Economic Impacts of the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank

Rutgers Office of Research Analytics

Kevin P. Sullivan
Peggy Brennan-Tonetta, Ph.D.
Lucas J. Marxen

May 2017

Economic Impacts of the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank

A Report Prepared for the
RCSB Protein Data Bank

by

Kevin P. Sullivan[1]
Peggy Brennan-Tonetta, Ph.D.[2]
Lucas J. Marxen[3]

Office of Research Analytics
Rutgers New Jersey Agricultural Experiment Station
88 Lipman Drive - Martin Hall
New Brunswick, NJ 08901

May 2017

[1] Assistant Director, Statistical Analysis, Office of Research Analytics, NJAES

[2] Director, Office of Research Analytics, NJAES; Associate VP for Economic Development, ORED; Associate Director, NJAES

[3] Associate Director, Research Technologies, Office of Research Analytics, NJAES

ACKNOWLEDGEMENTS

The authors wish to acknowledge the individuals at the RCSB-PDB who contributed information, guidance and data to this study - Dr. Stephen Burley, Director; Christine Zardecki, Deputy Director; and Dr. Helen Berman, Founding Director.

EXECUTIVE SUMMARY

Introduction

The Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB-PDB), the US data center for the Worldwide Protein Data Bank (wwPDB) archive, manages and provides large-scale data of experimental 3D structures of biological macromolecules. RCSB-PDB provides wwPDB data and other unique services freely as an open data resource that is utilized by researchers (from structural biologists to computational biologists), educators, and students all over the world. RCSB-PDB operates at Rutgers, The State University of New Jersey and the University of California San Diego/San Diego Supercomputer Center.

Methodology

The results of an analysis examining the value and economic impact of the RCSB-PDB and its work are presented below. Sources of data for this analysis include website analytics (RCSB-PDB user numbers and levels of use) and RCSB-PDB operational expenditures. The authors of this analysis also applied several key factors and conclusions drawn from a 2016 study of the European Molecular Biology Laboratory- European Bioinformatics Institute (EMBL-EBI), the site of PDB Europe.

Key Findings

The analysis revealed that RCSB-PDB data and services are utilized extensively, provide a significant value to its user community and creates significant societal economic impact far beyond its user community.

- **RCSB-PDB Users:** The RCSB-PDB is regularly and widely accessed, and it is estimated that there are more than 1 million unique users annually. However, for purposes of this study, a conservative estimate based on the average monthly unique users was utilized in calculating the economic impacts. Using this methodology, during fiscal year 2016, the RCSB-PDB website at rcsb.org supported an estimated **295,465 unique users.** Users hailing from 100+ countries logged more than **7 million sessions** and more than **32 million page views**.
- **Replacement Value:** While the costs of data creation and deposition are unknown, a reasonable **estimate to replicate the RCSB-PDB data archive is $12 billion** (assuming $100,000 avg. cost to replicate each entry).
- **Investment Value:** RCSB-PDB's operational costs, including the costs of data creation and deposition, annotating and adding value to the data, and other expenses total **$6.9 million** per year.
- **Economic Impacts to the State Economy:** Economic impact analysis reveals a total economic impact to New Jersey of **$8.5 million** (including multiplier effects), approximately **42 jobs** created with annual **wages of $4.7 million**, and estimated tax revenues (to local, state, and federal governments) of **$1 million annually.**
- **Access Value:** The value of time and money users spend obtaining RCSB-PDB data and services represents the economic value of their investment. The access value of RCSB-PDB data and services is estimated to be **$43.7 million** annually.
- **Use Value:** Use value represents the value (cost) of the time spent accessing and working with RCSB-PDB data. The use value of RCSB-PDB data and services is estimated at **$5.5 billion** annually, 800 times greater than RCSB-PDB's direct operating cost.
- **Contingent Valuation:** Contingent valuation involves estimating the value users place on a freely provided service. Utilizing the average contingent value gleaned from a study of EMBL-EBI, the contingent value of the RCSB-PDB data and services is estimated to be **$2,573 per user, $760 million** in total, 110 times greater than RCSB-PDB's direct operating cost.
- **Efficiency Impacts:** Efficiency impacts (i.e., productive time researchers gain from the efficiency associated with using RCSB-PDB data and services) allow researchers more time to do other work-related activities. The value of RCSB-PDB's efficiency impact is estimated at **$2.5 billion** annually.
- **Return on Public Investment in R&D:** A wider, more comprehensive measure of long-term societal impacts stemming from RCSB-PDB derives from the impact of publicly funded research that utilizes RCSB-PDB data and services. The estimated and conservative long-term value of the research and development using RCSB-PDB data is **$1.1 billion** annually. Over the next 30 years, the estimated return on investment is conservatively **$8.0 billion** in net present value.

INTRODUCTION

Established in 1971 (with 7 entries) as the first open access digital resource for biological data, the Worldwide Protein Data Bank (wwPDB) is now the single global archive of experimental 3D structures of biological macromolecules with more than 125,000 entries at the end of 2016. These biological macromolecules are critical to biomedical research; indeed, the users of wwPDB data are researchers from all over the world, from structural biologists to computational biologists and beyond. The Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB-PDB), the US data center for the wwPDB archive, provides PDB data at no charge to users through the RCSB-PDB website. RCSB-PDB operates at Rutgers, The State University of New Jersey and the University of California San Diego/San Diego Supercomputer Center.

Through a unique agreement, PDB data are also provided by other Worldwide Protein Data Bank partners (including the Protein Data Bank Europe, Protein Data Bank Japan, and the Biological Magnetic Resonance Bank). However, the RCSB-PDB team (operating at Rutgers University and the University of California San Diego), serves as the wwPDB archive keeper, and is also responsible for data integrity and disaster recovery.

The full economic benefits of the RCSB-PDB are extensive, but are not well understood. To add to public understanding of its economic benefits, the RCSB-PDB commissioned the Rutgers Office of Research Analytics to examine the economic impacts of the RCSB-PDB, including the value of the research conducted with RCSB-PDB data and services. Presented here are the results of a quantitative study of the economic impacts of the RCSB-PDB.

STUDY APPROACH

The Office of Research Analytics team conducted a quantitative study of the economic impacts of the RCSB-PDB, which are complementary to other types of impact measures (e.g., impacts on human health, scholarly citations, patents, etc.). In this section, study methods and data sources utilized in the study are described.

Study Methods

In 2016, Beagrie and Houghton (2016) released a study of the economic impacts of the European Molecular Biology Laboratory- European Bioinformatics Institute (EMBL-EBI), home of the PDB Europe. Because of the similarities between the data provided by EMBL-EBI and RCSB-PDB (the PDB data archive is identical), the Beagrie and Houghton (2016) study provides a reasonable methodological path for this study. However, it should be noted there are differences between the RCSB-PDB and the EMBL-EBI, most notably in the types of services offered. In addition, given the time constraints to conduct this impact assessment, a survey of RCSB-PDB users could not be administered, as was done in the Beagrie and Houghton (2016) study. Therefore, the team relied primarily on web analytics from the RCSB-PDB website, in addition to utilizing key factors from the Beagrie and Houghton (2016) study.

Analytical techniques intended to capture the full extent of impacts of open research data services are still evolving. Nonetheless, Beagrie's prior work on measuring impacts of "free and open data and services" (Beagrie and CSES 2012) has been cited as a particularly good example of an accepted method in the measurement of such impacts (Technopolis 2013). Following Beagrie's approach (Beagrie and CSES 2012, Beagrie and Houghton 2016), a range of methods for exploring the value and impact of RCSB-PDB's research data and services were utilized in this study. Given the similarity between the RCSB-PDB and the EMBL-EBI, and the desire to apply best practices in measuring value and economic impact of major research data services, the utilization of certain factors from the Beagrie and Houghton (2016) study is justified. To account for the level of uncertainty in the estimates due to the distinctive differences between the two programs, a range of low and high estimates for a number of the economic impacts were also calculated.

Similar to Beagrie and Houghton (2016), the following economic indicators were estimated for the RCSB-PDB: **access value** and **use value** (value of time and money users spend obtaining and using the product or service), **contingent value** (value of non-market goods and services), **efficiency value** (characterizes the relationship between costs and output), and **return on public investment**. In addition, an economic impact assessment was conducted of RCSB-PDB operations on the economy of New Jersey through input-output modeling. With the input-output model approach, expenditure and employment data from operations are used to estimate the **direct and indirect impacts of public expenditures on the regional economy**. This study estimates the direct and indirect impacts of RCSB-PDB expenditures on the New Jersey economy (in terms of jobs, wages, and output), and on overall tax revenues (i.e., local and state taxes, and federal taxes).

Data and Assumptions

Presented below is a discussion of the sources of data, key factors, and assumptions used to calculate estimates of value and economic impact. Table 1 below summarizes the key factors and their sources utilized in this study.

Table 1 – Key Factors Utilized in RCSB-PDB Impact Assessment

Factor	Value	Source
Mean time of access on RCSB-PDB site (minutes)	6.3	Analytics of PDB Web Users
Estimated annual RCSB-PDB accesses	7,120,959	Analytics of PDB Web Users
Estimated annual RCSB-PDB unique users	295,465	Analytics of PDB Web Users
Mean willingness to pay for RCSB-PDB data and services (in 2016 US dollars)	$2,573	EMBL-EBI Study
Efficiency impact factor (increased productivity due to RCSB-PDB data and services)	0.45	EMBL-EBI Study
Mean hourly cost (avg salary of researchers using RCSB-PDB data in 2016 US dollars)	$58.50	EMBL-EBI Study
Mean time with RCSB-PDB data per week (hours)	6.8	EMBL-EBI Study (20% of 34 hours performing research)
Average return to R&D (from publicly funded research)	20%	Economics Literature (ranges from 20% to 60%) EMBL-EBI study used 40%

Website analytics data was utilized to estimate the number of users of the RCSB-PDB and the average time spent by users of the website. As reported in the EMBL-EBI study (Beagrie and Houghton, 2016), there are potential difficulties when relying on website analytics data to estimate the number of unique users of a website or service. Most website analytics attempt to distinguish "unique visitors" by examining a user's IP address and cookies stored on the web browser. However, this approach has been shown to overestimate visitors (Fomitchev 2010) due to the dynamic nature of IP addresses (multiple users behind a single IP address and a single user having multiple IP addresses), the number of devices used by individual users, and the multiple locations used to access the website. RCSB-PDB's own internal analysis of website data showed that examining unique visitor data on a monthly basis minimized the effect of these factors and provided a conservative estimate of the number of unique RCSB-PDB website users.

Website analytics data from Google Analytics were captured for the time period July 1, 2015 through June 30, 2016 (Fiscal Year 2016). Monthly unique visitor data were compiled and the mean calculated to estimate the annual unique user count (295,465 users). It should be noted that this estimate of unique users is conservative, as the RCSB estimates that there are more than 1 million unique users annually. To calculate the total annual usage, the annual number of sessions and page-views were summed (7,120,959 and 32,786,301 respectively). Sessions per user were then calculated by taking the total number of sessions annually and dividing it by the estimated annual unique user count (24.1 session/user). The unadjusted mean session duration was arrived at by averaging the monthly average session duration as calculated by Google Analytics (6:19 minutes). The raw session duration data was analyzed to determine the potential impact of outliers on the average session duration estimate. It was determined that the number of small outliers far outweighed large outliers; therefore, the unadjusted mean provides a conservative estimate of session duration.

The monetary assumptions (i.e., willingness to pay, mean hourly cost) were taken from the EMBL-EBI study and were converted to US dollars from British pounds at the May 22, 2015 exchange rate, and adjusted to 2016 price levels using the consumer price index from US Department of Labor, Bureau of Labor Statistics (all urban consumers, US City Average). Based on labor assumptions in the Beagrie and Houghton (2016) study, it is estimated that users of the RCSB-PDB data and services (researchers and educators from around the world) work on average 34 hours per week, spend about 20% of that time using PDB data (6.8 hours/wk), and work 47 weeks per year.

To account for uncertainty associated with some of the above factors and differences in the programs, a sensitivity analysis was performed by providing a range of economic impacts. A low, middle, and high economic impact estimate is provided by changing the following factors in the formulas:

- mean time of access (+/- 20%)
- mean willingness to pay (+/- 20%)
- efficiency impact factor (+/- 20%)
- mean time with RCSB-PDB data (+/- 20%)
- average return to R&D (20% for low estimate, 60% for high estimate).

For proper context, all estimates are expressed as an annual value in current prices and at current levels of activity (i.e., 2016).

IMPACT ASSESSMENT

Investment Value

Investment value includes RCSB-PDB's operational costs, costs of data creation and deposition (from contributors), costs of annotating and adding value to the data (from collaborators), etc. While the costs of data creation and deposition are unknown, a reasonable **estimate to replicate the RCSB-PDB data archive is $12 billion** (assuming $100,000 average cost to replicate each entry). The annual operating costs of the RCSB-PDB are approximately $6.9 million per year ($5.1 million spent on operations in New Jersey and $1.8 million spent in California). These operating costs include all operating expenses (e.g., equipment, maintenance, supplies, salary and fringe, etc.). The operating costs can be considered the minimum annual investment value needed to keep RCSB-PDB archive available for public use. The annual operating costs in New Jersey of $5.1 million are used to estimate the economic impact to the New Jersey economy (presented below).

Economic Impacts to the New Jersey Economy

Economic impact analysis through input-output modeling is based on the premise that a regional economy is built upon inter-industry dependencies and that the expansion (or contraction) of economic activity in one industry will have ripple effects that move throughout the entire economic system. Technology firms, for example, utilize a wide range of business services (e.g., accounting, advertising, legal services, etc.). Changes in the business volume conducted by firms will impact these, and other, support industries. These are often referred to as "indirect effects." Economic activity by technology firms also impact other sectors of the economy through increased household spending. Wages paid by tech firms are spent and re-spent throughout the economy on a variety of goods and services ranging from dental and medical care to real estate to restaurant meals. These "induced effects" ripple throughout the economy. The multiplier effects (both indirect and induced) tend to diminish over time because each round of spending is reduced by the amount of money directed outside of the economic region. These "leakages" include payments to social security, income taxes, personal savings, and payments for imported goods and services. IMPLAN, a widely used input-output modeling system, provides a model of the New Jersey economy and is used to measure the economic impacts of the RCSB-PDB operation located in New Jersey.

The economic impact analysis that follows is based on a typical year of RCSB-PDB operations. In performing an economic impact analysis for the RCSB-PDB, 2016 expenditures of $5.1 million was used. Note that there are additional RCSB-PDB operations in California (approximately, $1.8 million in expenditures) which have local impacts to the California economy; however for purposes of this study, the research team focused only on the impacts to New Jersey. Table 2 shows the annual economic impacts to the New Jersey economy from the RCSB-PDB operations. The RCSB-PDB generates a total of $5.1 million in economic output annually, resulting in a total **economic impact to New Jersey of $8.5 million** (including multiplier effects). Twenty-four employees are directly employed by the RCSB-PDB in New Jersey and earn $2.7 million annually (including fringe benefits). The total number of jobs created (directly and through multiplier effects) is estimated to be approximately forty-two jobs with annual wages of $4.7 million.

Table 2 –Economic Impact of RCSB-PDB to the New Jersey Economy

Impact Type	Employment	Labor Income	Output
Direct Effect	24.0	$2,747,000	$5,113,000
Indirect/Induced Effect	18.6	$1,987,000	$3,381,000
Total Effect	**42.6**	**$4,734,000**	**$8,494,000**

Table 3 summarizes the tax impacts associated with the economic activity of RCSB-PDB's New Jersey operations. It is estimated that state and local tax impacts are $395,000 annually and the federal tax impact $630,000 annually. **Thus, the total tax impact of RCSB-PDB operations is an estimated $1 million.**

Table 3 – Summary of Tax Impacts from RCSB-PDB Operations

Impact Type	State/Local	Federal	Total
Direct Effect	$197,000	$308,000	$505,000
Indirect/Induced Effect	$198,000	$322,000	$520,000
Total Effect	**$395,000**	**$630,000**	**$1,025,000**

Access Value

Access value and use value (presented in the next section) represent the value of time and money users spend obtaining and using the product or service. Estimates of the economic value of the investment made by users in access and use reveal the minimum value that the product or service is worth to them. As previously stated, web analytics data reveal that the average time users spend accessing RCSB-PDB services is 6.3 minutes. Annual accesses to RCSB-PDB data and services in 2016 totaled 7.1 million. Assuming a mean hourly cost of $58.50, the **access value of RCSB-PDB data and services is estimated to be $43.7 million annually** (with a range of $34.7 – $52.7 million) (Table 4). **The estimated access value of RCSB-PDB data and services is 6 times greater than RCSB-PDB's direct operating cost of $6.9 million.**

Table 4 – Access Value of RCSB-PDB Data and Services

Factor	Low	Mid	High
Mean time of access (minutes)	5.0	6.3	7.6
Mean hourly cost ($)	$58.5	$58.5	$58.5
Annual accesses	7,120,959	7,120,959	7,120,959
Access Value ($)	**$34,715,000**	**$43,740,000**	**$52,766,000**

Access value = (mean time of access * mean hourly cost * estimated annual accesses)

Use Value

Use value can be estimated by the cost of the time that users spend accessing and using a product or service. The number of users and their average time spent accessing and working with RCSB-PDB data are the drivers of the use value estimate. It is assumed that 295,465 RCSB-PDB users spend 6.8 hours per week (on average) accessing and working with RCSB-PDB data. Based on this data, the **use value of RCSB-PDB data and services is estimated at $5.5 billion annually** (Table 5). Varying the hours per week users access and work with RCSB-PDB data +/- 20% yields use value estimates ranging from $4.4 to $6.7 billion annually. **The estimated use value of RCSB-PDB data and services is 800 times greater than RCSB-PDB's direct operating cost.**

Table 5 – Use Value of RCSB-PDB Data and Services

Factor	Low	Mid	High
Mean time/week with PDB data (hrs)	5.4	6.8	8.2
Mean hourly cost ($)	$58.5	$58.5	$58.5
Estimated users	295,465	295,465	295,465
Use Value ($)	**$4,386,857,000**	**$5,524,191,000**	**$6,661,524,000**

Use value = (mean time with RCSB-PDB data per week * mean hourly cost * weeks per year * estimated users)

Contingent Valuation

The value of public services (such as RCSB-PDB's data and services) are often not well understood because they are not quantifiable by using standard methods to value private market commodities. Such freely available products and services often require specialized non-market economic valuation methods to value public goods and services not traded in markets (Freeman 1993; Mitchell & Carson 1990; Perman et al. 1999). One such method, contingent valuation, involves estimating the value of non-market goods and services based on preference theory. Preference theory states that a good or service which contributes to human welfare has economic value, and something that satisfies an individual's preferences contributes to the individual's welfare. Further, individual preferences are revealed by one's willingness to pay for a good or service. In the case of non-market goods and services, individuals can be asked what they would pay (i.e., contingent value) for a good or service in a hypothetical market situation. Contingent valuation has been applied for decades to estimate costs and benefits of many public goods and services, and economists have found that the estimates provide reasonable and consistent starting points in evaluating the value of those public goods and services (Kopp et al. 1997). The underlying theory of contingent valuation is consistent with the challenge of measuring the value that results from RCSB-PDB's goods and services.

Utilizing the **mean contingent value of $2,573** from Beagrie and Houghton (2016) as the average willingness to pay for RCSB-PDB data and services among users (295,465), the **contingent value of the RCSB-PDB data and services is estimated at $760.2 million** (with a range of $608.1 – $912.4 million) (Table 6). **This estimated contingent value is 110 times greater than RCSB-PDB's direct operating cost.**

Table 6 – Contingent Value of RCSB-PDB Data and Services

Factor	Low	Mid	High
Mean willingness to pay ($)	$2,058	$2,573	$3,088
Estimated users	295,465	295,465	295,465
Contingent Value	**$608,067,000**	**$760,231,000**	**$912,396,000**

Contingent value = (mean willingness to pay * estimated users)

Efficiency Impacts

Efficiency characterizes the relationship between costs and output (or profit in the case of private industry). As an example, a researcher producing the same level of output at lower costs due to access to a new resource has achieved an efficiency gain. The RCSB-PDB data and services allow researchers and educators to do their work more efficiently, freeing up their time to do other work-related activities. The RCSB-PDB's efficiency impact can be estimated by measuring the value of the efficiency gains enjoyed by RCSB-PDB users (i.e., the value of time saved or the costs avoided by not having to create the data themselves or obtain it elsewhere). To calculate this impact, it is assumed that RCSB-PDB users are 45% more efficient (Beagrie and Houghton 2016) during the time they spend working with RCSB-PDB data. Analysis estimates the **efficiency impact of the RCSB-PDB to be $2.5 billion** annually (Table 7). Changing the efficiency impact factor of RCSB-PDB data +/- 20%, yields efficiency impact estimates ranging from $2 to $3 billion annually. **The estimated efficiency value of RCSB-PDB data and services is 360 times greater than RCSB-PDB's direct operating cost.**

Table 7 – Efficiency Impact of RCSB-PDB Data and Services

Factor	Low	Mid	High
Efficiency impact factor	0.36	0.45	0.54
Mean time/year with PDB data (hrs)	319.6	319.6	319.6
Mean hourly cost ($)	$58.5	$58.5	$58.5
Estimated users	295,465	295,465	295,465
Efficiency Impact	**$1,988,709,000**	**$2,485,886,000**	**$2,983,063,000**

Efficiency impact = (efficiency impact * time with data * mean hourly cost * estimated users)

Return on Public Investment in R&D

A more comprehensive measure of the long-term impacts of the RCSB-PDB data and services is the impact of the research which stems from them. Similar to Beagrie and Houghton (2016), the potential return on public investment from the research time spent with RCSB-PDB data and services using a modified Solow-Swan model (Houghton and Sheehan 2009) is explored.

Since return on investment impacts recur throughout the useful life of the data archive, return on investment impacts over time are modeled to estimate the overall value of the returns. There is no dearth of research on the economic impacts of public investment in research and development (Bernstein et.al. (1991), Bonte (2003), Fraumeni and Okubo (2005), Nordhaus (2003), and Toole (2007)). Houghton, J.W. and Sheehan, P. (2009) put forth 25% as a conservative estimate of the social return on public investment of R&D, and suggest a range of 20-60% as reasonable. To provide a range of impacts, return on investments of 20%, 40%, and 60% estimated results are provided in this report. It should be noted that research using RCSB-PDB data is both publicly and privately funded; however, the distribution of public vs private funding is unknown. The return on investment estimates below stem primarily from publicly funded research that utilizes RCSB-PDB data. In an effort to be conservative, ROI impacts estimated using a 20% return on investment rate are highlighted in the results.

As Table 8 shows, the conservative estimated value of public R&D using RCSB-PDB data is $1.1 billion annually (range of $1.1 – $3.3 billion). The value of the **returns from R&D is estimated at $8.0 billion in net present value** (over 30 years).[4] While this value cannot be attributed 100% to RCSB-PDB, these estimates give a measure of scale to the activities to which RCSB-PDB data and services make an important contribution.

Table 8 – Return on Public Investment in R&D using RCSB-PDB Data and Services

Factor	Low	Mid	High
Average return to R&D (%)	20%	40%	60%
Mean time/week with PDB data (hrs)	6.8	6.8	6.8
Mean hourly cost ($)	$58.5	$58.5	$58.5
Estimated users	295,465	295,465	295,465
Return on R&D	**$1,104,838,000**	**$2,209,676,000**	**$3,314,515,000**

Return on R&D = (mean time with PDB data per week * mean hourly cost * weeks pa * estimated users * average return to R&D)

Return on Privately Funded Investment in R&D (case studies)

The full extent that private companies utilize RCSB-PDB data is unknown (companies are able to download RCSB-PDB data anonymously). However, it is well known that RCSB-PDB data is utilized extensively in commercial applications and the returns (to the companies and society) can be enormous. Two examples of the impact of RCSB-PDB data (on chronic myeloid leukemia and human immunodeficiency virus) are presented below to highlight the nature of the return on investment from privately funded research utilizing RCSB-PDB data.

[4] To calculate net present value, the value of the data is depreciated at a rate of 5% per year over 30 years (geometric depreciation). The annual returns from each year's impact are allocated over 15 years using a normal distribution. A discount rate of 3.5% is applied to calculate net present value.

Chronic myeloid leukemia (CML) is a type of cancer that starts in certain blood-forming cells of the bone marrow. In 2001, Gleevec was released as a treatment for CML. However, resistance challenges exist with Gleevec and many patients have an inferior response to Gleevec, either initially or over time after a number of treatments. RCSB-PDB data were critical in understanding the molecular basis for the effectiveness of Gleevec and instrumental in the development of two superior treatments, Sprycel and Tasigna (released in 2006 and 2007, respectively). The 5-year survival of CML patients increased to 66.9% after the release of these two new treatments (versus 34.2% in 1995). In 2016, sales of Sprycel/Tasigna were approximately $3.5 billion worldwide.

The human immunodeficiency virus (HIV) is a type of retrovirus that causes HIV infection and leads to acquired immunodeficiency syndrome (AIDS). AIDS involves the progressive failure of the immune system, allowing life-threatening infections and cancers to develop and thrive. RCSB-PDB data were instrumental in the development of most of the current HIV treatments (i.e., Reverse Transcriptase and Protease Inhibitors). These treatments have dramatically increased the life expectancy of HIV infected persons (by more than 10 years), and resulted in remarkable improvements in quality of life. Moreover, Reverse Transcriptase and Protease Inhibitors accounted for more than $18 billion in sales in 2015.

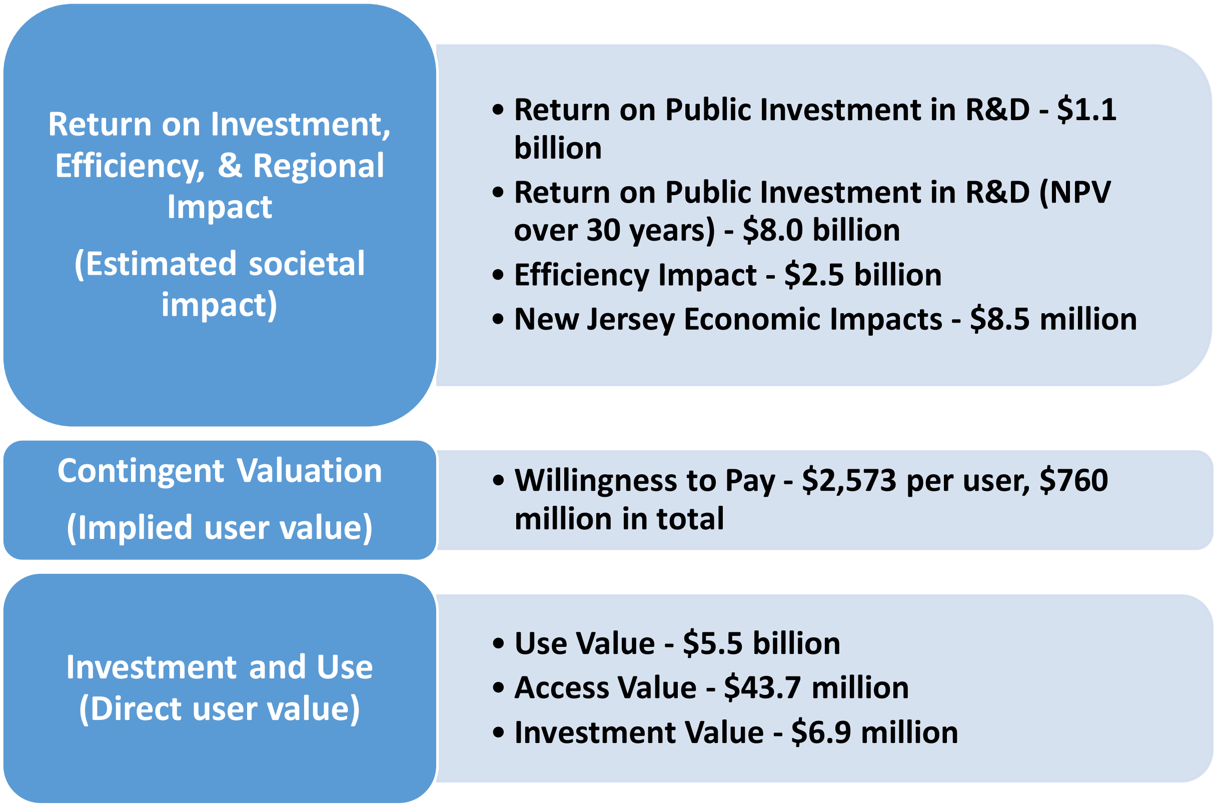
LIMITS OF THE STUDY

Given the time constraints to conduct this impact assessment, the study team was unable to administer a survey to RCSB-PDB users. Relying on key factors in the Beagrie and Houghton (2016) study adds a level of uncertainty to our estimates. To compensate for this uncertainty, a range (low and high) for many impact measures were estimated. Thus, it is recommend that a survey of RCSB-PDB users be administered in the future in order to verify key information and to update this impact assessment.

SUMMARY

A summary of the estimated economic impacts of the RCSB-PDB are shown in Figure 1. As stated previously, the estimates are driven mainly by RCSB-PDB user population (number of users and levels of use), as well as key assumptions from the Beagrie and Houghton (2016) study. As such, the estimates scale linearly with the number of users and their level of use. As large as the estimates are, they may be just a fraction of the full economic impact associated with the RCSB-PDB data and services. For example, the estimates do not directly link to the incalculable impact from the saved lives and improved quality of life that results from medical research discoveries. The estimate of return on public investment from expenditures on R&D is intended to capture some of this value. To illustrate this point, consider the millions of lives saved by improved mortality from new treatments for cancer and other devastating diseases over the past decades, as demonstrated in the two case studies. Overall, millions of people have been able to continue with their lives and contribute to the economy as a result of medical research that utilized the resources of the RCSB-PDB.

Figure 1 - Summary of Annual Impacts of RCSB-PDB Data and Services



Return on Investment, Efficiency, & Regional Impact

(Estimated societal impact)

- Return on Public Investment in R&D - $1.1 billion
- Return on Public Investment in R&D (NPV over 30 years) - $8.0 billion
- Efficiency Impact - $2.5 billion
- New Jersey Economic Impacts - $8.5 million

Contingent Valuation

(Implied user value)

- Willingness to Pay - $2,573 per user, $760 million in total

Investment and Use (Direct user value)

- Use Value - $5.5 billion
- Access Value - $43.7 million
- Investment Value - $6.9 million

REFERENCES

Beagrie, N. and The Centre for Strategic Economic Studies (CSES) University of Victoria (2012). "Economic Impact Evaluation of the Economic and Social Data Service", Economic and Social Research Council, Great Britain.

Beagrie, N. and Houghton, J.W. (2014). "The Value and Impact of Data Sharing and Curation: A Synthesis of Three Recent Studies of UK Research Data Centres", Joint Information Systems Committee (Jisc), Bristol and London.

Beagrie, N. and Houghton, J.W. (2016). "The Value and Impact of the European Bioinformatics Institute", Bristol and London.

Bernstein, Jeffrey and Nadiri, M. Ishaq (1991), "Product Demand, Cost of Production, Spillovers, and the Social Rate of Return to R&D", National Bureau of Economic Research Working Paper 3625, Cambridge, 1991.

Bonte, Werner (2003), "Does Federally Financed Business R&D Matter for U.S. Productivity Growth?", Applied Economics (October 2003), pp. 1619-1625.

Dey-Chowdhury, S. (2008), "Perpetual inventory method", Economic & Labour Market Review 2(9), September 2008, pp48-52. Office of National Statistics.

Fomitchev, M.I. (2010), "How Google Analytics and conventional cookie tracking techniques overestimate unique visitors", Academia.

Fraumeni, Barbara and Okubo, Sumiye (2005), "R&D in the National Income and Product Accounts: A First Look at its Effect on GN", pp. 275-316 in Corrado, Carol, Haltiwanger, John, and Sichel, Daniel (Eds.), "Measuring Capital in the New Economy", National Bureau of Economic Research, University of Chicago Press, Chicago, 2005.

Freeman, A. M. III. (1993), "The Benefits of Environmental Improvement: Theory and Practice", Washington DC: Resources for the Future.

Houghton, J.W. and Sheehan, P. (2009), "Estimating the Potential Impacts of Open Access to Research Findings", Economic Analysis and Policy 39(1).

Kopp, R. J., Pommerehne, W.W. & Schwarz, N. (eds.) (1997), "Determining the Value of Non-Market Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods", Boston: Kluwer Academic Publishers.

Nordhaus, William D. (2003), "The Health of Nations: The Contribution of Improved Health to Living Standards", pp. 41-73 in Murphy, Kevin M. and Topel, Robert H. (Eds.), "Measuring the Gains from Medical Research: An Economic Approach", University of Chicago Press, Chicago, 2003.

Mitchell, R. C. & Carson, R. T. (1990), "Using Surveys to Value Public Goods: The Contingent Valuation Method", Washington DC: Resources for the Future.

Perman, R., Ma, Y., McGilvray, J. & Common, M. (1999), "Natural Resource and Environmental Economics", 2nd edn, New York: Pearson Education, Inc.

Sveikauskas, L. (2007), "R&D and Productivity Growth: A Review of the Literature", Washington DC.: US Bureau of Labor Statistics Working Paper 408.

Technopolis (2013), "Big Science and Innovation, Report to the Department of Business, Innovation and Skills", London.

Toole, Andrew (2007), "Does Public Biomedical Research Complement Private Pharmaceutical Research and Development Investment?", Journal of Law and Economics
(February 2007), pp. 81-104.