Return on Investment of Streamlined Data Mining and Analysis

Pipeline Pilot Improves Research Efficiency, Productivity, and Costs, Delivering ROI of up to 7:1

TABLE OF CONTENTS

Data Overload

Scientific data is at the heart of all research and development organizations—guiding efforts, prompting decisions, and helping to create knowledge and build success. But for many organizations, as scientific data volume has grown exponentially, it has become more than just empowering; it has also become burdening.

Many organizations struggle to contend with vast amounts of diverse, non-standardized data. Their data is isolated in disconnected databases, not shared among research groups. As such, unlocking the true value of this data is costly, inefficient, and time-consuming.

Often, scripting and manual processes are needed to acquire and merge data; incompatible software programs or customized programming are needed to analyze data; results must be cut-and-pasted between databases and applications; reports must be manually created.

These obstacles waste time. They invite error. And, they limit the number of individuals who can perform certain jobs, often forcing computational experts or IT resources to become involved.

Clearly, wasted time and resources lead to unnecessary financial expenses; they also threaten the overall discovery process by delaying decisions and forcing researchers to make choices with incomplete or inaccurate data.

This white paper explores how a solution that streamlines the integration and analysis of vast quantities of diverse data can provide organizations with a significant ROI (in some cases over 700%). Details are provided on four use cases:

- High-throughput screening (HTS) data processing at a large pharmaceutical company
- Corporate database management and profiling at a biotech company
- Property profiling at a large pharmaceutical company
- Library enumeration at a large pharmaceutical company

While the experiences of pharmaceutical and biotech companies are detailed in this white paper, users in other industries, notably those in the chemical, materials, and manufacturing industries, have also experienced similar success stories. It is also important to note that while large companies are cited in some examples within this white paper, similar financial and intangible benefits could be realized at smaller companies as well.

Ask More of Your Data with Pipeline Pilot

A research and development organization can only be as effective and profitable as its data allows. As described, many organizations struggle because there are too many obstacles that prevent them from leveraging data, not because they don't have enough data. Pipeline Pilot helps organizations overcome these obstacles, letting them realize more value from their data.

Pipeline Pilot is a powerful client-server platform that streamlines the integration and analysis of the vast quantities of data flooding research informatics. It has been designed to meet the critical requirements of informatics professionals, offering:

- An open server architecture
- A vast array of componentized science
- An agile development environment
- Fast, secure, and flexible deployment options
- Minimal maintenance costs
- Application extensibility

With Pipeline Pilot, organizations can improve:

- **Collaboration** by uniting databases, applications, and algorithms from diverse fields, locations, and vendors within one platform.
- **Accuracy** by merging and processing diverse types of data (numeric, textual, chemical, biological, materials) in real-time, from original sources, without cutting-and-pasting.
- **Efficiency** by automating routines to retrieve, merge, analyze, and report data.
- **Resource utilization** by enabling users to create, share, and utilize workflows without complex scripting or programming.
- **Decision-making** by providing better access to the tools researchers need, as well as to interactive reports from past and present projects.
- **Exploration** by freeing researchers from mundane tasks and allowing more time for performing additional experiments and asking more 'what if' questions.

With Pipeline Pilot, organizations can improve the overall research process, potentially speeding up time to market. On top of that, Pipeline Pilot can lead to significant cost savings, which are realized through reduced expenses, licensing fees, and infrastructure costs. On the subsequent pages of this white paper, we detail the financial benefits that have been seen by four organizations that used Pipeline Pilot to create customized solutions for their research needs.

The Business Case for Pipeline Pilot—Financial and Intangible ROI

In collaboration with an independent research organization that specializes in ROI and cost-benefit analysis, Accelrys conducted an objective study of the financial returns generated by the use of Pipeline Pilot over a three-year period.

In all four scenarios examined in this white paper, the organizations experienced significant cost savings as a result of the automation of complex processes that previously took significant amounts of time, effort, and money. The companies also reaped significant intangible benefits that cannot be easily quantified. For example, as described in the following section, a company that used Pipeline Pilot for HTS data processing found that the whole process became sufficiently faster, allowing researchers to analyze data using a number of different hit thresholds, thereby helping those researchers attain better results. Similarly, in a subsequent example explored on the following pages, a company that used Pipeline Pilot for profiling found that the cost and time savings in turn made it possible for its chemists to perform more profiling, thereby helping them better understand the compounds they were studying.

Property Profiling

In the first scenario, bench chemists at a large pharmaceutical company frequently needed to attain property profiles for their compounds of interest. However, in order to get accurate profiles, the chemists had to hand off the work to computational experts. Because no automated profiling tools were in place, the computational experts had to use a complicated workflow to attain accurate profiles. The workflow involved eight applications and utilities, intermediate file conversion, result collation, and manual reporting of results to the chemists. The entire process took a whole day, was prone to error, and had to be performed once every other day to keep up with the frequent requests of chemists. Due to the sheer volume of requests, turnaround time was two to five days.

To streamline this workflow and better utilize resources, the company used Pipeline Pilot to create a profiling solution that enabled bench chemists to profile their compounds via an easy-to-use application that was accessed via a Web browser. By providing Web-based access to a Pipeline Pilot protocol that the computational chemist had designed and validated, the solution not only empowered the chemists and eliminated the burden on computational experts, it also helped ensure the company's best practices were still in place.

The new process takes approximately five minutes, eliminates turnaround time, and is more accurate than the previous manual procedure. It is now performed between five and 10 times per day.

By using Pipeline Pilot in this way, the pharmaceutical company has seen significant financial benefits—a three-year cumulative net benefit of $164,250. The project has an ROI of 121%, and a payback period of five months. Details are provided in Table 1 and Figure 1.

In addition to gaining substantial financial benefits, the company experienced many indirect benefits. For example, the error-prone and time-consuming process that was in place before the Pipeline Pilot solution caused chemists to often make decisions based on incomplete or inaccurate information. With the use of Pipeline Pilot, profiling can now be performed much more quickly and inexpensively, and as a result, it is used much more frequently. The more frequent use and thorough profiling mean chemists have more accurate and comprehensive information on which to base decisions.

It is also important to note that in this particular example the pharmaceutical company chose to replace and automate its entire profiling process with a Pipeline Pilot solution; this included replacing its existing software for scientific calculations. However, Pipeline Pilot can be used to automate the profiling process even if a company chooses to keep its existing third-party or in-house software for performing calculations.

Table 1: Financial ROI of a Pipeline Pilot Property Profiling Solution at a Large Pharmaceutical Company

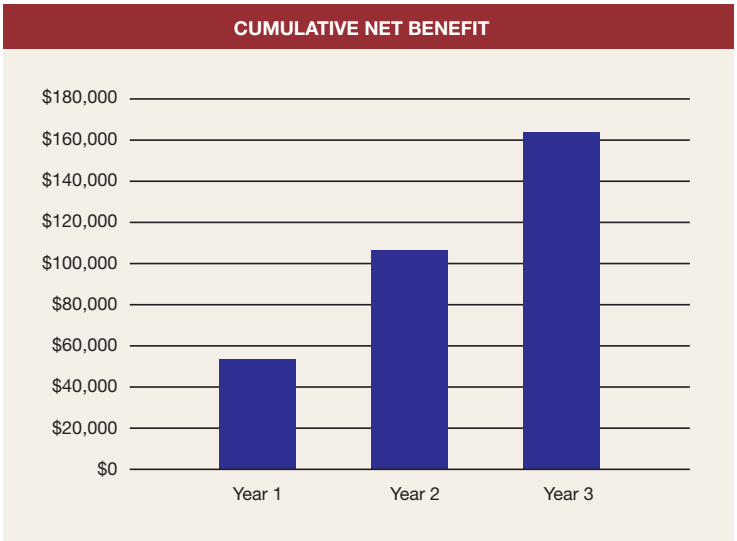
Project Summary*						
ROI: 121%						
Net Present Value: $135,197						
Cumulative Net Value: $164,250						
Project Costs**		**Start Up**	**Year 1**	**Year 2**	**Year 3**	**Total**
TOTAL PROJECT COSTS		$37,500	$10,417	$43,917	$43,917	$135,750
Benefits			**Year 1**	**Year 2**	**Year 3**	**Total**
Elimination of the Previous Manual Solution			$100,000	$100,000	$100,000	$300,000
TOTAL BENEFITS			$100,000	$100,000	$100,000	$300,000
Financial Analysis		**Start Up**	**Year 1**	**Year 2**	**Year 3**	
Net Value		($37,500)	$89,583	$56,083	$56,083	
Cumulative Net Value		($37,500)	$52,083	$108,167	$164,250	
Net Present Value	$135,197					
Payback Period (months)	5					
ROI	**121%**					

* For an explanation of the terms used in this and other charts, and an explanation of how the results were calculated, see the Appendix at the end of this paper.

** Project costs include implementation, time investment and server and client investment.

Figure 1: Cumulative Net Benefit—Pipeline Pilot for Property Profiling (Large Pharmaceutical Company)

After implementing a property profiling solution using Pipeline Pilot, the pharmaceutical company cited in this example will realize a cumulative net benefit of over $160,000 by Year 3.



HTS Data Processing

HTS data processing is an essential technology for pharmaceutical companies that want to accelerate the discovery, development, and manufacturing of potential new drugs. Unfortunately, rapid data collection, proper data analysis, accurate hit identification, and efficient quality control present significant challenges—these processes are both time-consuming and expensive.

Pipeline Pilot enables companies to quickly and easily create custom applications for HTS data processing. Its data gathering and integration tools can be applied across the entire high throughput discovery process. Data can be accessed and merged from several sources, such as libraries used in the screen, images, numerical data, lab notebook data, and Microsoft® Excel® files containing experimental design details and notes. Once data is collected and merged, it can be analyzed and reported using a wealth of integrated Pipeline Pilot tools, including those for chemistry, sequence analysis, R-statistics, data modeling, imaging, and reporting. With these tools, companies can create a truly customized solution that streamlines the entire high throughput discovery process. Alternatively, companies can choose to use the analysis software they already have in place, while using Pipeline Pilot to integrate and automate processes.

In this particular example, a large pharmaceutical company that was struggling with the challenges of HTS turned to Pipeline Pilot to create a custom solution. Before implementing Pipeline Pilot, the company was manually processing HTS data—a process typical of so many companies today. In this manual process, the HTS scientists copied data from a database and loaded it into Microsoft® Excel®. A series of macros and manual operations were then run on the data. The process was slow, time-consuming, and prone to error. In total, it took four days, and it had to be performed 20 times per year.

Using Pipeline Pilot, the process was automated. It now takes 14 minutes instead of four days, and it only needs to be performed 15 times per year, instead of 20 times per year. As a result of this increased productivity and reduced costs, the company stands to save $101,950 over three years, which equates to an ROI of 113%, and a payback period of six months. Details are provided in Table 2 and Figure 2.

In addition to the financial benefits, because data processing is now sufficiently faster, it has enabled researchers to examine the data using a number of different hit thresholds, helping them explore more samples and yielding potentially better results.

Table 2: Financial ROI of a Pipeline Pilot
HTS Data Processing Solution at a Large Pharmaceutical Company

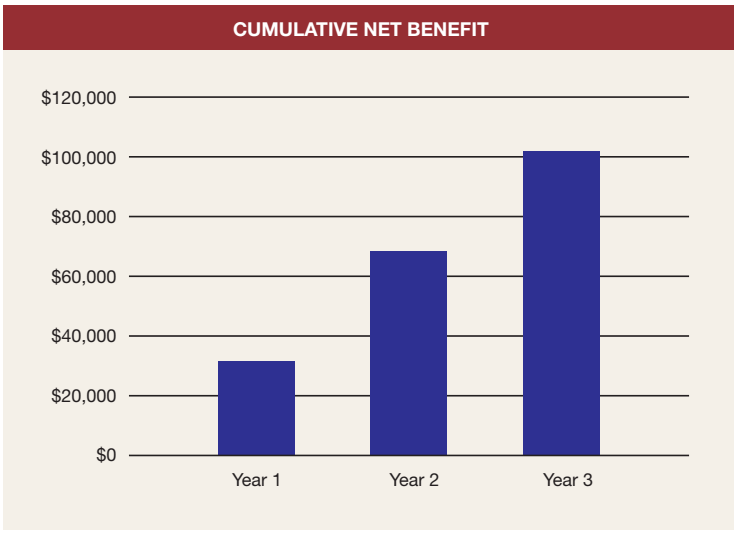
Project Summary*						
ROI: 113%						
Net Present Value: $83,315						
Cumulative Net Value: $101,950						
Project Costs**		**Start Up**	**Year 1**	**Year 2**	**Year 3**	**Total**
TOTAL PROJECT COSTS		$31,000	$350	$29,350	$29,350	$90,050
Benefits			**Year 1**	**Year 2**	**Year 3**	**Total**
Elimination of the Previous Manual Solution			$64,000	$64,000	$64,000	$192,000
TOTAL BENEFITS			$64,000	$64,000	$64,000	$192,000
Financial Analysis		**Start Up**	**Year 1**	**Year 2**	**Year 3**	
Net Value		($31,000)	$63,650	$34,650	$34,650	
Cumulative Net Value		($31,000)	$32,650	$67,300	$101,950	
Net Present Value	$83,315					
Payback Period (months)	6					
ROI	**113%**					

* For an explanation of the terms used in this and other charts, and an explanation of how the results were calculated, see the Appendix at the end of this paper.

** Project costs include implementation, time investment and server and client investment.

Figure 2: Cumulative Net Benefit—Pipeline Pilot for HTS Data Processing (Large Pharmaceutical Company)

After implementing an HTS data processing solution using Pipeline Pilot, the large pharmaceutical company cited in this example will realize a cumulative net benefit of over $100,000 by Year 3.



Corporate Database Management and Profiling

In the third scenario, the compound collection of a large, multi-site biotech company had grown through numerous corporate mergers. As a result, the collection contained many compounds that were either inappropriate, or were not suitable for screening. However, these compounds were not eliminated from analyses because the company had no efficient way of identifying which compounds should be tested and which should not. As a result, expensive, unnecessary tests were being run on compounds that should never have been screened in the first place. Consequently, the company was spending approximately $300,000 per year in unnecessary screening costs.

The company chose Pipeline Pilot to reduce the costs of unnecessary screening. To begin creating a customized solution, chemists were asked to provide example compounds that were active, and others that should be eliminated from screening. This data was then input into Pipeline Pilot, which used a Bayesian model to learn the consensus of the chemists, creating customized rules that could be used to identify which compounds should be kept in for analyses and which should be eliminated from the collection altogether. The model can be easily revised at any time.

It took only three hours to set up this initial process, which eliminated 20% of the compounds that the company was unnecessarily screening. As a result of this more in-tune process, the company will realize significant financial benefits, saving $751,000 over three years. The deployment had an ROI of 504% and a payback period of two months. Details are provided in Table 3 and Figure 3.

Over time, there may be even greater savings than projected in initial estimates because the ease of re-configuring the model will allow the company to continually refine the screening rules, which could make the model more accurate over time.

Table 3: Financial ROI of a Pipeline Pilot
Database Management and Profiling Solution at a Multi-site Biotech Company

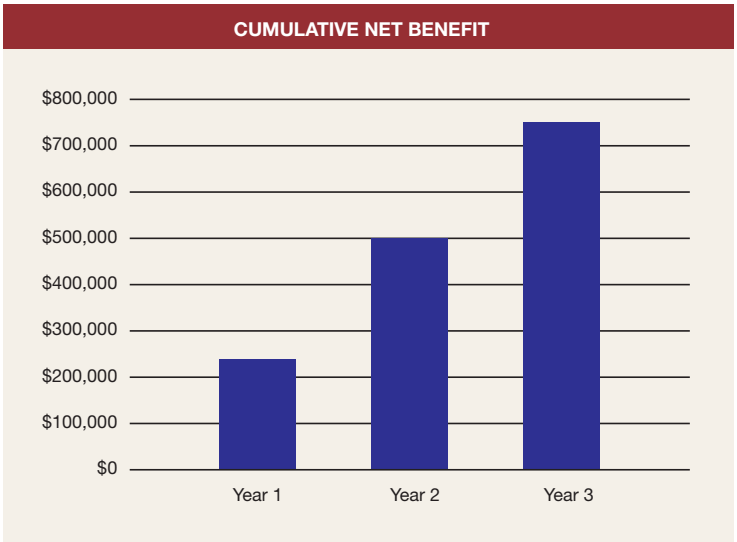
Project Summary*						
ROI: 504%						
Net Present Value: $629,309						
Cumulative Net Value: $751,000						
Project Costs**		**Start Up**	**Year 1**	**Year 2**	**Year 3**	**Total**
TOTAL PROJECT COSTS		$51,000	$0	$49,000	$49,000	$149,000
Benefits			**Year 1**	**Year 2**	**Year 3**	**Total**
Elimination of the Previous Manual Solution			$300,000	$300,000	$300,000	$900,000
TOTAL BENEFITS			$300,000	$300,000	$300,000	$900,000
Financial Analysis		**Start Up**	**Year 1**	**Year 2**	**Year 3**	
Net Value		($51,000)	$300,000	$251,000	$251,000	
Cumulative Net Value		($51,000)	$249,000	$500,000	$751,000	
Net Present Value	$629,309					
Payback Period (months)	2					
ROI	**504%**					

* For an explanation of the terms used in this and other charts, and an explanation of how the results were calculated, see the Appendix at the end of this paper.

** Project costs include implementation, time investment and server and client investment.

Figure 3: Cumulative Net Benefit—Pipeline Pilot for Corporate Database Management and Profiling (Multi-site Biotech Company)

After implementing a corporate database profiling and management solution using Pipeline Pilot, the biotech company cited in this example will realize a cumulative net benefit of over $700,000 by Year 3.



Library Enumeration and Design Services

This final scenario is similar to the one for property profiling. A large pharmaceutical company needed to design, enumerate, and profile proposed combinatorial libraries. Bench chemists were not able to perform the complex calculations required, and so they turned the work over to computational chemists, who used a complicated workflow involving several applications and utilities, intermediate file conversion, and the collation of results before sending the results back to the requesting chemist. This multi-step procedure took six days to complete, was done 55 times a year, and was prone to error.

As with the property profiling scenario, Pipeline Pilot was used to create a solution that enables chemists to directly submit their compounds for profiling via a Web browser. The chemists' easy-to-use web-based application actually calls a series of underlying Pipeline Pilot protocols that have been customized by the computational experts to ensure proper processes are upheld. There are protocols that link to databases to retrieve existing libraries, reagents, and reactions, as well as those that build and analyze new virtual libraries. These protocols use both Pipeline Pilot's native enumeration components, as well as other third-party applications that had been wrapped as SOAP services.

With the Pipeline Pilot solution in place, this process now takes approximately five minutes; the six-day turnaround time has been eliminated, and results are more accurate because manual, error-prone procedures have been eliminated. Each day, the new process is now performed between five and 10 times.

By using Pipeline Pilot in this way, the pharmaceutical company has seen significant financial benefits—a three-year cumulative net benefit of $701,625. The project has an ROI of 776%, and a payback period of one month. Details are provided in Table 4 and Figure 4.

The company has also experienced indirect benefits. For example, before the use of Pipeline Pilot, profiling took a long time and was also prone to error; therefore, chemists often made decisions based on incomplete or inaccurate information. With the use of Pipeline Pilot, profiling can be performed much more quickly, frequently, and accurately. As a result, chemists have more accurate and comprehensive information on which to base decisions. In addition, the new automated process also allows more time for subsequent design refinement when reagent replacement is needed during synthesis.

Like in the property profiling example, the pharmaceutical company in this example chose to automate the entire process with Pipeline Pilot, as well as replace existing software for scientific calculations. However, as with the property profiling scenario, companies that want to keep existing third-party or in-house software for performing the calculations can do such, using Pipeline Pilot to automate the process.

Table 4: Financial ROI of a Pipeline Pilot
Library Enumeration and Design Service Solution at a Large Pharmaceutical Company

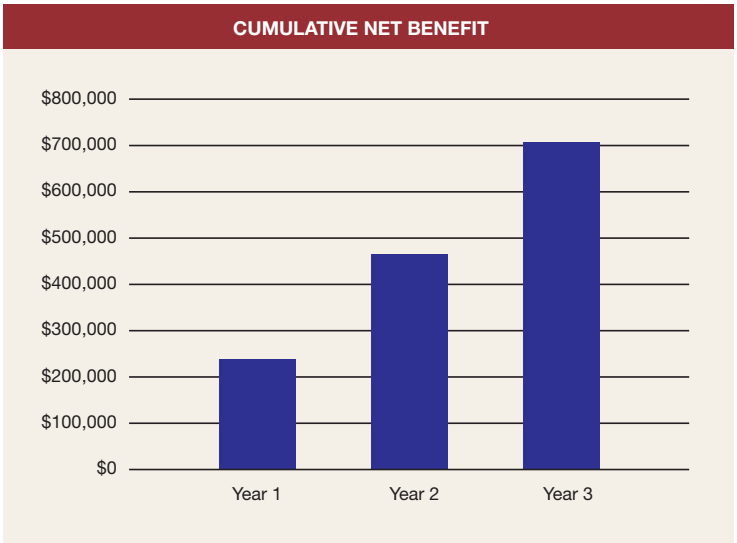
Project Summary*						
ROI: 776%						
Net Present Value: $589,300						
Cumulative Net Value: $701,625						
Project Costs**		**Start Up**	**Year 1**	**Year 2**	**Year 3**	**Total**
TOTAL PROJECT COSTS		$31,000	$458	$29,458	$29,458	$90,375
Benefits			**Year 1**	**Year 2**	**Year 3**	**Total**
Elimination of the Previous Manual Solution			$264,000	$264,000	$264,000	$792,000
TOTAL BENEFITS			$264,000	$264,000	$264,000	$792,000
Financial Analysis		**Start Up**	**Year 1**	**Year 2**	**Year 3**	
Net Value		($31,000)	$263,542	$234,542	$234,542	
Cumulative Net Value		($31,000)	$232,542	$467,083	$701,625	
Net Present Value	$589,300					
Payback Period (months)	1					
ROI	**776%**					

* For an explanation of the terms used in this and other charts, and an explanation of how the results were calculated, see the Appendix at the end of this paper.

** Project costs include implementation, time investment and server and client investment.

Figure 4: Cumulative Net Benefit—Pipeline Pilot for Library Enumeration and Design Services (Large Pharmaceutical Company)

After implementing a library enumeration and design services solution using Pipeline Pilot, the pharmaceutical company cited in this example will realize a cumulative net benefit of over $700,000 by Year 3.



Conclusion

As scientific data volume has grown exponentially, it has become more than empowering; it has also become burdensome. Making decisions and guiding research can be costly, inefficient, and time-consuming when data is non-standardized and unshared. In many organizations, researchers must undertake cumbersome processes and overcome several obstacles before they gain the insight they need to progress in their projects. As a result, the entire discovery process is hindered.

Pipeline Pilot eliminates the cumbersome processes often needed to leverage vast amounts of disconnected data. It has been designed to meet critical requirements of the informatics professional, as evidenced by its open server architecture, vast array of componentized science, agile development environment, fast, secure and flexible deployment options, minimal maintenance costs, and application extensibility. The power of data mining, analysis, and reporting tools offered by Pipeline Pilot make it easy for organizations to create customized research solutions that allow them to finally realize the full potential of their corporate data.

To illustrate and quantify the value of Pipeline Pilot, an independent research organization that specializes in ROI and cost-benefit analysis performed an objective study of the financial returns generated by the use of Pipeline Pilot in four real-life examples. In each scenario, significant financial ROI was realized, as were numerous intangible benefits. The analysis showed:

- Using Pipeline Pilot for property profiling by bench chemists, a large pharmaceutical company will gain a cumulative, projected three-year net benefit of $164,250, with an ROI of 121%, and a payback period of five months.
- Using Pipeline Pilot for HTS data processing, a large pharmaceutical company will gain a cumulative, projected three-year net benefit of $101,950, with an ROI of 113%, and a payback period of six months.
- Using Pipeline Pilot for corporate database management profiling to reduce the number of physical experiments it had to perform, a biotech company will gain a cumulative, projected three-year net benefit of $751,000, with an ROI of 504%, and a payback period of two months.
- Using Pipeline Pilot for library enumeration, a large pharmaceutical company will gain a cumulative, projected three-year net benefit of $701,625, with an ROI of 776%, and a payback period of one month.

Similar results can be gained by companies in the chemical, materials, and manufacturing industries, and significant benefits can also be seen in smaller size companies.

Appendix

Terms Used in the Charts

Return on Investment (ROI), Net Present Value (NPV), and **Payback** should be used in conjunction to understand the rate, size, and timing of the return.

For ROI and NPV the benefit is all money saved (cost avoidance relative to purchasing something else) plus productivity benefits. The cost is the initial investment and any other annual or monthly costs. Net benefit is total benefits minus total costs for a specified time period.

For time value of money equations (NPV), we use an interest rate (generally called the cost of capital) of 10% (just for the examples below). We use the cost of capital because it represents a fair hurdle to adjust future cash flows in order to determine if those cash flows are greater than the cost of the debt and equity required to make the investment.

ROI is the percentage return expected over a specified period of time. ROI is the total benefit divided by the total costs. This ROI metric is good for assessing the multiplier provided by the benefits relative to the total costs (including investment).

$$\text{ROI} = \frac{\text{Net Benefit}}{\text{Total Costs}} \times 100 \text{ (in percent)}$$

LEGEND
Net Benefit: The total savings minus total cost
Total Costs: The total cost

Example: Net Benefit = 1,000, Total Cost = 500

$$\text{ROI} = \frac{\$1{,}000}{\$500} \times 100$$

ROI = 200%

NPV represents the cumulative present value (today's dollars) of the expected net benefit (savings minus the cost over a specified period of time in years). This dollar figure provides visibility on the actual net benefit taking into consideration the time value of money (the ongoing benefit in today's dollars). An NPV greater than zero means there is a positive benefit of making the investment.

$$\text{NPV} = \frac{\text{Net Benefit 1}}{(1+r)^1} + \frac{\text{Net Benefit 2}}{(1+r)^2} + \ldots + \frac{\text{Net Benefit n}}{(1+r)^n}$$

LEGEND
Net Benefit: The savings minus cost for each year that the NPV is to be applied.
r: The interest rate (generally called the cost of capital) for the organization.
n: The total number of years for which the calculation is to be applied.

Example: Net Benefit = 1,000, r = 10%, n = 3

$$\text{NPV} = \frac{\$1{,}000}{(1+.10)^1} + \frac{\$1{,}000}{(1+.10)^2} + \frac{\$1000}{(1+r)^3}$$

NPV = $909 + $846 + $751 = $2,506

Payback is the number of months it takes for the project to yield a positive cumulative cash flow. Payback is a key measurement of risk but does not take into account cash flows after the break-even period.

Cumulative Net Value is the difference between Total Benefits and Total Project Costs (Total Benefits – Total Project Costs).

Case Study Forum is the leader in the writing and production of ROI-focused Case Studies and White Papers. Each ROI Case Study provides insight into the business impact —the revenue, productivity and cost savings the customer achieved as a result of the investment made in the solution or service. For more information, please contact Case Study Forum at 508-380-8886, or visit **www.CaseStudyForum.com**.