sciences is due to the maturity of the Internet and high-throughput automation technology, and scientific research is driven by data and generates data. The overall development of life sciences is increasingly towards large-scale data analysis. The establish-

ment of more databases, computing environments and virtual work networks has reduced the information barrier of life sciences and greatly improved the efficiency of scientific research and production. From the construction of national biological sample databases and data sequencing platforms, to various mobile phone application service platforms, in the era of big science, various platforms have made the organization of life sciences more diverse.

The United States, the European Union, and Japan were among the first countries to include the protection of genetic resources in their national strategic plans. The earliest databases in the world include the National Center for Biotechnology Information (NCBI) of the United States, the European Bioinformatics Institute (EBI) and the DNA Database of Japan (DDBJ), which are the three most influential national databases. They started in the 80s and 90s of the last century, and are relatively authoritative institutions in the storage, exchange and acquisition of national bioinformatics data.

Recently, the construction of the National Gene Bank in Shenzhen, China, was approved in 2011, adopting a pioneering operation model that includes a resource sample bank and bioinformatics. In terms of overall architecture, the Shenzhen National Gene Bank is structured with “three banks and two platforms”, including the “three databases” of biological sample resource bank, bioinformatics database and animal and plant resources living bank, and the “two platforms” of digital platform and synthesis and editing platform. In recent years, China has also built databases at different levels, such as bioinformatics, genomes, bioinformatics tools, knowledge, and databases based on data produced by some biotechnology companies.

The internationally renowned and diverse biobanks have also promoted the development of life science platforms. For example, the International Society for Biological and Environmental Repositories (ISBER), a branch of the American Society for Research Pathology, founded in 1999, has established the Animal Sample Bank, Environmental Sample Bank, Human Sample Bank, Microbial Sample Bank, Museum Sample Bank, and Plant/Seed Sample Bank. Equally influential was the National Cancer Institute (NCI), which was founded in 2005. NIH is also committed to seeing the establishment of various technology platforms as a priority in the life sciences development roadmap. For example, such as chemical small molecule screening center, imaging probes synthesis platform, etc (Wu, 2004).

In addition to the well-known sequencing platforms, there are also synthetic biology platforms built with the rapid development of synthetic biology in recent years. Several major sequencing platforms have strong capabilities in producing genetic data, such as the Wellcome Sanger Institute (<https://www.sanger.ac.uk/>) established in the United Kingdom in 1992, the Broad Institute (<https://www.broadinstitute.org/>) in the United States, and the “French Genome Medicine 2025 Project” 12 gene sequencing platform centers built in France, as well as the crowd sequencing technology service company that is developing in China. The synthetic biology platform takes DNA synthesis as the core support-

ing technology, and there are currently three prominent platforms in the world: Illinois Biological Foundry for Advanced Biomanufacturing (iBioFAB, <https://experts.illinois.edu/>) and The Edinburgh Genome Platform (The Edinburgh). Genome Foundry, EGF, <https://www.genomefoundry.org/>), Imperial College London DNA Synthesis Platform (Wang et al., 2019).

2.2. Small Platform for Information Services

Various mobile phone applet software and Internet platforms in professional fields are all manifestations of the platformization of life sciences in the micro field. During the new crown pneumonia period, some small medical informatization and digital platforms have played a great role in the treatment of the epidemic. For example, WeDoctor's "COVID-19 Current Affairs Relief Platform" has gathered more than 66,000 doctors across the country, serving 2.119 million people and nearly 150 million visits (Li, 2022). In the post-epidemic era, people's health concept has improved and the demand for healthy life has objectively accelerated the arrival of a new round of biotechnology revolution, industrial revolution and bioeconomy, and objectively accelerated the platform-based development of life sciences.

3. Ecology of Life Sciences

From an ecological point of view, energy flows between different levels and communities. A biome becomes an organic whole, and every part of the whole and its interactions have an impact on the whole. Based on the development of the above-mentioned life platform, multiple entities can share resources and information, and the construction of various platforms allows organizations and individuals with different identities to relate and achieve a win-win situation. Life sciences and biomedicine have formed new partnerships between academic, clinical, and industrial communities, and the emergence of new frontier leaders, citizen data scientists, and some community organizations. The diverse forms of life science research show the characteristics of ecology.

3.1. Diverse Participants in Scientific Research

- First of all, companies are the new players in scientific research.

BGI was originally established to represent China in the Human Genome Project. Today, BGI is a leading life sciences organization in China and the world's largest genomics R&D institution. Whether it is in scientific research, industry, or the transformation of achievements, BGI plays an important role in the development of life sciences. Different from the traditional scientific research led by universities and research institutes, some powerful private enterprises represented by BGI have shown outstanding scientific research achievements and demonstration effects, reflecting the important position of enterprises in scientific research in today's era of big science.

Industry has always been a new part of the main body of scientific innovation,

but in the past, China's life science research has not been as influential as today, and private enterprises have had such a great impact on science. Social welfare research and commercial competitiveness research have been divided through market allocation. Enterprise alliances, industrial common technology alliances and industrial clusters are developing rapidly.

BGI's founding team completed the "China Part" of the Human Genome Project (about 1% of the human genome), which is the sequencing of about 30 million base pairs on the severed arm of human chromosome 3. This work has enabled the only developing country participating in the International Human Genome Project, and has also enabled BGI to master the world's most advanced genomics technology for the first time. In terms of scientific research, BGI has published more than 200 papers in top international scientific research journals such as "CNNS" (Cell, Nature, New England Journal of Medicine, Science). At the same time, BGI also founded the journal "GigaScience", which publishes datasets and software articles in the field of life sciences and medicine, which is open and has a certain international influence.

In the field of application, BGI's non-invasive pre-genetic testing technology has now been widely used in medical diagnosis and treatment. BGI is already one of the most prolific non-invasive prenatal screening facilities in the world, and its scale is still expanding. Wang Jun, then executive chairman of BGI, once said, "What BGI is trying to do is to develop something useful, from scientific discovery to end market" (Cyranoski, 2012). BGI has developed a massively parallel sequencing system that can screen fetal DNA from the mother's plasma and diagnose chromosomal abnormalities.

- Second, citizen scientists have become the new participants in scientific research.

Now, scientific research crowdsourcing has become a new scientific organization model in the era of big science, also known as "citizen science". It originated as an organization in the business sector, and then crossed borders and "citizen science" to develop a model of scientific research crowdsourcing. Its essence is to outsource research tasks to non-specific members of the public through open recruitment.

For example, the Open Science Competition is a type of scientific research crowdsourcing. In the past two years, artificial intelligence AlphaFold and AlphaFold2 have been launched, solving the problem of protein spatial folding structure that has plagued structural biologists for nearly half a century. In fact, the solution to this problem lies in the wide participation of scientific research participants, that is, through the channel of the Global Critical Assessment of Protein Structure Prediction (CASP). Citizen scientists, industry professionals, and hobby enthusiasts have been given the opportunity to participate in scientific discovery. Through the competition, DeepMind's R&D team also learned about this problem, which was originally a small group of biologists in the academic world, and tried to solve it.

Of course, there are many other competitions of the same kind, such as the

CAMEO (Continuous Automated Model Evaluation) alongside CASP, which is known as the two most authoritative competitions in the field. CASP is a biennial competition, while CAMEO is held continuously, and almost every week, structural biologists are available for participants to compete in the latest protein problems, and the rankings are updated weekly. The RoseTTAFold software system designed by the team of Professor David Baker of the University of Washington's Protein Design Institute, inspired by AlphaFold, was entered in the CAMEO competition, and Baidu, Tencent, and Huawei all participated in the competition for algorithms. Professor Lan Yanyan from Tsinghua University's Institute of Intelligent Industry was also one of the many contestants, and she and her team's AIRFold have won the first place for four consecutive weeks.

The form of scientific outsourcing represented by the Science Open Competition provides a platform for many non-academic people who want to participate in scientific research, so that capable individuals, teams, companies, enterprises, communities, etc. can participate, reflecting the ecology of life science research organizations.

- Third, community science has become a new organization of scientific discovery.

The community of community scientific organizations has begun to form in the last century and has become a new trend in the development of scientific research. According to a study, well-organized community science networks are an important factor in promoting the development of science research in the life sciences. After the Human Genome Project, the National Institutes of Health (NIH) named a group of classic model organisms for biomedical research to facilitate scientific research. Designation as a model organism by the National Institutes of Health does not guarantee an increase in publication trends. For example, in yeast research, in 1993, the Cold Spring Harbor (CHS) Yeast Course was introduced, which established the Cold Spring Harbor genetics tradition and subsequently led to the Yeast Molecular Biology Conference. Since then, the number of attendees has doubled each year. Such courses and conferences create a sense of community, which in turn promotes yeast as a model organism (Dietrich et al., 2014). More and more such organizations are being formed, forming small-scale scientific communities with community resources, databases, networks, etc., and are supporting the life sciences as ecological complexes.

3.2. Cluster Convergence of Industries

Driven by big data, the life sciences sector has formed a reputation sub-institution that sits outside of the university's academic institutions. From the intersection and integration of disciplines to the joint action of multiple disciplines, the boundaries of scientific organization are more blurred. This is manifested in the fact that the emergence of a scientific problem usually requires knowledge, skills and methods that involve many fields, that is, "convergence" of life science research. Convergence is a multifaceted integration, i.e., people, disciplines, de-

partments, institutions. From a societal perspective, research extends beyond the scope of the original academic community to stakeholders such as industry and government ministries. For example, in November 2013, the U.S. federal government, together with state governments, businesses, universities and other relevant departments, launched a collaborative action on the use of big data to make knowledge discoveries. In the life sciences, the National Institutes of Health (NIH), IBM, Sutter Health and the G. Kissinger Health System of Pennsylvania have invested \$2 million to fund the development and application of new technologies and methods of data analytics to help medical care detect heart failure earlier, as well as a new project by the nonprofit Medic Mobile to quantify health and measure health to improve the health of millions of people (White House, 2013).

The research in the industry weakens the disciplinary boundaries and tends to transform the research results into applications. Form a collaborative network for scientific research transformation. For example, the Institute for Quantitative Biosciences (QB3) at the University of California is the result of the creation of the community. QB3 is funded by university discovery biology experts and funded by investors, while it trains researchers to run small startups. In addition, he has the support of the government and the private sector, start-ups also give back to the government to drive the economy, and the private sector provides technology and labor. This is a clustering effect of the life sciences industry (Research Council of the National Academy of Sciences, 2014). Similarly, taking Shenzhen BGI as an example, in addition to BGI, there are a number of new scientific research institutions in Shenzhen, such as Shenzhen Kuang-Chi Institute of Advanced Technology, Shenzhen Tsinghua University Research Institute, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, and Shenzhen National Innovation Energy Research Institute. Among them, a new type of organizational operation mode was born, such as the operation mode of “private government assistance” practiced by BGI and Guangqi Advanced Technology Research Institute (Lin, 2016). BGI also cooperated with South China University of Technology to launch the “Genome Science Innovation Class”, which implements the innovative training program of industry-university-research cooperative education. Students in the “Innovation Class” take courses at both the school and the BGI company, and their credits are mutually recognized. In the year since 2010, students have even published papers in top international scientific journals, including review articles in famous Chinese newspapers and the journal Nature (Mao & Sun, 2010).

The interaction and cooperation between new scientific research institutions in the industry, as well as with the government and universities, breaks through the traditional scientific research model in a single institution, and makes the cluster and convergence structure of life science organizations integrated across departments. This kind of research ecosystem is mostly found in academic centers, such as the vicinity of universities. In such collaborative networks, partner-

ships are closer and more conducive to communication.

4. Sharing of Life Science Research Methods

Natural genes are not patentable. Since the beginning of the Human Genome Project, the basic principle of “common, co-doing, and sharing” has been established, and the reference sequence of the human genome is stored in the form of FASTA on the websites of UCSC and NCBI, which is open to the world. This also lays the basic tone for the sharing of biological data resources driven by big data. There are many influencing factors in the formation of life science platforms, such as the collection of data in research design, the mode of collaborative co-authorship of article results, the open and shared digital platform system, and the culture of cross-knowledge convergence. In turn, platform-based life sciences also promote the openness and sharing of life sciences.

Life sciences and biomedicine usually rely on the quantitative analysis of experimental data, and massive data surpasses ordinary computer science scales in mathematical tools and methods for modeling and analysis, requiring high-throughput sequencing. These will further facilitate the standardization and generalization of biomedical data. In addition, the connection between the Internet and the external network, and the effective communication of the electronic team, are all convenient conditions for data sharing.

Peer-to-peer sharing and communication, such as publishing articles in journals or attending conferences and forums, is a form of scientific sharing. One of the evaluation indicators of scientific research and innovation is actual ability and achievement, including invited reports at important national and international conferences, publication in influential journals, and positions in important academic organizations. In the past, the content of authors and articles was demarcated by disciplines, and there were significant differences between disciplines, but in the integration of biotechnology and information technology, the boundaries of life sciences, biotechnology or information science disciplines are blurred, and many authors are compound talents with multidisciplinary knowledge or the published results themselves are co-authored. According to the relevant research summary, from the perspective of the citation frequency of the paper and the journal impact factor, the influence of co-authored papers is significantly higher than that of non-co-authored papers (Wang & Ding, 2018). This is a representation of scientific research knowledge sharing in the development of life sciences with the participation of big data, and it is not necessarily a necessary method for discipline integration and communication.

In the tradition of open databases for life science access resources, after the participation of big data, it is more about the storage and sharing of massive biological data information. For example, PubMed, Genbank and Protein Data Bank are open access to and access to various databases for scientists around the world, making scientific research more convenient and especially conducive to the progress of clinical medicine. Of course, data privacy and data security make it necessary to further clarify the ethical norms for data sharing.

Finally, the support mechanism of inclusive culture also drives the sharing of life sciences. Whether it is to cope with the cultural atmosphere of health and humanities, or the rigorous teamwork in the laboratory, a good interpersonal network of equality and respect and smooth communication is the great ecological environment for the development of scientific research in a teamwork atmosphere. Encourage young scholars to make bold assumptions and conjectures, provide more platforms for young people, encourage innovation, and tolerate failure. Free and innovative ideas come from an open and inclusive cultural environment.

5. Conclusion

Driven by high technologies such as big data, the field of life sciences has undergone a revolution, entered the paradigm of big data, and the organizational model of science has also undergone revolutionary changes. From a structural point of view, the organizational structure of life sciences shows three characteristics: platform, ecology and sharing. First of all, thanks to the development of the Internet and high-throughput technology, the construction of national databases, diverse biobanks, and sequencing platforms, as well as the development of biomedical programs and software on mobile phones, are all manifestations of the platform model. As a result, life sciences are beginning to develop towards platform-based science. Second, the platformization of science has promoted the popularization of scientific research. The participants in scientific research have expanded from scientists to stakeholders such as industry, citizens, and community organizations. In particular, the participation of industry is very active, and a variety of new scientific research institutions have produced a cluster convergence effect. The exchange and cooperation between various scientific research participants and scientific research institutions make life science present an ecological scientific research collaboration network environment. Third, the genome revolution has established a basic tone of open sharing of biological data information. Based on the platform-based, ecological, open and inclusive scientific organization culture, life sciences are developing towards shared science.

Acknowledgements

The author expresses sincere thanks to anonymous reviewers.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Cyranoski, D. (2012). Chinese Genomics Giant BGI Plots Commercial Path. *Nature Biotechnology*, 30, 1159-1160. <https://doi.org/10.1038/nbt1212-1159>
- Dietrich, M. R., Ankeny, R. A., & Chen, P. M. (2014). Publication Trends in Model Or-

-
- ganism Research. *Genetics*, 198, 792-793. <https://doi.org/10.1534/genetics.114.169714>
- Li, B. (2022). *Bioeconomy* (p. 340). China Democracy and Legal Publishing House.
- Lin, X. (2016). Private Government Assistance: An Effective Organizational form of the Government's Innovation-Driven Strategy. *Scientific Research*, 34, 386-394.
- Mao, D. W., & Sun, X. (2010). The Model Reform Has Shown Initial Results, and Talent Training Has Gradually Become a Feature—"Science and Nature", Students of the Genome Science Innovation Class of Huagong—BGI, Have Frequently Appeared in "Science" and "Nature", Which Have Attracted Attention. *Guangdong Science and Technology*, 19, 16-18.
- Research Council of the National Academy of Sciences (2014). *Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond* (X. L. Wang, Y. Xiong, & J. R. Yu, Trans., pp. 26-27). Science Press.
- Wang, B., Liu, F., Zhang, E. C. et al. (2019). National Gene Bank: Co-Ownership, Co-Ownership and Sharing. *Heredity*, 41, 761-772.
- Wang, M. X., & Ding, J. D. (2018). A Review of the Author-Author Cooperation Model of Scientific Research Papers. *Research Intelligence*, 38, 172-177.
- White House (2013). *Fact Sheet: Data to Knowledge to Action: New Announcements*. <https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/Data2Action%20Announcements>
- Wu, J. R. (2004). A Roadmap to the Future of Life Sciences. *Science*, 56, 24-26+2.