In this article we estimate the economic benefit of open source drug discovery (OSDD) programme for its decade-long effort of developing newer and novel therapeutics for TB by streamlining the upstream research and development (R&D) of the drug discovery pipeline. The OSDD programme, presently subsumed into the India TB Research and Development Consortium (ITRDC) project led by ICMR, has opened up a novel method of asset utilization and capacity building at a negligible cost, leading to positive economic impact. The OSDD programme has also become an instrumental part of shaping the National Intellectual Property Rights (NIPR) Policy, 2016. This calls for a policy directive as programmes such as OSDD can be a powerful strategy in dealing with the excruciating disease burden of TB and other neglected diseases that countries such as India are facing.

Keywords: Drug discovery, economic benefit, ITRDC, India NIPR, open source drug discovery, tuberculosis.
**Open source: from open source software to neglected diseases**

Open source is a term first used in the software sector, to describe a software for which the source code is publicly available and freely redistributable. The open source initiative describes open source as ‘a development method for software that harnesses the power of distributed peer review and transparency of process’; it details an open source definition that includes access to source code, the right to redistribute without charge, permission to create derived works, and no discrimination against users or fields of application. The four pillars of an open source approach are verification, collaboration, cost reduction, and the creation of a commons. As the source code is available to the community, it can be verified for errors and undesired features. Collaboration take place effortlessly across organizational barriers and attract contributors with differing monetary and non-monetary motivations. Over time, knowledge resource and capability can be created, since the software is open for others to use, learn from, build on, and adapt for local interpretation.

The concept of propriety software and customized operating systems gained massive momentum as they dominated a large part of the last century. The major supercomputing giants such as Cray-Supercomputers Inc., IBM, Hewlett-Packard (HP), SGI, etc. developed tailor-made operating systems for each of their supercomputers thus being a revenue generator. However, in the last two decades the rapid transition of computational architecture has also witnessed the adoption of free and open software’s and operating systems such as LINUX. Hence the open innovation movement can further be traced back to the concept of LINUX as open source software.

The development of LINUX as an alternative operating system is rooted in the monopolization of proprietary...
software’s and operating systems, such as that of Unix and Microsoft, thus launching the need for creating a network of virtual community volunteers. As of 2013 Linux accounted for 93.5% of the operating systems used on a supercomputer.

The open source movement penetrated any form of informatics sector including the Bio-Informatics sector. Notably the Bio-Linux operating system launched exclusively for biological science researchers is bundled with an array of over 250 open source softwares such as R and Bioconductor, Bio-Perl, Bio-Java, Bio-Python, etc. The impact of this can be measured as USD 5.6 billion generated from sales of supercomputers in 2015 alone indicating that the indirect revenue generation by open source software is nearly USD 5 billion18.

On the contrary, the pharmaceutical sector is a classic example of how a closed-door approach and maintaining strict intellectual property rights (IPRs) is causing the sector to be laggard and resistive to disruptive innovation. When translating open source ideas to global health R&D initiatives, there are quite a few similar features. Software and pharmaceutical R&D are both knowledge-driven and generating fields with a highly and involved global community. Like software development, pharmaceutical R&D has a large virtual element; this facilitates an internet-enabled collaboration, which is at the core of most open source applications. Both software and pharmaceutical R&D have shown a rapid pace of innovation driven by basic R&D and both fields have a diverse set of stakeholders19,20.

However, there are very significant differences, such as laboratory utilities and clinical trials are much more expensive than the capital and equipment required for software generation. Safety and regulatory components play a larger role in drug R&D and increase time, risk and cost. The R&D time frame and risk is larger for drug development as opposed to a software development project21. The driving force for the success of IT industry’s growth was that it was led by young minds, who were responsible for the disruptive innovations. The revolution of social media is a testament to the power of youth. This involvement is however lacking in the pharmaceutical sector. The role of youth remains restricted to lower bounds of scientific hierarchy. This non-inclusiveness is one of the key factors for lack of innovation in the pharmaceutical sector. Although, through outsourcing several components of drug discovery by the pharmaceutical sector goes to academia, no comprehensive development of new chemical entities (NCEs) has occurred.

In order to circumvent this problem for neglected diseases, a disruptive approach was needed, specially for upstream science (R&D) lead identification. Also, the large scale failure of leads attrition is due to the absence of a systems-level approach, despite the large amount of omics data availability. There crowd sourcing of young minds without the worry of IPR restrictions has become the need of the hour. An open source approach for neglected-disease R&D must comprise the following three functionalities22. (i) Open access (to data), (ii) open collaboration (across organizational and geographical boundaries) and (iii) open rules (that enable various forms of openness).

**Open source drug discovery – Linux for Pharma**

The article aims to evaluate the economic benefit of the OSDD programme, over a ten-year period, since its launch in 2007. Envisioned in 2006, to prioritize the need for R&D in infectious diseases, the Government of India launched the OSDD programme, which leveraged the power of crowdsourcing coupled with the creation of a

---

**Figure 3.** Drug R&D model costs to yield 1 new molecular entity (NME).
virtual laboratory space. Using a constructivist approach, OSDD provided a uniform platform which allowed for an intellectual, analytical and technological reinvigoration for drug discovery. The OSDD pipeline provided a new lens of study for business model innovation whilst inculcating the concepts of the national innovation systems (NIS). The founding principle of NIS is that for innovation to happen knowledge must flow freely between all actors and linkages. OSDD facilitated the creation of an open network between national laboratories, government agencies, research enterprises, scientists and students globally. OSDD aimed at developing a platform based on systems biology which allowed sharing of risks, rewards and resources. OSDD engaged a community of students, scientists, clinicians, academicians and institutions, which collaborated through an online portal and offline laboratory work with open-notebook. The public financing of R&D costs, allowed OSDD to work with generic manufacturers thereby mitigating treatments at close-to-marginal cost, and creating higher affordability of drugs. OSDD shared resources through an online collaborative platform that operates through a ‘clickwrap license’ in which participants agree not to use knowledge generated from the online intellectual commons for proprietary benefits. OSDD also maintained publicly accessible databases, including an integrative genomics map of Mycobacterium tuberculosis.

The OSDD initiative also shared financial and non-financial rewards, both at the individual level and collective levels. Individual rewards ranged from activities which encourage women scientists with small prizes in the form of credit for phone usage and internet access. The OSDD’s collaborative platform tracked the consistent and significant contributors of the online community and rewarded those individuals in different ways, such as: authorship and acknowledgement in eventual publications. The best performers in OSDD community have leveraged their participation into competitive applications for international fellowships and trainings in various prestigious programmes.

OSDD showcased the business concept of crowdsourcing and asset optimization with minimal overhead expenditure. Assets in three independent entities were available, namely physical, instruments and latent human resource. OSDD positioned these entities to create a robust and sustainable business model. The OSDD model is based on collaboration and knowledge sharing to solve complex scientific challenges. The first disease target of OSDD was tuberculosis; the economic burden of which has been discussed in the previous section.

**Physical asset**

The OSDD programme synced together national laboratories, universities, colleges and research institutions, creating a virtual institution in a resource limiting environment. OSDD brought together 10 CSIR national laboratories, 14 industry partners and over 50 universities and colleges from the country and several individuals and laboratories outside India, thus creating a virtual laboratory through these federated spaces. The computational and laboratory facilities of these institutions were repurposed to host and train scientists and students.

**Instrument asset**

OSDD utilized the instrumental infrastructure of existing physical assets reducing capital investments on instruments. The existing infrastructure was made to accommodate work packages from the pipeline to increase the utilization time of high-end facility. C-DAC Garuda provided their supercomputing facility to college and university students who had no access to such computing power. Garuda grid provides an unprecedented e-infrastructure for OSDD applications. OSDD members access the Garuda grid via the Garuda portal or galaxy workflow and launch their jobs on the grid nodes. The job executes in one of the several compute nodes of the grid and users can then download and share their results with their community members. Some of the major highlights of the infrastructure utilization are: (i) State-of-the-art high-performance computing (HPC) clusters provided to run drug discovery problems; (ii) Secure access to high-end resources for scientists and students even from remote locations; (iii) OSDD virtual organization created in Garuda; (iv) Galaxy workflow is provided for genomics and proteomics applications; (v) Parallelizing specific applications; (vi) Supporting bio-tools required by the community.

**Latent human resource**

OSDD’s human resource could be divided into three verticals, namely administrative, scientific and student. The administrative core managed and coordinated the functioning of OSDD. The backbone of OSDD was the mobilization of large student workforce who engaged in a virtual intellectual setting with scientists from across the world. Over 9000 participants from nearly 130 countries are registered on OSDD, most are from India followed closely by United States of America and United Kingdom. Through various academic activities the students were trained rigorously to achieve scientific competence and produce high quality scientific output which would be further analysed by the scientific cadre. The OSDD pipeline also created a niche for a Public–Private Partnership (PPP). Contract research organization (CROs) was made an integral part of the system to utilize their process strength with intellectual inputs of academic institutions. The Sir Dorabji Tata Trust awarded a grant to enable...
OSDD for the TATA CSIR-OSDD Fellowship (TCOF) to support students and young researchers involved in crowdsourcing. The Trust pledged USD 500,000 to enable the youth’s contribution in drug discovery for neglected diseases such as TB and malaria. The women scientist workforce, particularly those talented women who are (may or may-not-be) homemakers and are interested in working on the neglected diseases problem, were involved as a part of the crowdsourcing research exercise.

**Economic value analysis of OSDD**

With the open platform of Sysborg⁹, OSDD opened the scientific activity’s innovation and interaction base, thus tapping into India’s USD 45 billion worth of higher education sphere and nearly 200,000 direct scientific research manpower. Asset utilization and optimization were established using a government investment of USD 12 million²². The entire OSD framework created a functioning ecosystem that could also leverage the innovative power of youth while actively engaging social and digital media. The framework also showcases how open source can be an effective model for drug discovery. The major features are as follows:

OSDD volunteers undertook the mammoth task of re-annotating the genome of *Mycobacterium tuberculosis*. Though the genome had been annotated over a decade ago, yet most genes remained unknown, complicating the process of discovering new drugs for TB. The OSDD task of re-annotating the whole genome packed nearly 300 man-years into four months at zero additional cost⁴⁰. This was done exclusively through crowd-sourcing and open source tools. A benchmark effort on similar lines would be that of the ENCODE project, the pilot phase of which cost an estimated USD 55 million, with the third phase attracting nearly USD 150 million. Therefore, the value generation in terms of scientific manpower output approximates nearly USD 60 million.

Participants from the crowd sourcing activity of OSDD phase-I were surveyed in 2016. Of the 146 active participants 92 responded to their current education or occupational status at the time of their joining OSDD. Thirty four participants were Masters level students, 6 were Bachelors level students, 1 was a Bachelor of Medicine Bachelor of Surgery (MBBS) student and 1 was a Diploma student. We aimed to see the value addition of these participants, through OSDD, in terms of educational and occupational opportunities. Thirty six participants had enrolled in a Ph D programme (24 in India and 12 outside India). Twenty participants leveraged their skills developed with OSDD to be associated with industries, 20 went on to hold scientific positions in various institutions in the country. Seven participants who were registered as Ph Ds at the time of being associated with OSDD, all, held post-doctoral positions outside India. The rest are pursuing Masters level education. The human capital value created by OSDD generated an economic value of USD 2.5 million through Ph D students and USD 400,000 through Masters students alone (note 2). The actual impact may be more as only a fraction of the student’s value addition has been computed.

**Scientific outcomes of OSDD**

Utilizing the information generated about the enhanced functional annotations of the *Mtb* genome, OSDD developed a novel systems biology spindle map of metabolism (SBSM) and comprehended its static and dynamic structure using various computational approaches based on simulation and design. The study proposed the novel concept of metabolic persistor genes (MPGs) as a potential combination of drug targets for existing antibiotics⁵². Further, systems level mapping of the metabolic complexity in *Mtb* revealed possibilities of targeting bacterial NDH-I with existing FDA approved drug for type-II diabetes, Metformin, along with existing front-line antibiotics as a potential combination therapy for TB³¹. Based on a data-intensive genome level analysis and the principle of conservation of the evolutionarily important genes, OSDD further proposed a novel systems biology based virtual drug discovery model for the prediction of non-toxic metabolic targets in *Mtb*, thus suggesting a more effective starting point for generating new chemical leads³⁴.

OSDD outcomes include open biological and chemical repositories, namely

- **OSDDChem**–OSDDChem is the open access chemical repository that houses various compounds synthesized for screening against TB and Malaria. The open portal provides a platform for students and institutions to synthesize and submit well-characterized, pure potential anti-TB molecules or compounds. There are nearly 2000 freely accessible virtual and synthesized molecules submitted to OSDDChem portal by researchers from about 40 participating institutes³⁷.

- **Computational Resources for Drug Discovery (CRDD)** (note 3) – OSDD’s *in silico* module CRDD is a one-stop platform that integrates computational tools used in drug discovery.

- **Open access cloning repository** – OSDD established a centralized-high quality clone repository of *M. tuberculosis*. Using the drug targets from various experimental and *in silico* methods, nearly 200 targets were selected for cloning. OSDD today maintains a repository of 116 sequence confirmed *Mtb* clones³⁶.

- Naturally occurring plant based anticancerous compound-activity-target (NPACT) database³⁷ – NPACT is a curated database of plant-derived natural compounds that exhibit anti-cancerous activity, containing
1574 entries. Each entry provides information on their structure, properties (physical, elemental and topological), cancer type, cell lines, inhibitory values, molecular targets, commercial suppliers and drug likeness of compounds. NPACT concentrates on anti-cancer natural compounds found in plants only. NPACT is unique in providing bioactivities of these natural compounds against different cancer cell lines and their molecular target.

- SnCRe – structural interactome computational resource for Mycobacterium tuberculosis (Mtb)\textsuperscript{38} – the database provides an integrated platform to allow easy access and interpretation of data and results obtained by all groups in the Cambridge-Bangalore (CamBan) collaboration under OSDD. The academic collaborators of CamBan develop their own algorithms and databases which are used to create Mtb targeted datasets. These datasets are hosted on the database thereby providing the structural perspective to the studies on tuberculosis.

These are a few of the large number of open source databases and open access repositories provided on the OSDD platform. The reach and access of these databases are in competition with proprietary software’s and databases. The open access (domain analysis) estimate of the resources developed by OSDD–CRDD alone is nearly USD 2 million annually (note 4).

**Impact of OSDD**

In a separate study conducted by the present authors, we have shown how an open source approach to technology development leads to a product capable of creating a market disruption whilst impacting the lives of those at the bottom of economic pyramid\textsuperscript{49}. The Soleckshaw development had a huge impact on the health of rickshaw pullers. The rickshaw pullers using the Soleckshaw were generating a monthly income of USD 330 as opposed to USD 170 on a traditional rickshaw, this improved income marginally uplifted them from poverty and allowed them access to better nutrition improving their immunity and lowering the risk of contracting TB infection.

As of 2017, OSDD has completely developed the \textit{in silico} systems biology platform for non-toxic target identification of \textit{Mtb}. Also, FDA approved drugs have been repurposed and six drugs have been identified as metabolic targets\textsuperscript{34}. Currently the \textit{in silico} lead identification is underway, and several new international initiatives have begun. Overall through the development of human resource and computational resources the OSDD programme has shown an economic benefit of over USD 10 million.

The vision of OSDD is aligned with the sustainable development goal 3, Target 3.3 which is the eradication of TB by 2030.

In the past decade, the most remarkable policy directives to be introduced in India are

- The Open Access Policy 2015.
- India TB Research and Development Consortium (note 5) (ITRDC) under ‘Stop TB’.

The point of interest in NIPR is that it states \textit{India will continue to utilize the legislative space and flexibilities available in international treaties and the TRIPS Agreement. These flexibilities include the sovereign right of countries to use provisions such as Section 3(d) and Compulsory Licensing for ensuring the availability of essential and life-saving drugs at affordable prices. Section 2.10 of the NIPR 2016 states:}

‘Encourage R&D including open source based research such as open source drug discovery (OSDD) by the Council of Scientific and Industrial Research (CSIR) for new inventions for prevention, diagnosis and treatment of diseases, especially those that are life threatening and those that have high incidence in India’.

The informal adoption of open access policy aims to bring about reliability and transparency of public funds and their utilization in public funded research.

The ITRDC is a programme setup by the Government of India and spearheaded by the Indian Council of Medical Research (ICMR), which brings the various stakeholders, i.e. government bodies, private sector, NGOs, etc. under the aegis of open collaboration to meet the national and global goal of TB eradication\textsuperscript{40}.

As on 5 August 2016 Bill Gates, a visionary who built his empire on proprietary software, vouched that open innovation was the way forward for drug discovery. The coming decade will be the age of massive transformations, as the digital divide that shrinks the pharmaceutical industry will witness the need to embrace the open source ideology towards providing health for all. The major stakeholders of health particularly the big pharma would play a pivotal role in felicitating this. The various open source movements, such as open source software and open education (MOOCs), have showcased that billion-dollar industries can be built and sustained through the open source model thereby creating dynamic optimal business ventures, and the same would apply to the pharmaceutical industry.

**Conflict of interest:** The authors declare no conflict of interest. All authors have read and agreed on the content of the manuscript.

**Notes**

2. The CSIR system estimates for human resources through differential value calculations for PhDs and Masters students were USD
GENERAL ARTICLES

70,000 (USD 2016 current) per student over a period of ten years, and USD 41,000 (USD 2016 current) for Masters students over a period of four years. The calculations consider an 8% discounting over the respective years.

3. www.crdd.osdd.net

4. The overall open impact was calculated using an in-house calculation of OI = 85 * total hits. CRDD receives a daily hit of 5456 (https://sitereview.co/crdd.osdd.net.html) thereby generating USD 2 million.


22. Masum, H. and Harris, R., Open Source for Neglected Diseases: Magic bullet or Mirage, Results for Development Institute, 2011.


35. OSDD, Open Source Drug Discovery; www.osdd.net


ACKNOWLEDGEMENTS. The authors are grateful to the Department of Science and Technology (DST), India for the J. C. Bose National Fellowship awarded to S.K.B. and the Council of Scientific and Industrial Research (CSIR), India for providing financial support to N.C. (Grant No. P-81-101).

Received 5 December 2017; revised accepted 5 February 2018.

doi: 10.18520/cs/v115i10/0458-1864