# DEVELOPING A REVENUE MODEL FOR BIOINFORMATICS R&D IN INDIA USING MANAGEMENT APPROACH

**SHASHANK SHEKHAR**

**SCHOOL OF INTERDISCIPLINARY RESEARCH (SIRE)**

**INDIAN INSTITUTE OF TECHNOLOGY DELHI**

**JULY 2024**

# DEVELOPING A REVENUE MODEL FOR BIOINFORMATICS R&D IN INDIA USING MANAGEMENT APPROACH

by

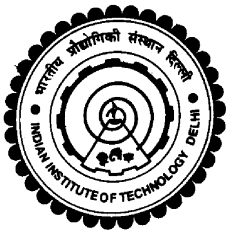

## SHASHANK SHEKHAR

**School of Interdisciplinary Research (SIRe)**


submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the





**Indian Institute of Technology Delhi**


**July 2024**

# *Certificate*

This is to certify that the thesis entitled "Developing a revenue model for bioinformatics R&D in India using management approach," being submitted by **Mr. Shashank Shekhar** to the Indian Institute of Technology, Delhi, for the award of the degree of **Doctor of Philosophy** in **School of Interdisciplinary Research (SIRe)** is a record of bonafide research work carried out by him. Shashank Shekhar has worked under our guidance and supervision and has fulfilled the requirements for submitting this thesis, which has reached the requisite standard to our knowledge.

The results of this dissertation have not been submitted in part or full to any other University or Institute for the award of any degree or diploma.


Prof. Sanjay Dhir
Department of Management Studies
Science Indian Institute of Technology Delhi
India

Prof. James Gomes
Kusuma School of Biological
Indian Institute of Technology Delhi
India

Prof. B. Jayaram
Department of Chemistry
Indian Institute of Technology Delhi
India


Dated:

# *Acknowledgments*

*My Ph.D. research has been an incredible journey in which I have encountered and experienced a range of emotions at various points in time. I wholeheartedly thank my supervisors, Prof. Sanjay Dhir, Prof. James Gomes, and Prof. B. Jayaram, for guiding, inspiring, motivating, and supporting me all through. It has been an honor to work under their mentorship. I am grateful to them for their invaluable advice, unwavering support, and encouragement throughout this exciting passage of research, which enabled me to advance as a researcher. The luxury of receiving ideas and inputs from experts from three different areas has been a unique value proposition for me and my research. I am deeply indebted to my supervisors for showing confidence in my abilities and helping me complete the research part of my Ph.D., followed by thesis writing as a part of the formalities for the award of my Ph.D. degree.*

*I am grateful to students' research committee (SRC) members Prof. Nalin Pant, Prof. Vivekanandan Perumal, and Prof. Arpan Kar, apart from other faculty and staff members of the Department of Chemistry, School of Biological Sciences, Department of Management Studies, and School of Interdisciplinary Research (SIRe), Indian Institute of Technology Delhi, for their help and support at different stages of my Ph.D. program. I thank the Department of Biotechnology, Department of Electronics & Information Technology (DeitY), Govt. of India, Centre for Development of Advanced Computing (CDAC), Pune, for their financial support during this research.*

*I thank all my past and current lab members of SCFBio (my workplace) and the School of Interdisciplinary Research (where I have enrolled in this Ph.D. program). I thank my fellow lab mates, especially Akshata, Smriti, Dinesh, and Devendra, for their technical and scientific inputs. Also, A. Mohan Rao and Puneeta were always there to provide administrative support whenever required.*

*I am also thankful to various groups and individuals, such as the respondents to my survey, the interviewees from academia and industry who shared their views in developing insights on my research topics, and program officers from various funding agencies in India who helped me in understand the financial support mechanism for bioinformatics research and all those who helped and contributed directly or indirectly during the commencement, continuation, and completion of beautiful journey of my Ph.D. and because of whom I have come to this stage where I am representing my work through this thesis.*

*I thank all my family members, especially my younger brothers (Saumitra and Saharsh), my uncle (Chacha), and my in-laws (especially my brother-in-law, Parijat, who kept asking me that when I am going to submit my thesis) for their love, affection, blessings, and unconditional support throughout the expedition of my doctoral research. I am immensely thankful to my wife, Vandana, for her love, affection, compassion, support, and sacrifices. My two lovely daughters, Shreya and Swara (Strawberry), have always been my motivation and source of inspiration as they more often create a learning environment at home, which pushed me to focus on my research alongside my routine office work, which has many facets. Also, being the senior most member of the research center, I have had the responsibilities and administrative duties required for the smooth functioning of the research center, which is among the best in the country.*

*My parents are in the heavenly abode, and I miss them a lot, but their blessings have been the driving force for me during my Ph.D. journey. Last but not least, I thank the almighty for everything.*

**Shashank Shekhar**

# Abstract

In recent years, research in biological sciences has witnessed a significant shift in the research approach while moving from in vitro and in vivo experimentation to adopting in silico methodologies, which fall under the bioinformatics domain. This interdisciplinary area works on ways to understand and analyze biological data, especially when the data sets are extensive and complex in terms of size and characteristics. This prominent inclusion of *in silico* experimentation depends mainly on the development and efficacies of various research methods, computational tools, software, and databases to deal with scientific problems related to various organisms.

The rapid growth in bioinformatics research resulted in advancing scientific methods and developing relevant technologies, culminating in products and services beneficial to society. Considering the significance and relevance of various efforts in the life science area, the bioinformatics and computational biology based approaches attract researchers and professionals belonging to different subject areas, such as computer science, information technology, chemistry, biological sciences, mathematics, and several other related areas to work together through a common platform to create an innovation ecosystem of greater value eventually and to address prevailing and potential agricultural and healthcare issues affecting various stakeholders.

This thesis examines the scenarios and factors that enable academia, where most basic research and development occurs, to embed commercialization aspects to make academic research self-sustainable in the long run. This exploration gives academia new opportunities and enables them to think beyond the traditional approach of developing and disseminating novel knowledge with stakeholders. Most scholars and decision-makers

agree that entrepreneurship is the future and vital to society. In this thesis, bioinformatics research management is examined, and a growth and revenue model is proposed.

The thesis is divided into nine chapters. Chapter 1 discusses the current status of bioinformatics research in the country and how it strengthens the life sciences and healthcare-based research ecosystem with the advancement in information technology-driven applications. Apart from that, management perspectives are introduced and discussed to formalize and streamline the research pipeline. This is followed by adding up the self-sustainability aspects by exploring various revenue generation streams and realizing the vision of having a full-fledged revenue model within the academic boundaries.

Chapter 2 focuses on the systematic literature review related to the thesis topic, considering both the research and management aspects of bioinformatics activities in India and the best practices adopted at the global level. This part of the thesis investigates the empirical studies published in the literature on synchronizing research and management aspects. Based on the findings, an attempt is made to develop a sustainable model in bioinformatics research and development in India.

Chapter 3 has a detailed description of measurement scales, various factors, sub-factors included in this study, and the methodology adopted for analyzing the data collected either through literature review or from the respondents of the questionnaire survey and interviews to conduct the study. This is to help develop a revenue model in sync with the objectives of this thesis.

Chapter 4 discusses how bioinformatics researchers from academia and entrepreneurs in the life science and healthcare industries join hands to develop a conducive environment to complement each other's work. The adoption of management-based methodologies and approaches further bolsters this effort.

Chapter 5 illustrates and discusses the role of developing a knowledge-based ecosystem for bioinformatics R&D by strengthening human resource development programs that help create a pool of trained bioinformatics professionals in the country who can drive and manage this highly specialized sector linked with public health and agricultural productivity.

Chapter 6 highlights the role and significance of a High-Performance Computing (HPC) technology-based revenue model for bioinformatics R&D in academia and how the various factors and their interdependencies play their part in creating a technological ecosystem that paves the way for developing a framework from within the academic boundaries with the help of a management-based approach.

Chapter 7 addresses developing and managing bioinformatics technologies, including user-friendly and interactive software/tools and databases inside academic boundaries where most of these research activities are conducted and which are the genesis of most of the successful commercial ventures when it comes to bioinformatics software suites offered by commercial companies.

Chapter 8 dealt with triangulation and synthesis aspects, which are important research principles that provide authenticity and significance to the academic work. It culminates in multiple viewpoints and data sources to provide comprehensive knowledge. This chapter also includes the implications of the research work undertaken in this thesis.

Chapter 9 discusses the summary, a few perspectives emerging from the thesis work, limitations, and future scope.

# सार

हाल के वर्षों में, जैविक विज्ञान में अनुसंधान ने इन विट्रो और विवो प्रयोग से लेकर सिलिको पद्धति को अपनाने के प्रतिमान में बदलाव देखा है, जो जैव सूचना विज्ञान डोमेन के अंतर्गत आता है। यह अंतःविषय क्षेत्र जैविक डेटा को समझने और उसका विश्लेषण करने के तरीकों पर काम करता है, खासकर जब डेटा सेट व्यापक और जटिल होते हैं। यह प्रमुख समावेशन मुख्य रूप से विभिन्न जीवों से संबंधित वैज्ञानिक समस्याओं से निपटने के लिए विभिन्न कम्प्यूटेशनल उपकरणों और सॉफ्टवेयर के विकास और प्रभावकारिता पर निर्भर करता है।

जैव सूचना विज्ञान अनुसंधान में तेजी से वृद्धि के परिणामस्वरूप वैज्ञानिक तरीकों को आगे बढ़ाने और प्रासंगिक प्रौद्योगिकियों को विकसित करने में मदद मिली, जिसका परिणाम समाज के लिए लाभकारी उत्पाद और सेवाएं हैं। जीवन विज्ञान क्षेत्र में विभिन्न प्रयासों के महत्व और प्रासंगिकता को ध्यान में रखते हुए, जैव सूचना विज्ञान और कम्प्यूटेशनल जीव विज्ञान आधारित दृष्टिकोण विभिन्न विषय क्षेत्रों, जैसे कंप्यूटर विज्ञान और सूचना प्रौद्योगिकी, रसायन विज्ञान, जैविक विज्ञान, गणित और अन्य संबंधित क्षेत्रों से संबंधित शोधकर्ताओं और पेशेवरों को आकर्षित करते हैं। प्रचलित स्वास्थ्य देखभाल मुद्दों से जुड़ी जैविक अनुसंधान समस्याओं के समाधान के लिए अंततः अधिक मूल्य का एक नवाचार पारिस्थितिकी तंत्र बनाने के लिए एक आम मंच के माध्यम से हाथ मिलाना और एक साथ काम करना।

यह थीसिस उन परिदृश्यों और कारकों की जांच करने और समझने की कोशिश करती है जो शिक्षा जगत को, जहां सबसे बुनियादी अनुसंधान और विकास गतिविधियां होती हैं, आगे सोचने और इन गतिविधियों के साथ व्यावसायीकरण पहलुओं को शामिल करने और लंबे समय में खुद को आत्मनिर्भर बनाने में सक्षम बनाती हैं। यह परिदृश्य शिक्षा जगत के लिए कई अप्रयुक्त अवसरों को खोलता है और उन्हें हितधारकों के लिए नए ज्ञान को विकसित करने और प्रसारित करने के पारंपरिक दृष्टिकोण से परे सोचने में सक्षम बनाता है। अधिकांश शिक्षाविद और निर्णय-निर्माता इस बात से सहमत हैं कि उद्यमिता एक नया रास्ता है और बड़े पैमाने पर समाज की उन्नति और कल्याण के लिए आवश्यक है। यह थीसिस जैव सूचना विज्ञान में अनुसंधान के प्रबंधन पहलू का पता लगाने और उस दिशा में एक व्यवहार्य विकास और राजस्व आधारित मॉडल विकसित करने का प्रयास करती है।

थीसिस नौ अध्यायों में विभाजित है। अध्याय 1 देश में जैव सूचना विज्ञान अनुसंधान की वर्तमान स्थिति पर चर्चा करता है, और यह सूचना प्रौद्योगिकी-संचालित अनुप्रयोगों में

प्रगति के साथ जीवन विज्ञान और स्वास्थ्य देखभाल-आधारित अनुसंधान पारिस्थितिकी तंत्र को कैसे मजबूत कर रहा है। इसके अलावा, अनुसंधान पाइपलाइन को औपचारिक बनाने और सुव्यवस्थित करने के लिए प्रबंधन दृष्टिकोण पेश किए जाते हैं और उन पर चर्चा की जाती है। इसके बाद विभिन्न राजस्व सृजन धाराओं की खोज करके आत्मनिर्भरता पहलुओं को जोड़ा जाता है और अकादमिक सीमाओं के भीतर से ही एक पूर्ण राजस्व मॉडल रखने की दृष्टि को साकार किया जाता है।

अध्याय 2 भारत में जैव सूचना विज्ञान गतिविधियों के अनुसंधान और प्रबंधन दोनों पहलुओं और वैश्विक स्तर पर अपनाई गई सर्वोत्तम प्रथाओं पर विचार करते हुए, थीसिस विषय से संबंधित व्यवस्थित साहित्य समीक्षा पर केंद्रित है। थीसिस का यह हिस्सा मूल रूप से अनुसंधान और प्रबंधन पहलुओं के सिंक्रनाइज़ेशन पर साहित्य में प्रकाशित अनुभवजन्य अध्ययनों की जांच करता है, और निष्कर्षों के आधार पर, भारत में जैव सूचना विज्ञान अनुसंधान और विकास में एक स्थायी मॉडल विकसित करने का प्रयास किया जा रहा है।

अध्याय 3 में इस अध्ययन में शामिल माप के पैमाने, विभिन्न कारकों, उप-कारकों और साहित्य समीक्षा के माध्यम से या प्रश्नावली सर्वेक्षण और साक्षात्कार के उत्तरदाताओं से अध्ययन के संचालन और अंततः एकत्र किए गए डेटा का विश्लेषण करने के लिए अपनाई गई पद्धति का विस्तृत विवरण है। इस थीसिस के उद्देश्यों के अनुरूप मॉडल विकसित करने में मदद करता है।

अध्याय 4 में चर्चा की गई है कि कैसे जैव सूचना विज्ञान शोधकर्ता शिक्षा जगत से संबंधित हैं और जीवन विज्ञान और स्वास्थ्य सेवा उद्योगों के उद्यमी एक-दूसरे के काम के पूरक के लिए अनुकूल वातावरण विकसित करने के लिए हाथ मिलाते हैं। प्रबंधन-आधारित पद्धतियों और दृष्टिकोणों को अपनाने से इस प्रयास को और बल मिलता है।

अध्याय 5 मानव संसाधन विकास कार्यक्रमों को मजबूत करके जैव सूचना विज्ञान अनुसंधान एवं विकास के लिए एक ज्ञान-आधारित पारिस्थितिकी तंत्र विकसित करने की भूमिका को दर्शाता है और चर्चा करता है जो देश में प्रशिक्षित जैव सूचना विज्ञान पेशेवरों का एक पूल बनाने में मदद करता है जो सार्वजनिक स्वास्थ्य से जुड़े इस अत्यधिक विशिष्ट क्षेत्र को चला और प्रबंधित कर सकते हैं।

अध्याय 6 शिक्षा जगत में जैव सूचना विज्ञान अनुसंधान एवं विकास के लिए उच्च-प्रदर्शन कंप्यूटिंग (एचपीसी) प्रौद्योगिकी उन्मुख राजस्व मॉडल की भूमिका और महत्व पर प्रकाश डालता है और कैसे विभिन्न कारक और उनकी अन्योन्याश्रयताएं एक तकनीकी पारिस्थितिकी तंत्र बनाने में अपनी भूमिका निभाती हैं जो एक रूपरेखा विकसित करने का

मार्ग प्रशस्त करती है। प्रबंधन-आधारित दृष्टिकोण की सहायता से शैक्षणिक सीमाओं के भीतर से।

अध्याय 7 शैक्षणिक सीमाओं से उपयोगकर्ता-अनुकूल और इंटरैक्टिव सॉफ़्टवेयर/टूल्स और डेटाबेस सहित जैव सूचना विज्ञान प्रौद्योगिकियों के विकास और प्रबंधन पर केंद्रित है, जहां इनमें से अधिकतर शोध गतिविधियां की जा रही हैं और जो मूल रूप से अधिकांश सफल वाणिज्यिक उद्यमों की उत्पत्ति हैं। वाणिज्यिक कंपनियों द्वारा पेश किए गए जैव सूचना विज्ञान सॉफ्टवेयर सुइट्स में आता है।

अध्याय 8 त्रिकोणासन और संश्लेषण पहलुओं से संबंधित है, जो महत्वपूर्ण शोध सिद्धांत हैं जो अकादमिक कार्यों को प्रामाणिकता और महत्व प्रदान करते हैं। यह व्यापक ज्ञान प्रदान करने के लिए कई दृष्टिकोण और डेटा स्रोतों का समापन करता है। इस अध्याय में इस थीसिस में किए गए शोध कार्य के निहितार्थ भी शामिल हैं।

अध्याय 9 सारांश, थीसिस कार्य से उभरे कुछ दृष्टिकोण, कुछ सीमाएँ और भविष्य के दायरे पर चर्चा करता है।

# Table of Contents

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF ABBREVIATIONS

| Abbreviations | Definitions |
|---|---|
| **BTISNet** | Biotechnology Information System Network |
| **DBT** | Department of Biotechnology |
| **CDAC** | Centre for Development of Advanced Computing |
| **DeitY** | Department of Electronics and Information Technology |
| **DST** | Department of Science and Technology |
| **E7 Countries** | A group of emerging market economies from Brazil, China, India, Indonesia, Mexico, Russia and Turkey. |
| **EU** | European Union |
| **G7 Countries** | A group of advanced economies of Canada, France, Germany, Italy, Japan, the UK, and the US |
| **HPC** | High Performance Computing |
| **IIT** | Indian Institute of Technology |
| **IISc.** | Indian Institute of Science, Bangalore |
| **IoT** | Internet of Things |
| **MD/VP** | Managing Director/ Vice President |
| **MICMAC** | Matrice d'impacts croisés multiplication appliquée á un classment |
| **M-TISM** | Modified Total Interpretive Structural Modeling |
| **PI** | Principal Investigator |
| **R&D** | Research and Development |
| **R&T** | Research and Technology |
| **SAP-LAP** | Situation, Actor, Process–Learning, Action, Performance |
| **SCFBio** | Supercomputing Facility for Bioinformatics & Computational Biology, IIT Delhi |
| **TRL** | Technology Readiness Level |