NIST Advanced Manufacturing Series
NIST AMS 100-80

A Framework for Economic Decision Making in Advancing Manufacturing Industry Competitiveness

A proposed system of continuous improvement for investment analysis to advance competitiveness in U.S. Manufacturing

Douglas S. Thomas

This publication is available free of charge from:
https://doi.org/10.6028/NIST.AMS.100-80

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

NIST Advanced Manufacturing Series
NIST AMS 100-80

A Framework for Economic Decision Making in Advancing Manufacturing Industry Competitiveness

Douglas S. Thomas
Office of Applied Economics
Engineering Laboratory
National Institute of Standards and Technology

This publication is available free of charge from:
https://doi.org/10.6028/NIST.AMS.100-80

June 2026

U.S. Department of Commerce
Howard Lutnick, Secretary

National Institute of Standards and Technology
Arvind Raman, NIST Director and Under Secretary of Commerce for Standards and Technology

NIST AMS 100-80
June 2026

This manuscript was edited with the assistance of multiple AI tools, including OpenAI (2026) and Google Gemini (2026).

Publication History
Approved by the NIST Editorial Review Board on June 4th, 2026

How to Cite this NIST Technical Series Publication
Thomas, Douglas. (2026) A Framework for Economic Decision Making in Advancing Manufacturing Industry Competitiveness. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Advanced Manufacturing Series (AMS) NIST AMS 100-80. https://doi.org/10.6028/NIST.AMS.100-80

Author ORCID iDs
Douglas S. Thomas: https://orcid.org/0000-0002-8600-4795

Contact Information
douglas.thomas@nist.gov

Abstract

This report proposes an organization-wide system of continuous improvement to increase impact in advancing U.S. manufacturing competitiveness. To maximize the impact of each dollar invested in research and development (R&D), change agents (i.e., those who invest in advancing manufacturing competitiveness such as governments or trade organizations) will likely need to transition from intuition-based selection to rigorous measurement science. By treating each R&D project as an experiment with a clear hypothesis and measurable outcome, organizations can create a feedback loop that systematically improves future investment accuracy.

Keywords

Change agent; Competitiveness; Impact; Manufacturing.

Table of Contents

Executive Summary vi
1. Background: Competitiveness, Research & Development Investments, and Predictions 1
1.1. Existing Programs and Frameworks 1
1.2. Planning 3
1.3. We all make Predictions 6
1.4. Predicting with Accuracy 8
1.5. Real World Application of Increasing Forecast Accuracy 11
1.6. System Level Experiential Advancement 14
1.7. Approach for Economic Investment Analysis and Design for Manufacturing R&D Projects 15
2. An Economic Framework for Impact 17
2.1. Part One: Opportunity Mapping and Hypotheses Generation 19
2.2. Part Two: Validation and Recalibration 21
2.3. Part Three: Enterprise-Wide Utilization 21
2.4. Notes and Cautions 23
2.5. Adoption, Diffusion, and Long Time Horizons 27
2.6. Metrics and Units of Observation 30
3. Platforms 35
3.1. Platforms for Hypotheses and Economic Decision making 37
3.2. Data/Information Distribution Platforms 40
3.3. Data Collection Platforms 40
4. System of Continuous Improvement Across Change Agents 42
References 44
Appendix A. Opportunity Map (Example) 48
Appendix B. Hypothesis and Impact Data Variables (Example) 52
Appendix C. Project Entry Example: Data Infrastructure for Critical Material Recovery 55
Appendix D. Project Entry Example: Digital Twins for Advanced Manufacturing with the ISO 23247 Product 59

List of Tables

Table 2.1: Three Functions of Economics in Manufacturing R&D......17
Table 2.2: Forecasts of U.S. Additive Manufacturing Shipments by Varying Market Potential......28
Table 3.1. Example of a Framework: Risk Matrix for Evaluating Hazards......36
Table A. 1: Opportunity Map: NAICS 336100 Motor vehicle Manufacturing......49
Table B. 1: Industrial Investment Classification (IIC) System (single digit level)......52
Table B. 2: Example Illustrative Variable Set for Hypothesis and Impact Data......53
Table D. 1: Project Entry Example: Digital Twins for Advanced Manufacturing and ISO 24247......61

List of Figures

Fig. 0.1: Illustrated System of Continuous Improvement......vi
Fig. 0.2: Illustrated Hub-and-Spoke Model......viii
Fig. 1.1. Apollo manned lunar landing: Ground Operations Mission Support System Mission Profile (NASA 1969)......5
Fig. 1.2. Illustration of Prediction Accuracy......7
Fig. 1.3. Cumulative Net Present Value by Percent of ARC Cost Categories (Industrial Training and Assessment Centers 2026)......14
Fig. 2.1: Change agent R&D Investment Strategy: Flat vs. Power Law Distribution......18
Fig. 2.2: Data Cube Illustration of Manufacturing Costs......20
Fig. 2.3: Hub-and-Spoke Model Illustrated......22
Fig. 2.4: Adoption Conditions for the Framework......26
Fig. 2.5: Forecasts of U.S. Additive Manufacturing Shipments, by Varying Market Saturation Levels...29
Fig. 2.6: Rogers Variables Determining the Rate of Innovation Adoption......29
Fig. 2.7: Change Agent Path to Growing Both Real Per Capita GDP and Consumer Utility......31
Fig. 2.8: Illustration of Data Layers and Data Flow......33
Fig. 3.1: A 20th Century Barn Raising, Toronto, Canada......35
Fig. 3.2. Example of a Framework for Evaluating Projects or Investments......38
Fig. 3.3. Graphing Costs and Benefits: Examining Benefit-Cost Ratio (BCR)......39
Fig. 3.4. Graphing Costs and Benefits: Examining Net Present Value......39
Fig. 4.1. System of Continuous Improvement in Impact Implementing the Economic Framework Process......42

Glossary of Terms

- Assessment Recommendation Code (ARC): A classification system developed by the Department of Energy's Industrial Training and Assessment Center consisting of one to five digits used to categorize recommendations for manufacturing investments, such as energy management or productivity enhancements.
- Benefit-Cost Analysis: a systematic approach to evaluating the economic efficiency of a project or decision by comparing the total benefits it generates to the total costs incurred, often to determine its feasibility or social value. For more information see Boardman et al. (2018).
- Benefit-Cost Ratio (BCR): A metric that identifies the relationship between costs and benefits.
- Change Agent: An organization, such as a government agency, university, or trade organization, that invests in advancing manufacturing competitiveness.
- Data Cube: A conceptual model of manufacturing cost data representing the total of all costs, where individual smaller "cubes" represent detailed cost items (see Section 2.1).
- Economic Rate of Return: the rate of return on an investment or project that considers both financial and broader economic costs and benefits, reflecting its overall societal and economic impact. For more information see Boardman et al. (2018).
- Economic Impact: The measurable change outcomes attributable to a project, standard, or technology, expressed in monetary terms—typically, benefit cost analysis, net present value, economic rate of return, or internal rate of return—and realized through its adoption and use.
- Internal Rate of Return (IRR): A financial metric used to estimate the profitability of potential investments. For more information see Thomas (2017).
- Net Present Value (NPV): A financial calculation used to determine the total current value of an investment over a set period (e.g., 10 years), used to identify the highest performing investments. For more information see Thomas (2017).
- Opportunity Map: A dataset (e.g., the data cube) for revealing potential impact through extensible observations of manufacturing costs, used to guide the design and selection of future projects.
- Pareto Principle: (also called the 80/20 rule) states that a small share of causes often accounts for a large share of effects—classically, about 80% of outcomes come from 20% of inputs.
- Platform: A standardized framework or system developed by experts for use by non-specialists to facilitate data collection, economic decision-making, and hypothesis testing.
- Power Law Distribution: A power law distribution is a pattern where a few occurrences are extremely large or impactful, while most are small, and the probability of an outcome decreases as a power of its size. For more information see Section 2.
- Predicted Impact: A testable hypothesis or forecast of a project's future performance, which is compared against actual outcomes to improve future accuracy.
- Project: For the purposes of this report, a project is a grouping of change agent products and/or activities based on three criteria:
  - It influences the same industry or adoption community.
  - It uses the same mechanism to improve productivity/efficiency.

    - It groups related products and activities (e.g., a standard plus its supporting workshops and articles).
- Recalibration Factor: A factor used to systematically reduce prediction error over time through learning from previous project results. For more information see Section 2.2.
- Statistical Value of Life: an economic measure of how much people are willing to pay to reduce the risk of death, aggregated across individuals—not the value of any specific person's life.

Executive Summary

This report proposes an organization-wide system of continuous improvement for any organization trying to increase impact in advancing U.S. manufacturing competitiveness. The proposed system is applicable to other domains but is tailored to applied research in manufacturing. It is proposed that to maximize the impact of each dollar invested in research and development (R&D), change agents (i.e., those who invest in advancing manufacturing competitiveness such as governments or trade organizations) will likely need to transition from intuition-based selection to rigorous measurement science. By treating each R&D project as an experiment with a clear hypothesis (ex ante prediction) and measurable outcome (ex post estimate), organizations create a feedback loop generating actionable information that can be used to systematically improve future investment accuracy.

Three Functions of Economics in Manufacturing R&D

Economics serves three critical roles in this framework to drive realized impact:

- **Motivate**: Encourages manufacturers to or not to adopt new innovations by providing credible estimates of cost-effectiveness.
- **Guide**: Informs change agent R&D decisions regarding high-impact projects by evaluating potential costs and benefits before selection or design.
- **Justify**: Assesses whether agency investments generated sufficient economic impact by providing empirical evidence on realized outcomes and return on investment.

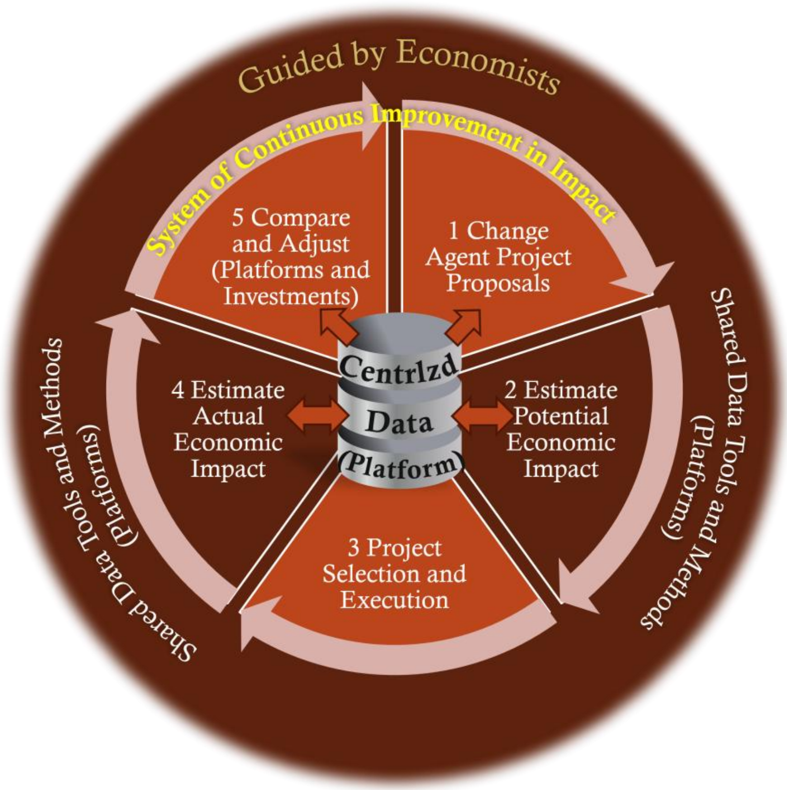


Fig. 0.1: Illustrated System of Continuous Improvement

Growth Engine for Impact

The proposed framework consists of three parts designed to move an organization from a "flat" distribution of impact toward a "power law" distribution, where returns are both higher and earlier per dollar of expenditure. While the focus of the framework is to ***guide*** change agents toward high impact projects, it also subsequently generates observations that ***motivate*** manufacturers and ***justify*** budgets; thus it largely achieves all three of the functions of manufacturing economics. The three parts of the framework include the following:

1. **Opportunity Map and Hypotheses**: Utilizing a "Data Cube" of manufacturing costs (stratified by industry, activity, firm size, and other factors) to identify where the greatest potential for impact exists and making hypotheses for impact. The hypotheses *guide* change agents in project design while the data can further be used to *motivate* manufacturers to adopt innovations.
2. **Validation and Recalibration**: Comparing predicted impacts against estimated actual results to calculate prediction error, which is then used to recalibrate and improve future forecasting models leading to designing and selecting higher impact projects. This process further refines the ability to accurately *guide* change agents toward high impact projects while the estimated impacts can be used to both *motivate* manufacturers to adopt innovations and ideally *justify* investments.
3. **Enterprise-Wide Utilization**: Implementing standardized platforms and shared data tools that allow the organization to perform high-accuracy analysis, harnessing economies of scale.

Without a systematic approach, change agents risk relying on traditions and individual biases, which often lead to inefficient outcomes. By adopting this measurement science approach, change agents can move toward the "strategic target" of maximum impact, ensuring that U.S. manufacturing remains globally competitive.

In its simplest form (i.e., tracking impact and hypothesis testing for impact), this framework essentially represents the minimum requirements for measurement science of impact and economics. As stated previously, it turns each project into an experiment where there is a hypothesis and a structured test, allowing learning from results to improve future predictions and impact. Without tracking or hypothesis testing, decisions are driven by intuition rather than evidence, increasing the likelihood of inefficient or ineffective outcomes. To reach the maximum impact per R&D dollar, change agents must apply measurement science for impact. The framework presented here lays out the fundamentals of that measurement science.

Implementing the framework requires navigating several logistical challenges, including balancing between ideal data/methods and those that are feasible to implement in practice. Thus, in addition to presenting the economic framework, this report delves into the logistics of developing standardized systems to implement it across an organization. It includes potential platforms for those who have limited decision-science expertise, which will be referred to as non-specialists, to make hypotheses and economic decisions; platforms for distributing economic data to non-specialists; and platforms for collecting data on economic investments to be used in guiding future investments/projects. The platforms for this method could be developed and guided by investment analysis experts (e.g., economists) but are potentially intended for experts and non-specialists (i.e., users), as the large majority of an organization consists of those who have not extensively studied decision science. The result is akin to a hub-and-spoke model (see Fig. 0.2) for data collection, dissemination, and analytical methods, where the hub maintains the data and standards platforms. The users are those entities (e.g., divisions, departments, or staff members) throughout an organization that apply the methods, data, and tools. The users pull data from the hub, feed data to the hub, and utilize the standards/methods maintained by the hub.

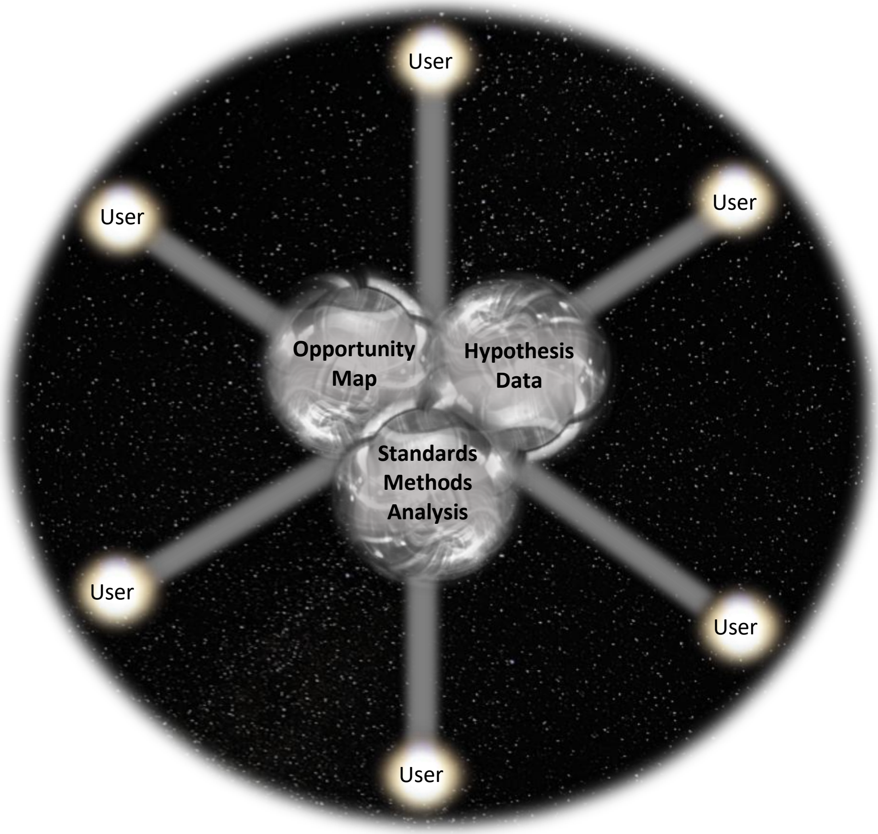


Fig. 0.2: Illustrated Hub-and-Spoke Model

1. Background: Competitiveness, Research & Development Investments, and Predictions

Change agents' (i.e., those who invest in advancing manufacturing competitiveness such as governments or trade organizations) R&D investments are central to national competitiveness, yet there is a challenge in creating comprehensive systematic planning for measurable impact. This pattern likely reflects structural characteristics of the work itself — including long time horizons, uncertainty, system-level coordination requirements, interdisciplinary impact planning, and difficulty in measuring spillover effects — which complicate strategic planning and evaluation. Because strategic planning depends on forecasting costs, benefits, and adoption, weaknesses in prediction directly undermine impact.

This section argues that improving economic forecasting and evaluation practices can lead to greater change agent impact. First, it discusses existing approaches for R&D programs (Section 1.1). Then, it provides the evidence and reasoning that motivate the framework discussed in the report: there is a challenge in systematically planning change agent R&D, forecasting can be biased and inconsistent, and outcomes may not be tracked—preventing organizations from learning from past investments (Section 1.2). Strategic planning relies heavily on predictions, but these predictions come with varying levels of accuracy (Section 1.3). There are methods available to improve prediction accuracy (Section 1.4), and real-world examples demonstrate how organizations have generated significant value by utilizing these approaches (Section 1.5). Additionally, patterns in R&D returns show that focusing on high-impact opportunities can accelerate and increase the overall impact (Section 1.6). Together, these insights highlight the need for an economic framework to guide decision-making in change agent R&D investments (Section 1.7).

1.1. Existing Programs and Frameworks

A recently discussed method for R&D management was presented to the Advanced Manufacturing Office (AMO) at the U.S. Department of Energy where an introspective performance assessment methodology with verification and validation of R&D Projects was discussed (Sandor 2019; Advanced Manufacturing Office 2019). This approach creates a standardized, transparent system for evaluating the effectiveness of federally funded manufacturing research projects. It proposes a two-tier assessment framework that tracks technical, economic, and energy-performance metrics and validates results.

The approach that was presented aims to improve accountability, measure progress, and improve decision making. Moreover, it has a feedback loop mechanism, but its stated objective is to, "Develop and codify a methodology, process and procedures (MP&P) to provide AMO a consistent, transparent and defensible accounting of anticipated benefits of currently funded technologies and supporting R&D projects." Thus, a primary focus of the AMO method is accountability and accuracy of measurement. In contrast, a focus of the framework in this report is to determine which project configurations have the highest return on investment and how do we get better at predicting that over time in order to select project components and design higher impact projects. Consider the following examples:

- If a change agent publishes a standard, does pairing it with a complementary report increase adoption?
- If a change agent develops a methodology, does creating a website around it increase its impact? Would a cluster of related websites be even more effective?
- When disseminating a standard or technology, is a change agent more successful in certain industries than others?
- Do innovations that reduce larger manufacturer costs have greater impact than those that affect smaller manufacturer costs?
- If a standard or complementary report is not being accessed, does that necessarily mean it is not being adopted? If so, are there interventions that could increase adoption?

The AMO system tends to be more of a measurement learning system while the framework in this report uses measurement and prediction to optimize economic impact. Thus, evaluation of predictive accuracy is a core component of this report's framework, including tracking forecast error and calibration of future predictions, which will in turn help answer questions about the effect of various project components on impact.

Another evaluation program can be found at the Defense Advanced Research Projects Agency (DARPA) program. This program evaluates proposals through scientific and technical review that focuses on identifying the projects that are the most advantageous to the government, which can include economic or defense utility (DARPA 2024). The evaluation process appears to focus heavily on execution of technical objectives rather than rigorously showing economic impact. The DARPA program has a feedback loop that is programmatic and milestone-driven.

Another research laboratory, the Oak Ridge National Laboratory, is one of 10 Department of Energy (DOE) laboratories evaluated by DOE's Office of Science (U.S. Department of Energy 2026a). The 10 laboratories are managed by contractors and evaluated annually by DOE's Office of Science (SC) based on scientific, technological, managerial, and operational performance. It uses a scoring system structured around eight Performance Goals (U.S. Department of Energy 2026b):

- Mission Accomplishment (Delivery of S&T)
- Design, Construction and Operation of Research Facilities
- Science and Technology Project/Program Management
- Leadership and Stewardship of the Laboratory
- Integrated Environment, Safety and Health Protection
- Business Systems
- Facilities Maintenance and Infrastructure
- Security and Emergency Management

Many universities also have advanced methods for evaluating work, such as the University of Manchester's "Evaluation Cycle Framework" (University of Manchester 2026) or Stanford's Program Evaluation Metrics and Tools (Stanford 2026).

Evaluation programs often have a suite of things they achieve; however, the framework here is not intended to be a system for program review but rather a system for tracking and identifying the conditions and configurations that systematically produce higher or lower impact. It involves:

- Explicit prediction before project execution (ex-ante ROI estimation as a core requirement)
- Systematic tracking of prediction error (not just outcome reporting, but formal forecast accuracy measurement)
- Continuous recalibration of the prediction model based on past errors (the model itself improves over time)
- Portfolio-wide learning across projects (insights are aggregated across all projects, not kept at program or project level)
- Project “configuration-level” analysis (identifying which design choices or structures drive impact, not just whether a project worked)
- Use of evaluation as an optimization tool, not just an accountability or reporting mechanism

Though the system collects many of the items a program review might require, it is not the focus. Many programs have some of these components, but few have all of them. The framework presented here is designed to integrate the elements listed above with a focus on generating actionable information to support the design and selection of higher-impact projects.

1.2. Planning

Change agent activities often require complex planning, have goals that are inherently challenging, and face difficulties in measuring impact— all common challenges for advancing manufacturing competitiveness. These challenges are illustrated in evaluations of change agent programs. For instance, a commission to evaluate European legislation for strengthening semiconductor ecosystem found that the legislation, “lacked an analysis of any trade-offs involved, and possible alternative actions and their potential impact.” For another change agent program, a review of the National Institute of Standards and Technology (NIST) Engineering Laboratory (EL) by the National Academies for Sciences Engineering and Medicine stated that “there is room for improvement in two specific areas: strategic planning and impact” (Panel on Assessment of the National Institute of Standards and Technology Engineering Laboratory 2025). An evaluation of a National Institute of Health (NIH) program found that assessments were often ad hoc and relied on qualitative judgments, highlighting the need for built-in planning and systematic measurement to capture program impact (Institute of Medicine (US) Committee for Assessment of NIH Centers of Excellence Programs, Manning, F. J., McGeary, M., & Estabrook, R. 2004).

Importantly, these critiques should not be interpreted as indicating a complete lack of measurement systems or planning. Many change-agent organizations—including NIST, NIH, and similar institutions—use well-established metrics such as publications, standards development, technology transfer outputs, stakeholder usage, and other indicators of activity and adoption. Rather, the central issue is that these metrics are typically not integrated into a unified framework that systematically links intermediate outputs to downstream economic outcomes or benefit-cost impacts in a way that supports predictive decision-making and project configuration optimization. Without these characteristics, planning does not strategically map to impact.

The types of critiques discussed above collectively suggest a common challenge: although change-agent R&D investments can generate substantial impact and are supported by structured planning and measurement systems, they often lack integrated approaches for translating measured outputs into quantitatively grounded estimates of downstream economic impact (e.g., benefit-cost analysis) that are systematically used in planning and resource allocation. This gap can reduce the effectiveness of resource allocation and the ability to maximize long-term impact.

It is a recurring challenge rooted not only in management challenges (e.g., predicting and attributing diffuse long-term impacts or addressing misaligned incentives between individual staff objectives and organizational R&D impact goals) but also in the structural complexity and uncertainty inherent in advancing national competitiveness. This report addresses that challenge by focusing on the economics portion of the development and evaluation stages of R&D, technology development, and standards development that tend to be conducted by change agents.

If change agent R&D is to be more strategic, it will need to be able to reliably forecast and maximize the economic value of investments. Typically, R&D facilitates sustainable increases in long run per capita GDP through efficiency, productivity, and resource preservation. One might argue that for manufacturing, it primarily frees up or preserves resources either for the manufacturer, user, or another stakeholder.

There are a number of processes that often occur when research is successful in advancing the manufacturing industry, which can include the following:

- Development of R&D project proposals
- Proposal evaluations and selection
- R&D execution
- Implementation of findings and results
- Dissemination of findings and results across manufacturers

Generally, each of these processes involve different types of expertise, making the whole undertaking somewhat complex. The advancement of industry competitiveness resembles a mission like Apollo 11, which put the first human on the moon, not because of its symbolism, but because of its structural complexity: multiple interdependent stages (e.g., launch, lunar orbit, and lunar descent), collaboration of many specialized experts (e.g., structural engineers, programmers, and medicine specialists), and tightly coupled systems where failure at any stage (e.g., launch, flight, or lunar descent) undermines the mission (see Fig. 1.1) (NASA 2024). Reliably advancing U.S. manufacturing competitiveness is a similarly complex task, as it involves many thousands of establishments with millions of people producing many types of complex

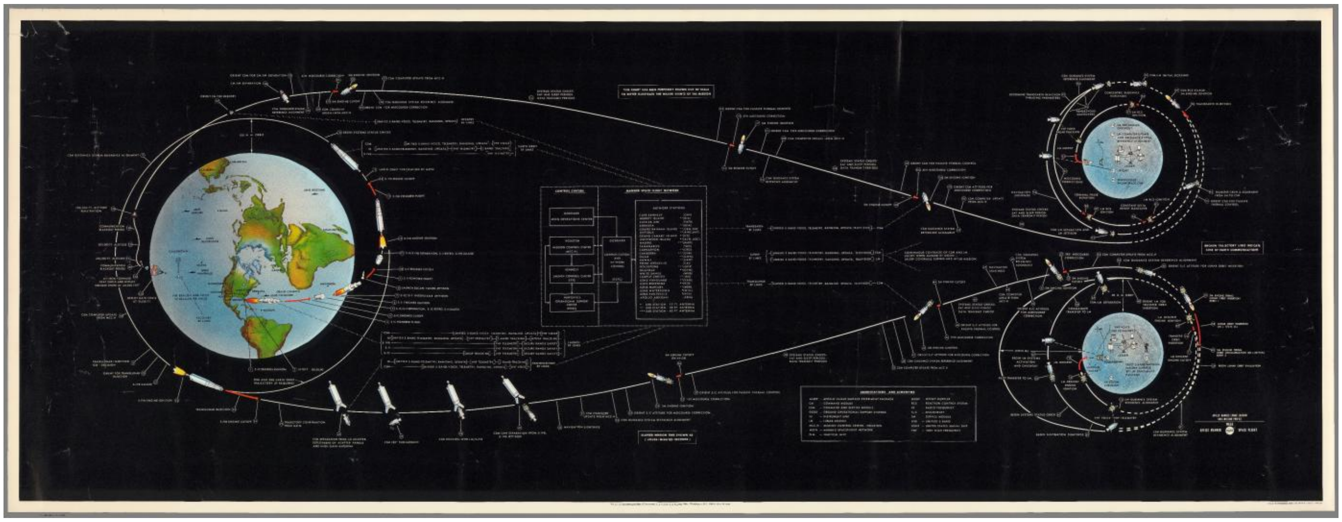


Fig. 1.1. Apollo manned lunar landing: Ground Operations Mission Support System Mission Profile (NASA 1969)

goods using complicated supply chains that often stretch around the globe. Additionally, both involve significant engineering challenges. Thus, manufacturing has many systems and, as discussed above, successful change agent R&D has many stages/processes.

Beyond these systems-level coordination challenges that parallel the Apollo mission, manufacturers themselves carry out a great deal of R&D; however, a significant portion (e.g., basic R&D or standards development) can be neglected due to the challenging nature of capturing the benefits (e.g., profit) of such work. For instance, standards often benefit multiple manufacturers, making it difficult for any particular manufacturer to capture the benefits of investing in them, resulting in fewer standards than is optimal. There are also many stakeholders and many types of stakeholders where one stakeholder’s benefit can be another stakeholder’s cost.

Additionally, there are incentives to conceal activities and processes, as revealing them can result in a loss of competitive advantage. These factors along with others can result in a complex system where inefficiencies are not easily observable. For instance, as Thomas (2022) points out in NIST AMS 100-48, it is difficult for manufacturers to differentiate products by their expected lifespan, resulting in less incentive to make long lasting products. The result is products such as appliances that have shorter and shorter lifespans. This inefficiency affects both consumers and manufacturers, as consumers cannot accurately identify and purchase long lasting products that they want, and manufacturers struggle to benefit from producing such products. This environment creates a clear role for change agents to invest in R&D and standards that produce broad societal benefits that are otherwise underprovided by the private sector.

To make change agent R&D more strategic, decision makers will need to be able to forecast the economic value of investments. In other words, strategic planning requires accurate predictions of costs, benefits, and adoption—predictions that are currently uneven in quality and often underdeveloped.

1.3. We all make Predictions

Typically, every investment (e.g., R&D investments or projects) or purchase involves making a prediction. When companies or governments invest in research, when an individual purchases a car or house, and even when purchasing basic items such as food, there is typically a prediction being made. When the decision maker chooses whether to invest or purchase an item, they are predicting whether the benefits of the purchase/investment will exceed the costs (both financial and non-financial). Because strategic planning requires forecasting, every investment decision ultimately involves a prediction about future costs and benefits. Currently, many of these predictions are not quantified, explicit, or formally stated. To maximize impact per dollar in change agent manufacturing R&D, decision makers will likely need to make formalized predictions.

There are a number of approaches to making predictions, which are each associated with different levels of accuracy and different levels of effort/cost to develop. An example to illustrate this tradeoff can be found in trying to predict how many windows there are in Seattle (see Fig. 1.2). To estimate this, one might take a qualitative approach, saying that Seattle is a large city; thus, there are many windows, or one might guess there are millions of windows (see Fig. 1.2). This approximation has a large level of error, but it did not require a high level of resources or cost to make the prediction.

To increase the accuracy of a prediction, one might use a quantitative method such as breaking the prediction into manageable subcomponents. One could, for instance, combine data about the population of Seattle with an approximation of the average number of windows each person is associated with. This approach reduces error; however, it requires more effort and cost to make the prediction. Finally, if one wanted a high accuracy estimate, a team of people could fly to Seattle and walk street by street counting all the windows. This would substantially reduce error but would significantly increase the cost of the prediction.

Predictions of future events can have similarly different levels of accuracy based on the level of investment in data gathering, data standardization, and model development. For instance, a weather prediction could be based on a qualitative evaluation such as looking at the sky, a quantitative method using local measurements (e.g., temperature, pressure, and humidity), or it can be a high-cost prediction based on models using satellite imagery. Higher levels of error in prediction result in higher risks of selecting or designing investments, projects, or products that have lower returns or benefits. One might ask what proportion of investment decisions for advancing manufacturing competitiveness are made without any quantified economic values, as these predictions are likely associated with lower levels of accuracy.

In order to improve the accuracy of predictions, an assumption of the proposed framework in this report is that at least some prediction errors are systematic rather than completely random. At first glance, it may seem like many manufacturing R&D investments are very heterogeneous. However, manufacturing R&D projects are typically focused on relatively narrow and measurable applications. For example, AI projects may target quality-control imaging or reductions in manufacturing changeover times. In these contexts, manufacturers' cost structures, operational constraints, and potential productivity gains can often be reasonably bounded and estimated. The manufacturing domain itself exhibits substantial structural persistence. At any given time, there is a finite set of industries, production activities, cost categories, and operational processes that public manufacturing R&D programs can realistically target for efficiency/productivity improvements. In addition, manufacturing capital assets, production systems, and organizational methods frequently remain in use for decades. Further, as discussed below, prediction errors associated with human judgment are often systematic rather than random. Moreover, given the extensive research on errors in human judgment, along with the structural stability of many manufacturing activities, it is likely that at least some prediction errors are systematic.

How many windows are there in Seattle?

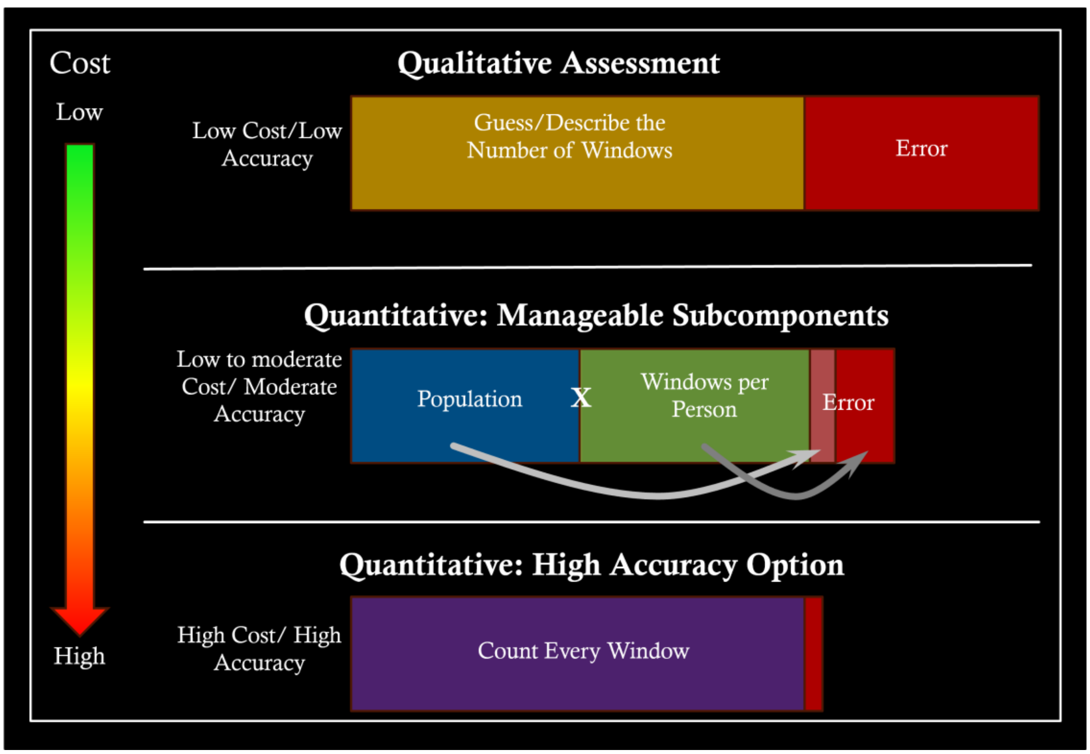


Fig. 1.2. Illustration of Prediction Accuracy

An additional consideration is that if every purchase and every project involves making a prediction, as argued earlier, then one challenge in improving prediction accuracy is the sheer volume of economic questions that need to be addressed. These questions include predicting the impact of both the projects that are to be executed and the alternatives that were ultimately not chosen; tracking impact progress and estimating realized impact for all projects;

and studying the error in predictions in order to improve prediction methodology. Predicting the impact of alternative projects alone is an almost endless effort, given the vast array of possibilities. The consequence is that there are more questions than there are economists or data available to answer them quickly and accurately enough for timely decision-making. While advancements in data collection and analysis tools help, the gap between the demand for rapid predictions and the availability of resources to produce them remains a significant challenge. Thus, the question becomes whether it is possible to make a step increase in prediction accuracy such that it increases impact/returns by selecting or designing more impactful projects. Additionally, the method used might need to be useable by those who do not have decision science expertise.

To illustrate this concept in manufacturing, one might consider a proposed project to develop a material standard for additive manufacturing (AM). To assess the impact of such a project, one could have a low-cost/low-accuracy impact assessment; moderate-cost/moderate-accuracy impact assessment; or a high-cost/high-accuracy assessment. The low-cost/low-accuracy assessment might include a qualitative narrative describing the impact that discusses the reduction in costs for validating part performance of AM products and increasing additive manufacturing adoption that results in lowered costs for customized and/or lightweight products. A moderate-cost/moderate-accuracy impact assessment might include making a quantitative assessment that breaks the impact prediction/hypothesis into manageable subcomponents. In this instance, the prediction might be broken into the number of adopters and average impact per adopter. Publicly available industry level data along with additive manufacturing reports could be utilized to estimate cost savings and potential adopters. Finally, a high-cost/high-accuracy assessment might include using industry data, manufacturing reports, sending out surveys, holding workshops, and contacting manufacturers to collect the data needed to more precisely estimate the potential impact of the AM standard.

1.4. Predicting with Accuracy

As discussed above, reliably identifying high return industry research investments often requires making forecasts and investment decisions with incomplete and imperfect information. Consequently, some forecasts can have greater uncertainty and might even be considered back-of-the-envelope estimates. Despite these shortcomings, it has been shown that some methods for making forecasts have greater accuracy than others. The Intelligence Advanced Research Projects Activity (IARPA), part of the Office of the Director of National Intelligence, set out to improve American intelligence, which often includes making forecasts such as estimating the probability of one country attacking another. To determine and improve the performance of their intelligence forecasts they created a forecasting tournament (Tetlock and Gardner 2015). This competition created a great deal of information on what works and how well. In conclusion, some methods are more accurate than others, even to the point of outperforming professional intelligence analysts with access to classified information (Tetlock and Gardner 2015). Unfortunately, these methods are frequently underutilized due to bias toward one's own individual insight and other such challenges. Thus, creating predictions with high levels of accuracy not only presents a challenge in developing rigorous data and methods for both decision science experts and non-specialists, but it can also involve overcoming psychological obstacles as well.

Even when better methods exist, humans often default to intuition, which reduces forecast accuracy. There can be a temptation or even a tendency for decision makers to use their instinct or intuition to determine their investments or projects. To some extent this should be resisted, as humans are vulnerable to being heavily influenced by immaterial feelings and emotions (Lewis 2004; Kahneman 2011; Ariely 2008). For instance, a project/investment might seem more appealing to a decision maker because of familiarity with the principal investigator, or maybe because they contributed to the idea. Research has shown that even having heard uninformative numbers can influence our judgement (Wilson et al. 1996). Many economic researchers have investigated these phenomena, including Daniel Kahneman, a Nobel Prize winner in economics. Despite one's best efforts, it has been shown that humans are not able to fully separate emotions from logical decision making (Kahneman 2011; Ariely 2008).

When intuition is broad and fuzzy it is more vulnerable to being based on unsound reasoning. *If it is necessary to use intuition, it should apply to assessing specific factors of an investment and be grounded on informed judgement*. For instance, if a particular cost of an investment is unknown (e.g., cost of energy), one might use individual insight to estimate the value of this individual cost. The unknown cost should not lead to or justify using intuition to evaluate the full merits of the project because one cost is unknown. When conducting an investment analysis of potential industry R&D projects, there are frequently many unknown values. These values need to be forecasted or predicted for an analysis to be completed. In the literature on forecasting, including that of Tetlock and Gardner (2015), some common themes arise:

Problem Clarification: It is important to ensure that relevant questions are being asked. Frequently, pursuits of information can be sidetracked by topics that are only adjacent to the true concerns at hand. The purpose of the forecasts discussed here is to identify the highest performing change agent investments to advance manufacturing competitiveness, given the constraints and resources available. For instance, the belief regarding the importance of a technological advancement might be irrelevant as to the scale of its economic impact compared to other change agent investments, as it may or may not actually translate into large returns, may only apply to a small number of firms, may not be viable, or it requires overwhelming sums of investment.

Extraneous Information and Overweighting: When forecasting, it is important not to overweight particular types of information (Tetlock and Gardner 2015). Sometimes too much can be read into a piece of evidence or relevant information. For instance, a piece of evidence may only suggest something is possible rather than it being probable. In manufacturing competitiveness, artificial intelligence (AI) provides a useful example. AI is changing a great deal of the industry landscape, making it seem as though AI project impact has limitless boundaries; however, manufacturing competitiveness projects are often more narrowly bounded. For instance, the application of AI in quality control imaging operates within a constrained evaluation boundary. There are only so many detectable defects and there is only a finite range of achievable cost reductions. As a result, the potential gains are valuable but are not unlimited. This example illustrates that change-agent applied R&D is primarily a process of engineering and economic optimization within bounded operational systems, rather than a speculative portfolio-selection problem.

Another example of overweighting is the tendency to overweight the importance of a solution being the most technologically advanced. Manufacturing industry change agents can fall into an assumption that technological sophistication equates to economic impact potential. The Palm Pilot, which held 70 % of the personal digital assistant (PDA) market in 1999 (Wiggins 2004), represents a historical example of the inadequacies of this assumption. At the time, PDA manufacturers were often focused on advanced complex technology; however, the Palm Pilot succeeded not by delivering the most advanced technology, but by effectively solving common user problems. People had important personal information scattered across paper planners, sticky notes, Rolodexes, and desktop computers — but no easy way to carry, update, and synchronize it all. The Palm Pilot gave people a pocket-sized place to synchronize and backup this information in a simple and easy way. Its success demonstrates that economic impact is often driven more by usability, constraint alignment, and effective problem solving than by technological sophistication alone.

Overconfidence: A common challenge that is faced when forecasting is overestimating one's own knowledge and abilities. For instance, in a survey of 161 students, 93 % estimated that they were above the median in driving ability, which is not possible. One might weigh their own evaluations/methods more than others simply because it is their own method/idea.

Manageable Sub-Problems: The values to be forecasted can be broken into manageable sub-problems to obtain a more accurate estimate. To illustrate, consider a survey that asks someone to estimate the hours per year they spend driving their car compared to one that asks each component of their driving time (e.g., number of hours per day they spend driving to and from work). An aggregated question such as one on the total hours per year they spend driving is difficult to answer, as they must consider all at once the different places that they drive. Someone is much more likely to estimate with accuracy the amount of time they spend driving to work and other individual components of their total driving. To the extent possible, break large problems into smaller sub-problems, utilize known values, and utilize comparisons to other similar circumstances.

Expert Opinion: In some situations, expert opinion can be quite useful; however, there are some limitations that are often overlooked. Expert opinion provides increased accuracy in a forecast when the subject is within the expert's domain of expertise; thus, it is important to identify the boundaries of one's knowledge. For instance, a physicist can provide a great deal of expertise in physics, but that may not translate into understanding the economic and social benefits that result from advancements in physics. Additionally, experts can perform poorly when there is limited feedback; that is, when their predictions are rarely confirmed or refuted (staal et al 2024; Bolger and Önkal-Atay 2004).

Testable Predictions: Forecast accuracy is improved in the presence of feedback—that is, when predictions can be empirically evaluated and compared against observed outcomes. This implies that forecasts should be specified in operationally measurable terms.

For instance, predictions that a change agent investment will have “large” or “important” benefits are not well-defined predictive statements, because they do not specify a measurable quantity or threshold that allows consistent ex-ante prediction and ex-post evaluation. While such statements may be interpreted or assessed retrospectively, their evaluation is not standardized and can vary across evaluators and contexts.

In contrast, a forecast that specifies a range of measurable outcomes—such as cost savings, productivity improvements, adoption rates, or other quantifiable impacts—constitutes a testable prediction. At a later point in time, observed outcomes can be compared against predicted ranges to assess forecast accuracy and improve prediction methods.

Accuracy in Numbers: It is well established that forecast accuracy is increased by including more people’s input into the forecast (Tetlock and Gardner 2015) . Therefore, it is beneficial to have a team of people or to survey individuals in order to estimate and answer manageable sub-problems. These lessons in forecasting suggest that change agent R&D decision-making can improve by adopting structured prediction methods and feedback loops.

Learn by Doing: Increasing prediction accuracy comes from practitioner experience. Just as one cannot learn to ride a bicycle from reading a physics book, one cannot generate high accuracy predictions without generating and testing predictions (Tetlock and Gardner 2015).

1.5. Real World Application of Increasing Forecast Accuracy

The benefits of improved forecasting can be dramatic. The following real-world examples illustrate how better prediction methods can reduce cost and increase performance. The value of improved forecasting accuracy can be illustrated by three organizations: Oakland A’s, Toyota, and UPS.

Oakland A’s: In the book “Moneyball: The Art of Winning an Unfair Game,” Michael Lewis writes about how professional baseball had become a game where the richest teams win; however, one team was able to take a small group of undervalued professional baseball players and executives, who were essentially “rejected as unfit for the big leagues” and create a highly competitive team (Lewis 2004). Manager Billy Beane and his staff at the Oakland A’s achieved this feat with a “willingness to rethink baseball: how it is managed, how it is played, who is best suited to play it, and why… [It is] what amounted to a systematic scientific investigation of their sport” (Lewis 2004). They found bargains by identifying prejudice in baseball’s traditions. Other teams were paying on average $3 million per win while the Oakland A’s were able to reduce that cost to just $675 000 or 22.5 % of their competitors cost.

The Oakland A’s reduced the cost of winning a baseball game by 77.5 % by increasing their forecast/prediction accuracy. Their methods more accurately predicted the number of wins the team would have with different players/strategies, allowing them to more reliably select higher return investments/strategies/players. It is important to note that both the new approach and old approach use data. For instance, in the old approach scouts would consult batting average, home runs, or runs batted in, but the new method used the data more effectively. It is also important to note that the old approach yielded positive results, as those approaches tended to result in winning more games. However, the Oakland A's approach to predicting outcomes was more accurate, resulting in an increased ability to accurately identify high returns.

If manufacturing industry change agents were able to reduce the average cost of research impact by a similar percentage as the Oakland A's experienced in baseball, researchers might increase their impact by 444 % - because a 77.5% reduction in cost means the same budget could produce 4.44 times as much impact assuming a constant cost of impact. This might be achieved by increasing the accuracy of impact predictions for projects, allowing change agents to identify projects that are both affordable and highly impactful.

There are some additional insights that might be gleaned from the MoneyBall example. Billy Beane along with the assistance of Paul Depodesta, a Harvard graduate in economics, are largely credited with the Oakland A's feat. According to Lewis, Depodesta saw a number of issues in baseball:

> "There was, for starters, the tendency of everyone who actually played the game to generalize wildly from his own experience. People always thought their own experience was typical when it wasn't. There was also a tendency to be overly influenced by a guy's most recent performance: what he did last was not necessarily what he would do next. Thirdly—but not lastly—there was the bias toward what people saw with their own eyes or thought they had seen. The human mind played tricks on itself when it relied exclusively on what it saw, and every trick it played was a financial opportunity for someone who saw through the illusion to the reality. There was a lot you couldn't see when you watched a baseball game" (Lewis 2004).

The observation that "the human mind played tricks on itself" is aligned with some of the items discussed in Section 1.4, including overconfidence in one's knowledge, the effect of extraneous observations, and being sidetracked by information that is only adjacent to the issue at hand. Additionally, Billy Beane's and Paul DePodesta's new scientific approach created significant tension/conflict with the scouts for the Oakland A's, who embraced the previous approaches and strategies. In the end, the new approach increased the accuracy of predictions, resulting in a more competitive team at a lower cost, but the change came with controversy, which might be common in shifting prediction methods.

Current approaches for project design and selection within change agents often rely on intuition and individual insight, similar to the scouts at the Oakland A's. This makes them vulnerable to the same issue where the human mind plays tricks on itself. A more rigorous system of impact forecasting can reduce this effect in the same way that it did for the Oakland A's and it is largely consistent with more formal observations of increasing forecasting accuracy discussed in Section 1.4.

Toyota's Total Production System: Another example of where increased accuracy in predictions resulted in high returns is in Toyota's Total Production System (Liker 2004). Toyota's system increases returns by improving prediction accuracy through standardized work, real-time feedback, and pull-based production, which reduce variability and uncertainty in demand and processes, enabling more efficient use of resources and lower costs. These improvements in operational predictability allow Toyota to better match production with actual demand, reducing excess inventory and minimizing downtime. By systematically anticipating potential bottlenecks and variations in the production process, the company can focus resources on more critical efforts such as increasing both productivity and overall value generated per unit of input. This example illustrates that structured forecasting and data-driven decision-making can generate substantial returns in complex operations—paralleling the potential benefits for change agent R&D investments.

UPS' Orion system: UPS' Orion system is yet another example of generating impact from increased prediction accuracy (Team Ascend 2025). ORION increases returns by improving prediction accuracy at scale—using standardized operations, real-time data, and continuous learning to reduce uncertainty in routing decisions, which lowers costs and increases productivity. By continuously analyzing routing data and adjusting plans in real time, ORION reduces uncertainty and improves the efficiency of every delivery. The system's predictive algorithms allow UPS to allocate vehicles and personnel more effectively, minimizing wasted miles and fuel while maximizing on-time performance. This demonstrates how accurate forecasting can scale operational improvements across a complex network—an approach that parallels how change agent R&D could benefit from better prediction of project costs, adoption, and impact.

Although R&D does not equate to baseball, production lines, or package delivery systems, the examples of Moneyball, Toyota, and UPS are intended to illustrate a narrower principle: that complex systems with interdependent components can benefit from structured feedback, measurement, and iterative updating of decision models under conditions of partial control and uncertainty.

In these systems, decision-makers do not control all relevant variables. For example, Toyota does not control consumer demand, UPS does not control traffic conditions, and the Oakland A's do not control player market prices. Instead, they operate by continuously updating decisions based on observed outcomes and feedback signals. Similarly, in applied R&D contexts, while outcomes are more uncertain and attribution is noisier than in engineered systems, decision points still exist where resources are allocated, and projects are selected. At these points, feedback from observed performance can inform future decisions, even if the mapping between actions and outcomes is imperfect and evolving.

Attribution noise can increase the marginal value of improved predictive systems by raising the importance of signal extraction, measurement design, and iterative model improvement, although this value is subject to cost and diminishing returns. Prediction value for R&D costs/benefits is likely highest in settings where decision stakes are high (e.g., national defense, industrial competitiveness, or high-leverage investments such as artificial intelligence, additive manufacturing, or predictive maintenance), expenditures are large, and small improvements in allocation efficiency can compound over time across a portfolio of projects. In these cases, even modest improvements in forecasting accuracy or project selection can yield disproportionately large downstream impacts.

1.6. System Level Experiential Advancement

If improved prediction methods can generate value in sports and logistics, the same principle should apply to change agent manufacturing R&D—provided outcomes are defined, measured, and analyzed. Much of the framework presented in this report is to facilitate system level experiential advancement. Simply put, the framework enables us to learn from the past. Without tracking hypotheses and outcomes, there is limited ability to study the investments and projects that have been executed. As a consequence, decision makers can struggle to systematically identify all of the characteristics of high impact/return investments.

To illustrate the types of data that can guide future investments, consider the Department of Energy's Industrial Training and Assessment Center (ITAC) Program, which makes recommendations for efficiency investments in manufacturing. Fig. 1.3 graphs the estimated net present value calculated using data from the ITAC's Assessment Recommendation Codes (ARC), a classification system for recommendations (e.g., recommendations for maintenance). It reveals that there is a pattern in the savings such that 20 % of the recommendation codes represent 80 % of the benefits (i.e., net present value), consistent with the Pareto Principle, which is based on a power law distribution; that is, that a small portion of the opportunities produces most of the value. This finding reveals that impact can likely be both increased and realized sooner by targeting technical solutions with high impact. That is, this data confirms that increasing the accuracy of change agent forecasts can result in identifying and prioritizing higher impact projects, as some projects account for more impact than others. If this distribution did not exist, then the impact of one project might have little difference from another, resulting in no benefits from increased accuracy in predictions.

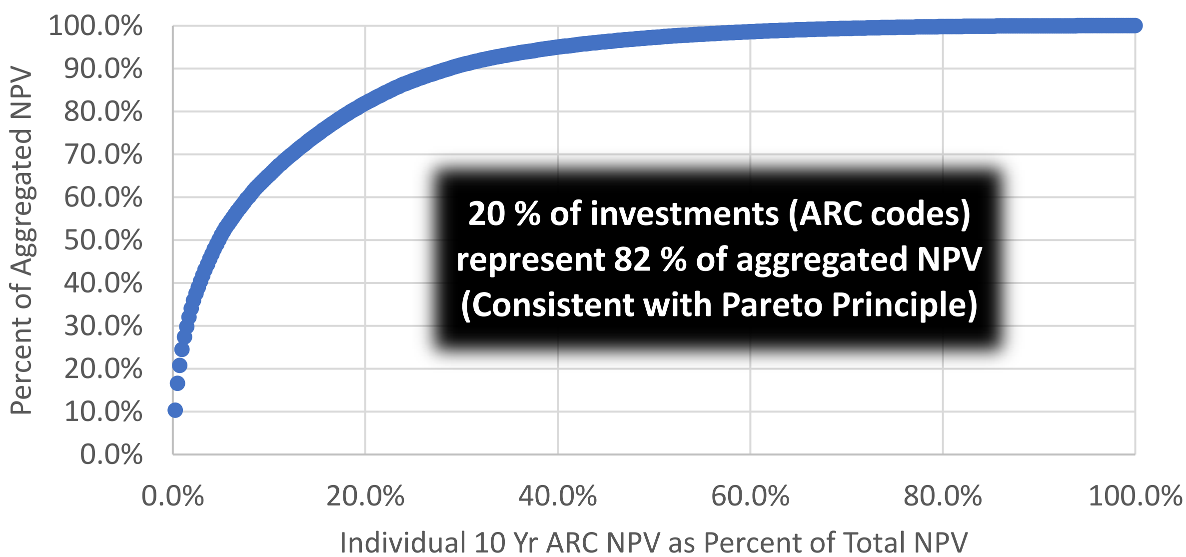


Fig. 1.3. Cumulative Net Present Value by Percent of ARC Cost Categories (Industrial Training and Assessment Centers 2026)

An example that illustrates how data could guide R&D might be found in data from the ITAC ARC codes on biopharmaceuticals (i.e., NAICS 325411 and 325412: Medicinal, botanical, and pharmaceutical preparation manufacturing). Using the ITAC data one can find that 61 % of facilities can benefit from using higher efficiency lighting with the investment having an average 10-year net present value of $79 000 per establishment and a 292 % internal rate of return. Further, one might see that 43 % of facilities can benefit from eliminating leaks in inert gas and compressed lines. Investments in this area could be expected to yield a 10-year net present value of $93 000 per establishment. This information can be used by other ITAC evaluations to streamline their assessments. The estimates can be made because the ITAC program has platforms for estimating costs and benefits along with those for tracking, collecting, and sharing data.

When combined with an ex post estimate of impact, the information gained can increase the accuracy of predictions, resulting in identifying project characteristics that result in higher impact. Other items might be tracked as well to identify characteristics of high return. For instance, investments in train-the-trainer programs can potentially have greater effect than those that directly train students. Projects that might result in disseminating information or spreading like contagion among many manufacturers might have greater impact than those that only affect a single manufacturer. These patterns suggest that tracking project characteristics could reveal what types of interventions spread most effectively.

1.7. Approach for Economic Investment Analysis and Design for Manufacturing R&D Projects

Together, these insights suggest a need for a systematic, data-driven framework that improves forecasting and learning over time. To do this, this report proposes an organization-wide (or potentially multi-organization) system of continuous improvement to increase impact in advancing U.S. manufacturing competitiveness. To maximize the impact of each dollar invested in research and development (R&D), change agents will likely need to apply rigorous measurement science. The framework presented here defines the core principles of that science, treating each project as an experiment with a clear hypothesis and measurable outcomes. It can be broken into three parts: opportunity map and hypotheses; validation and recalibration; and enterprise-wide utilization. Together, these three items largely achieve the previously discussed functions of economics in manufacturing R&D: *motivate* manufacturers to adopt innovations; *guide* change agents to high impact projects; and *justify* budgets. The result of the framework is a system of continuous improvement where standardized data is collected and shared across the organization.

The core of the framework and the logic behind it is presented in Section 2. The logistics of achieving the framework includes developing platforms or standardized means for making economic observations, assessing impact, and collecting data, as discussed in Section 3. This series of standardized frameworks includes those for: (1) non-specialists (i.e., those who do not have expertise in decision science) to make economic decisions, (2) disbursing economic data to non-specialists, and (3) for collecting data on economic investments to be used in guiding future investments. The approach further proposes comparing predicted economic impacts to impacts measured after the implementation of a project to adjust or refine methods for economic decision making. The platforms for this method would need to be developed and guided by investment analysis experts (e.g., economists), but are largely meant for non-specialists. Standardized platforms facilitate clarifying and refining the prediction question along with reducing the effect of extraneous information or predisposed beliefs of the decision maker. It can also implement methods that increase accuracy such as breaking the prediction into manageable subcomponents, utilizing expert insight, and accuracy in numbers. Finally, the proposed approach creates testable predictions allowing for improvement in accuracy.

In summary, change agent R&D faces a systemic challenge: without strategic planning and accurate forecasts, investments risk producing limited impact. Forecasting research shows that structured methods outperform intuition, and real-world examples demonstrate the economic value of improved prediction. Finally, historical data can reveal patterns of high-impact investments. Together, these insights motivate the framework presented in Section 2.

2. An Economic Framework for Impact

Economics plays a central role in generating impact within change agent organizations (e.g., federal agencies, universities, trade organizations) through three primary mechanisms (see Table 2.1). In this context, economics is used as a decision-support tool that complements engineering, technical, and domain-specific expertise. First, economics **motivates** manufacturers to adopt innovations by serving as a tool for estimating cost effectiveness. The second is that economics **guides** change agent R&D to high-impact projects by evaluating the costs and benefits for each potential project. These first two mechanisms highlight the core role of economics: generating actionable information for decision-making. The guiding mechanism—the focus of this report—illustrates a hierarchy of influence: change agents drive innovation in manufacturing while economists drive innovation within the change agents themselves, often by critically evaluating priorities and assumptions. The last mechanism for economics is that it **justifies** change agent R&D budgets by evaluating the impact that occurred. Together these three items — motivate, guide, and justify — encompass the mechanisms for how impact is realized through economics.

Economic Functions	Target Audience	Affected Actions
Motivate: It encourages manufacturers to adopt new innovations by providing credible estimates of cost-effectiveness.	Manufacturers	Adoption of Cost-Effective Innovations
Guide: It directs change agent R&D resources toward high-impact projects by evaluating potential costs and benefits before selection.	Change Agents	Design and Selection of High Impact Projects
Justify: It validates agency budgets by providing empirical evidence of the economic impact achieved.	Decision Makers	Funding Decisions

Table 2.1: Three Functions of Economics in Manufacturing R&D

The framework proposed below has three parts: 1) Opportunity Mapping and Hypotheses Generation; 2) Validation and Recalibration; and 3) Enterprise-wide Utilization. The data and methods in the framework facilitate the three primary mechanisms (Motivate, Guide, and Justify). Thus, it achieves multiple aspects of how economics achieves impact. In its simplest form, this framework is essentially the minimum requirements for measurement science for impact. It turns each project into an experiment where there is a hypothesis and a structured test, allowing learning from results to improve future predictions and impact.

The foundation of this framework—hypothesis testing and impact tracking—is essential to realizing a change agent’s full potential. To illustrate, imagine an alternate history in which, a century ago, efforts to maximize automobile fuel efficiency proceeded without systematic measurement or hypothesis testing. Factors such as transmission gear ratios, aerodynamics, and operating temperatures and pressures would go unexamined, making today's levels of fuel efficiency virtually unattainable.

A similar argument applies to other domains: maximizing crop yields, reducing disease transmission, or achieving successful space flight all depend on systematically testing hypotheses and tracking outcomes. Without such evidence-driven approaches, decisions are guided by intuition rather than data, increasing the likelihood of inefficient or ineffective results. Moreover, realizing maximum performance—whether in fuel efficiency, agriculture, healthcare, space exploration, or change agent impact—requires applying the principles of measurement science: careful tracking and rigorous hypothesis testing.

The benefits of hypothesis testing for change agent impact rely on the idea that some change agent investments have a higher return per expenditure dollar than others. The distribution of potential impact likely follows some power law distribution, such as that of the Pareto Principle, as illustrated in Fig. 2.1 and as evidenced by data previously discussed in Fig. 1.3. On the x-axis of Fig. 2.1 are projects ranked from highest impact to lowest. On the y-axis there are potential impacts either in dollars or return on investment. If there is no knowledge about which projects have a higher impact, then the distribution of impact approaches a flatter distribution, as change agents do not know which project to invest resources nor do they know which ones require more resources than others. Alternatively, in a world where we approach perfect

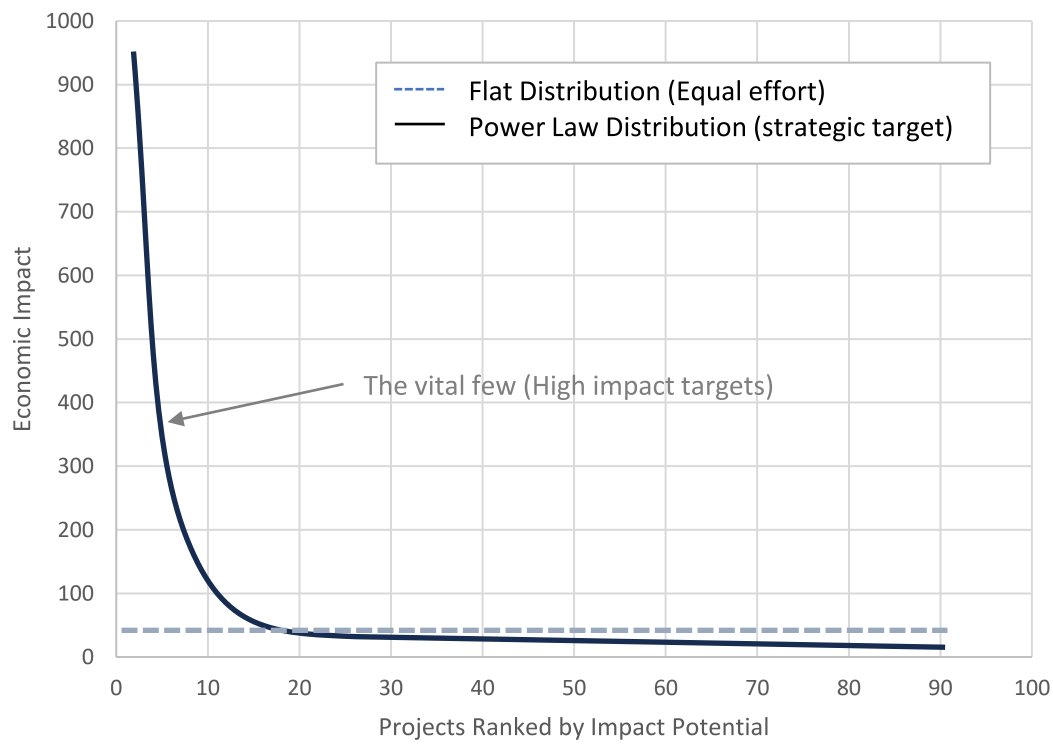


Fig. 2.1: Change agent R&D Investment Strategy: Flat vs. Power Law Distribution

information about potential impact, the distribution of impact becomes closer to a power law distribution where impacts are both higher and earlier per dollar of R&D expenditure. To maximize impact per dollar of expenditure, change agents likely need to move further away from the flat distribution toward the power law distribution to capture high return projects by increasing the accuracy of impact predictions.

The impact of a change agent organization for manufacturing can be calculated as shown in Equation 1:

Equation 1

$$Impact \approx \sum_{j=1}^{J} r_j * c_j$$

Where:

r_j = percent reduction in cost (manufacturing or user cost) resulting from project j

c_j = manufacturer and/or user cost associated with project j

Note that impact is evaluated over a fixed time horizon using standard discounting methods. Equation 1 is primarily for conceptual purposes. To move from the flat distribution in Fig. 2.1 toward the power law distribution, one needs increasingly more accurate predictions of impact, which implies the need for more accurate predictions of r_j and c_j for each project j. Facilitating these predictions is discussed in the framework below.

2.1. Part One: Opportunity Mapping and Hypotheses Generation

The first part of the framework is to develop an opportunity map, which consists of data characterizing potential impacts. Cost data on the manufacturing industry can be thought of as a cube of cubes (see illustration in Fig. 2.2). All the little cubes together represent the total of all manufacturing costs/losses. Each little cube represents an individual cost with particular characteristics such as the industry, firm size, and other factors (not shown). It is important to note that the framework presented here is intended for applied R&D in manufacturing, making the data cube an invaluable resource. As one moves closer to basic R&D, the uncertainty of benefits increases. There may be an application in basic R&D for the framework and data cube; however, it is currently tailored for use in applied R&D.

As illustrated in Fig. 2.2, the granularity of the data cube can be increased over time providing increasing ability to accurately predict the costs that manufacturers face related to a particular project proposal (i.e., c_j from Equation 1). Note that in this context, projects should include items that achieve impact through similar mechanisms and pathways (e.g., a standard and guides for adopting the standard); thus, they may not coincide with organizational project definitions. Alternatively, one might need to break a project into groupings of products. The cost data represented by the cube can be used to increase the accuracy of predicting the manufacturing (or user) cost(s) relevant for an R&D project (i.e., c_j in Equation 1) that **guides**

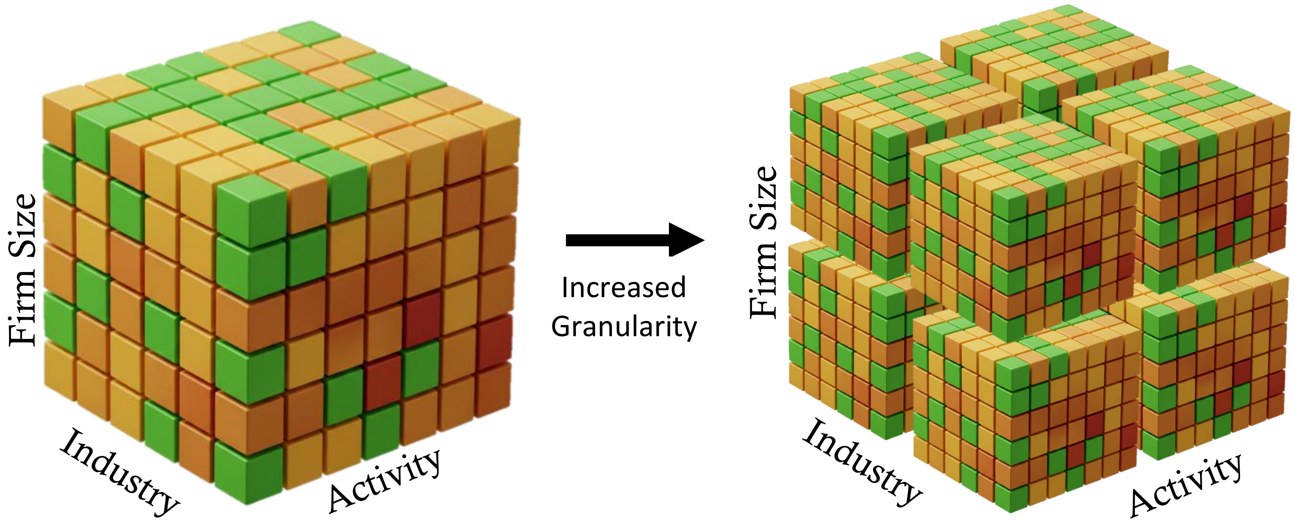


Fig. 2.2: Data Cube Illustration of Manufacturing Costs

project design and selection. The data can also be used to make a rough prediction for manufacturers' return on investment for **motivating** the adoption of innovations. Note that a change agent does not need to necessarily fill the entire cube, just those items needed for prediction.

The granularity, factors of the data (e.g., industry and firm size), and accuracy must be strategically selected and managed to maximize impact that results from prediction accuracy while managing database costs. Also, for the sake of efficiency and effectiveness, the data cube needs to be extensible. The extensibility might be compared to the map that geologists would use to locate oil reserves. Each time that they go to identify the next location to drill, they refer back to the existing maps and build off of them rather than starting completely anew. The resulting data cube for manufacturing costs represents an extensible set of observations made before making a hypothesis about impact for a proposed project.

Industry-level estimates of c_j such as that found in the data cube are useful for assessing potential scaled impact and prioritizing projects with high leverage, but realized impact is ideally measured using firm-level c_j for participating manufacturers. Industry-level c_j is therefore used for portfolio-level forecasting and scenario analysis, while firm-level c_j supports evaluation, calibration, and learning.

In addition to predicting c_j, change agent organizations will want to aim for accurate predictions of r_j. Predicted percent cost reduction is estimated using bottom-up, mechanism-based models informed by reference classes, early pilot evidence, and uncertainty bounds. Estimates are expressed probabilistically and updated over time, with evaluation emphasizing calibration and learning rather than point accuracy. The reduction in cost comes from manufacturing process efficiencies, input substitution, capital utilization, labor productivity, and quality/reliability. Estimates of reduction might map these categories where each one is associated with a range of potential reductions.

2.2. Part Two: Validation and Recalibration

After a project or selection of products are completed, the resulting impact needs to be evaluated for some portion of them. The impact that is measured can be used to **justify** budgets, evaluate prediction accuracy, and evaluate the effectiveness of different aspects of a project/product (e.g., diffusion efforts). To evaluate prediction accuracy, the estimate of realized (ex post) impact is compared to the predicted (ex ante) impact from Part One described in Section 2.1. The result is the error in predictions:

Equation 2

$$E_t = \frac{1}{J}\sum_{j=1}^{J}\left|Impact_j - Pred\_Impact_j\right|$$

Where
E_t = Mean absolute error of projects 1 thru J.
$Impact_j$ = The estimated actual impact of project j
$Pred\_Impact_j$ = The predicted impact of project j.

Each round of projects will result in revealing an estimated level of error associated with the approach for predicting impact. As time goes on, the error can be systematically reduced through learning and eventually it might be minimized:

Equation 3

$$E_{t+1} = E_t - (E_t * RCL)$$

E_t = error in ROI estimation at time t
RCL = Recalibration factor reduction

The recalibration factor is the proportion of error that is reduced as a result of observing error in predictions and altering prediction methods to increase accuracy; thus, as time t increases, error is minimized. The result is improved prediction accuracy, enabling change agent organizations to better forecast impact and identify high-impact projects. This approach has been demonstrated to work even in non-fixed systems, such as in the previously discussed examples at UPS, the Oakland A’s, and Toyota. Note that when an organization is initially adopting this framework, it can use past data, successes, and failures to backtest impact prediction approaches.

2.3. Part Three: Enterprise-Wide Utilization

The last part of the framework is to build and utilize the opportunity map, the prediction and tracking of impact, and hypothesis testing organization wide. This can generally be seen as harnessing economies of scale and standardizing metrics. As more change agents collaborate, there is more data being shared moving away from episodic learning that happens in silos to enterprise-wide learning that compounds. This results in a sharing of the costs for collecting information and sharing knowledge, resulting in more knowledge per dollar of expenditure. This can start an engine of growth for impact due to the breadth and depth of effects on R&D projects, including:

- Cross-project effects
- Multi-process effects within each project
- Long-term continuous improvement year after year
- Double feedback loop effects (continuous validation) for
    - Prediction metrics
    - Project performance

Enterprise-wide utilization also standardizes economic metrics. There are many measures and estimates of economic activity where a great deal of it uses methods that essentially embellish the potential impact. An enterprise-wide opportunity map contributes to standardizing these measures, reducing the potential effect of embellishment.

Achieving enterprise-wide utilization is likely to require a centralized hub that maintains data (opportunity map and hypotheses data) and maintains the standard classifications and methods with the result being a hub-and-spoke model (see illustration in Fig. 2.3) where the spokes feed and extract data via the hub. The hub includes logistics management and a centralized economics office that develops and maintains methods, platforms (discussed in Section 3), data, data classifications, guidance, and conducts research to advance impact. The spokes are operating units that develop projects for

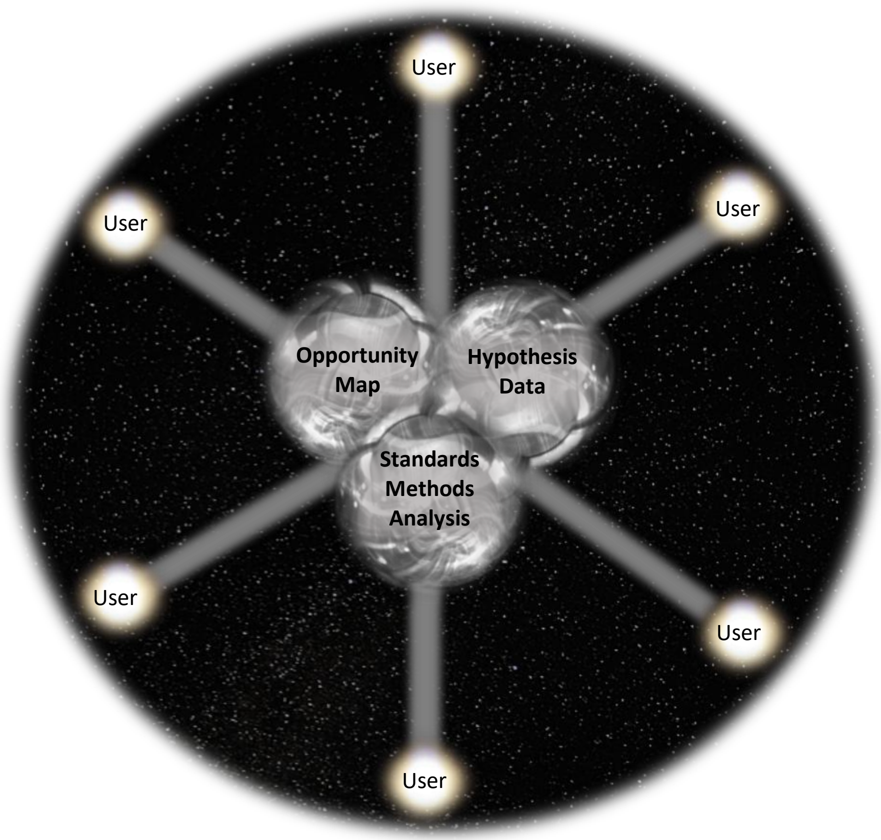


Fig. 2.3: Hub-and-Spoke Model Illustrated

advancing manufacturing competitiveness. They feed data into and pull data out of the hub while adhering to standards and methods.

The consequence of the framework replicates approaches in other highly impactful programs such as those discussed in Section 1 where they increase forecast accuracy to realize high impact/returns. The potential growth in impact is likely substantial for R&D programs that do not track impact or generate feedback loops for impact.

2.4. Notes and Cautions

This approach is grounded in measuring impact and testing hypotheses about what drives impact, an approach that equates to applying scientific methods leading to impact growth. A change agent organization is unlikely to approach or reach its maximum impact potential without impact tracking and hypotheses testing, the foundation of this approach.

It is important to note that a change agent organization likely still needs a diversified portfolio due to diminishing returns, risk of project failure, and avoiding strategic lock in where the organization loses various research capabilities. There is also a need to maintain intellectual humility, as predictions and measures of impact are not absolute. Additionally, the intended purpose of the framework above is primarily to ensure that each group or project moves closer to its highest impact potential and not to justify shifting funds from one group to another, which may have unintended consequences. Moreover, selecting the highest-impact projects alone may not necessarily lead to a change agent reaching its potential impact, as additional operational factors must be considered. Thus, impact predictions/forecasts are better treated as strategic guides rather than automated lock-in decisions.

There are a number of conditions that can affect the success of the framework, including:

- Incentive structures
- Incorporating economics into project development and selection
- Economic knowledge among staff
- Resources allocated to manufacturing economics
- Overall good faith cooperation with economic staff

If incentive structures are such that economic analyses are suppressed or altered, it affects the success of the framework. Also, if the organization does not incorporate economics into project decisions, it will likely be problematic. There is also the natural presence of incentives for inflating impact assessments; thus, to counter this it is likely necessary to reward accuracy in predictions. Staff that determine the projects that move forward need to have knowledge about and understand economic issues. The economics staff will likely need enough funding to facilitate analysis. Finally, there is a need for overall good faith cooperation between economic and technical staff.

There are a few assumptions/structures that the framework tends to rely upon. The first is that it assumes the cost of building the data cube is lower than the value gained from the increased prediction accuracy. Note that a change agent does not need to necessarily fill the entire cube, only those places where data is needed. Data accuracy can also be reduced in places where it is not needed. The data cube can be filled over time rather than creating it all at inception.

The framework also assumes staff have the incentive to report their failures (e.g., errors in prediction) honestly and decision-makers can maintain intellectual humility to accept that their predictions may be wrong and require adjustment. It also assumes the organization is willing to allocate sufficient resources specifically to manufacturing economics and data infrastructure. Without resources and good faith cooperation, the feedback loop breaks. The accuracy of the Recalibration Factor (RCL) is entirely dependent on the quality of the error data fed into it.

Ensuring that the feedback loop does not break involves addressing the psychological and structural barriers that prevent honest reporting. If a change agent only rewards high-impact results, researchers may be tempted to over-forecast potential impact and then over-report actual impact. Additionally, staff may fear that reporting a large gap between predicted impact and actual impact will result in reduced funding or negative performance reviews. This effect is similar to that experienced in safety. In settings with elevated safety risk, rewarding zero incidents often leads to the suppression of near miss data, making the environment more dangerous by hiding risks. To ensure the feedback loop functions, the organization will need to shift its value system from outcome-only to accuracy-first:

Reward the "Near Miss" in Forecasting: Just as safety cultures reward the reporting of hazards before they cause injury, an impact culture can reward the identification of prediction errors. An analyst who identifies why a project underperformed provides more value than one who obscures a failure.

Intellectual Humility: The framework assumes that decision-makers maintain humility to accept that their initial predictions might be wrong. This can be formalized by evaluating staff based on the error of their portfolio over time, rather than the success of a single moonshot. There is also a need for humility in impact forecasts: when a project faces high uncertainty or unknown unknowns, intuition may outperform formal impact assessments. This suggests maintaining a secondary pathway for project selection and development that allows intuition-driven opportunities to be pursued alongside data-driven projects. Note that this would likely account for only a minority of projects, which would be highly novel or exploratory in nature.

Decoupling Reporting from Funding: It is crucial that the framework is used to optimize project design rather than to justify shifting funds between groups in a punitive manner, which would immediately trigger defensive (and dishonest) reporting. Stated another way, if funding is coupled with impact predicting and reporting, the power of the framework is diminished.

Leading Indicators as a safety valve: The framework's success depends in part on the assumption that realized impact can be measured with sufficient speed to inform subsequent cycles of project proposals. To address the inherent time lag of long-term R&D, the system can employ a dual-layered approach. Noisy leading indicators (e.g., publication downloads, citations, and page visits coupled with subject context, value potential, and case study evidence) provide rapid, real-time signals to guide near-term decision-making. Meanwhile, more accurate lagging indicators serve as the definitive ground truth for periodic structural recalibration. Recognizing that not all R&D produces impact on the same timeline, projects are evaluated on clocks consistent with their expected impact latency, ensuring that slow-burn but high-value work is not penalized. In this way, lag time is managed not by eliminating it, but by structuring learning so speed informs decisions without redefining success prematurely. Using noisy leading indicators allows for quicker feedback. If staff can report a likely failure during the execution phase without penalty, the organization can pivot resources more effectively than waiting for a lagging "actual impact" report. The full dissemination and adoption of change agent standards and technologies can take years to fully mature. Change agents, however, likely want early signals as to whether a standard or technology is successfully being implemented.

Framework Adoption: one might present the conditions for reaching a change agent's potential impact as a pyramid, as illustrated in Fig. 2.4. It requires the right culture and incentive structure, standardized methods, tracking and testing, optimizing and researching impact configurations, and implementing the findings that lead to reaching a change agent's impact potential. This process is itself an exercise in scientific reasoning, presenting challenges that necessitate continued discovery and methodological advancement. Achieving the right conditions is likely difficult and critical, as there can potentially be a misalignment of incentives. There is likely a tendency by staff at any organization to prefer a system with less accountability and lower burden, unless it produces individualized benefits of greater value. The reason being is that it is natural to want to conserve effort and reduce exposure to personal risk. The proposed framework comes with significant benefits (e.g., enabling impact growth) but they tend to be at the organizational level and not at the individual level. The framework also comes with additional burden and more accountability at the individual or local level. Moreover, there is the potential for tension between reducing personal burden/risk for individual optimization and maximizing enterprise-level productivity and accountability for organizational optimization.

Incentive misalignment challenges might be addressed by emphasizing the framework's potential benefits: impact growth, improved ability to communicate impact, earlier indicators of progress, and reduced bias in impact evaluation, making it more likely that meaningful contributions are recognized and less likely that ineffective efforts are mistakenly credited. Further, an organization can increase the probability of successful implementation by providing tools to facilitate adoption (e.g., guides and software); reducing individual risks by, to the extent possible, decoupling funding from impact evaluation; recognizing and praising adoption; and spreading costs of adoption out by modularizing the framework. Modulization might include breaking the framework into smaller components where each component has benefits. For instance, the modules might include the following:

- Project Characteristics Tracking: This module implements a systematic approach for tracking and characterizing projects and the components of projects (e.g., categorization). The benefit of this module is that it provides a clean way to group and discuss projects.
- Rough Order of Magnitude Impact Tracking: This module implements the tracking of fast-moving noisy impact indicators. The benefit of this module is that it provides

leading indicators of impact that can be used to guide decisions and communicate potential impact.

- Deep impact analysis: This module includes implementing impact analyses of projects. The benefit is that it can aid in understanding the accuracy and precision of fast-moving indicators and provides a more robust estimate for communicating impact.
- Prediction Data System (Industry Data): This module includes developing industry data to guide in the design and selection of projects.
- Formalized Predictions and Tracking: This module includes formalizing and tracking predictions. The benefit being that it aids in understanding, designing, and selecting projects.
- Analysis and Recalibration: This module includes the analysis of collected data to generate actionable insight with the potential benefit being that impact is increased over time.

This modularization spreads out costs into bite size pieces where each piece has its own benefits. The next section discusses predicting and tracking the adoption/diffusion of change agent standards, innovations, and technologies.

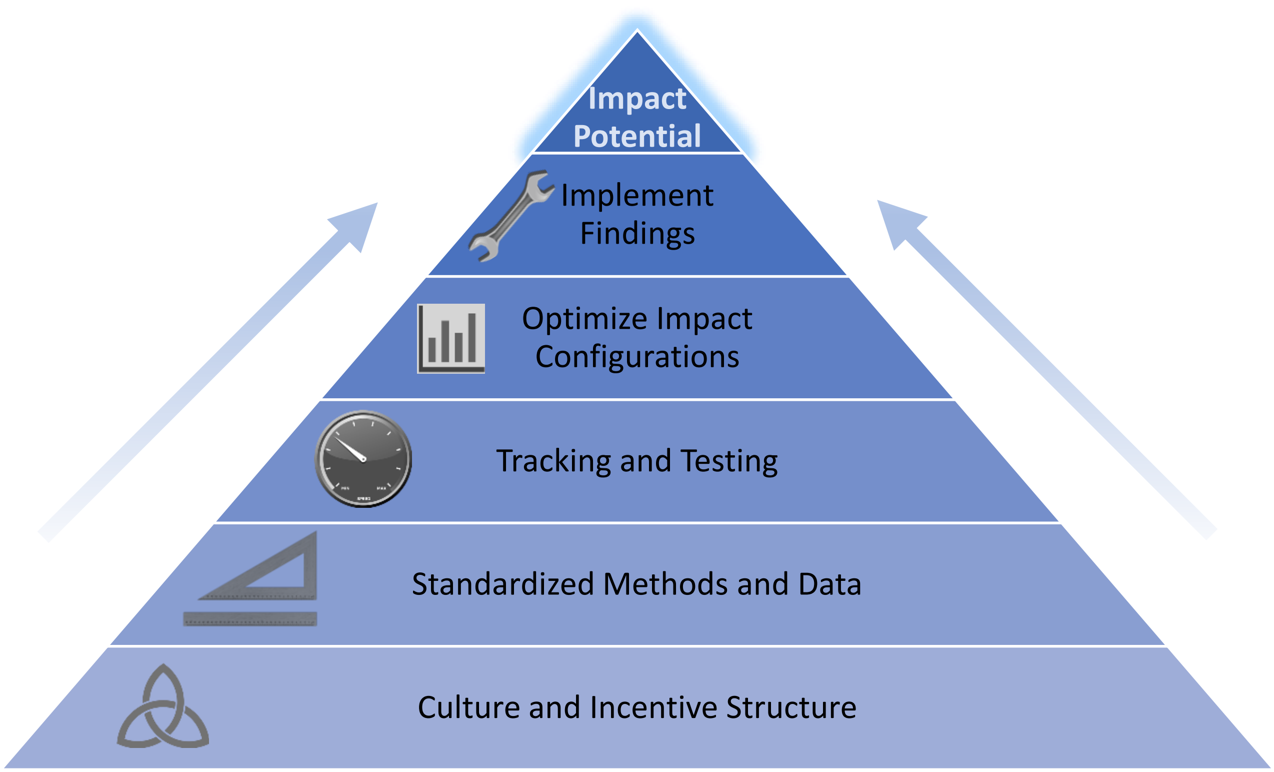


Fig. 2.4: Adoption Conditions for the Framework

2.5. Adoption, Diffusion, and Long Time Horizons

Adoption and diffusion describe how innovations move from initial development to widespread use across firms and markets. Adoption in this context refers to the decision by individual manufacturers to implement a new technology, standard, or practice, while diffusion captures the aggregate pattern of those decisions over time—often following an S-shaped curve, with slow early uptake, rapid acceleration as benefits become visible and networks form, and eventual saturation. Understanding diffusion is essential for estimating long-term economic impact, since the value of an innovation depends not only on its technical merit, but on how quickly and broadly it is adopted across an industry. The most widely accepted model of technology diffusion was presented by Mansfield (1995):

$$p(t) = \frac{1}{1 + e^{\alpha - \beta t}}$$

where

$p(t)$ = the proportion of potential users who have adopted the new technology by time t;

α = location parameter; and

β = Shape parameter $(\beta > 0)$.

An example of applying this can be found in Thomas and Gilbert (2014). To examine the adoption of additive manufacturing, they assumed that the proportion of potential units sold by time *t* follows a similar path as the proportion of potential users who have adopted the new technology by time t. In order to examine shipments in the industry, they assumed that an additive manufacturing unit represents a fixed proportion of the total revenue; thus, revenue will grow similarly to unit sales. The proportion used was calculated from 2011 data. The parameters α and β are estimated using regression on the cumulative annual sales of additive manufacturing systems in the U.S. between 1988 and 2011. United States system sales are estimated as a proportion of global sales. This method provides some insight into the current trend in the adoption of additive manufacturing technology. Unfortunately, there was little insight into the total market saturation level for additive manufacturing; that is, there was not a good sense of what percent of the relevant manufacturing industries would produce parts using additive manufacturing technologies versus conventional technologies. To address this issue, a modified version of Mansfield's model was adopted from Chapman (2001):

$$p(t) = \frac{\eta}{1 + e^{\alpha - \beta t}}$$

where

η = market saturation level in percent.

Because η was unknown, it was varied between 0.15 % and 100 % of the relevant manufacturing shipments, as seen in Table 2.2. The 0.15 % was derived from Wohlers estimate that the 2011 sales revenue represents 8 % market penetration, which equates to $3.1 billion in market opportunity and 0.15 % market saturation. At this level, additive manufacturing was forecasted to reach 50 % market potential in 2018 and 100 % in 2045, as seen in the table. A

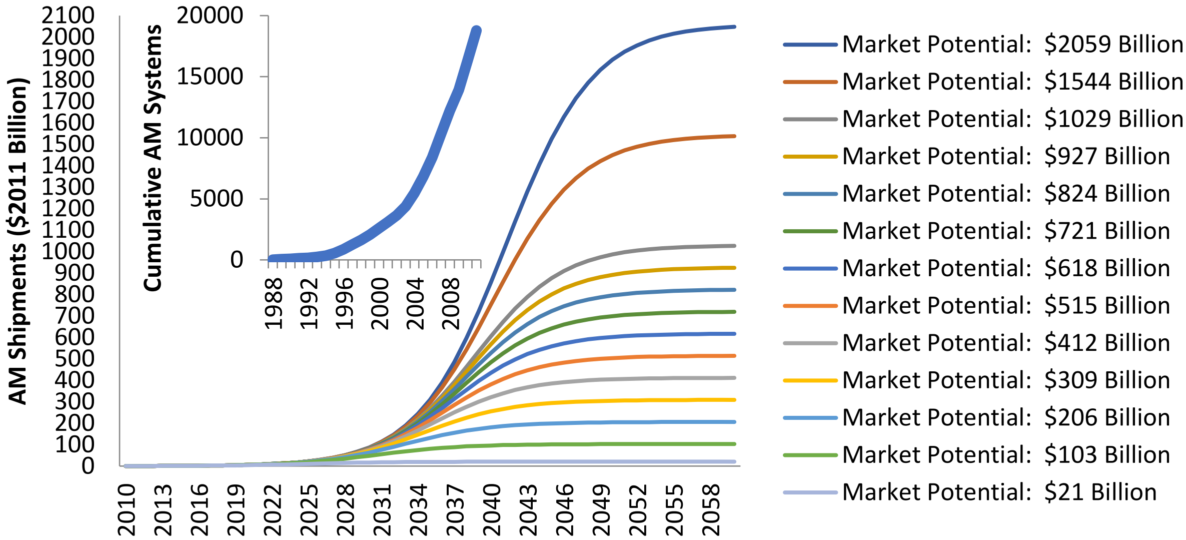
Table 2.2: Forecasts of U.S. Additive Manufacturing Shipments by Varying Market Potential

Market Potential of Relevant Manufacturing (percent of shipments)	Market Potential, Shipments ($billions 2011)	Approximate Year 100% of Market Potential Reached	Approximate Year 50% of Market Potential Reached	Approximate Year $100 Billion in Shipments is Reached	Approximate Year $50 Billion in Shipments is Reached	R^2
100.00	$2 058.9	2069	2042	2031	2028	0.948
75.00	$1 544.2	2068	2041	2031	2028	0.948
50.00	$1 029.5	2067	2039	2031	2029	0.948
45.00	$926.5	2066	2039	2031	2029	0.948
40.00	$823.6	2066	2038	2031	2029	0.948
35.00	$720.6	2065	2038	2031	2029	0.948
30.00	$617.7	2065	2037	2031	2029	0.948
25.00	$514.7	2064	2037	2032	2029	0.948
20.00	$411.8	2063	2036	2032	2029	0.948
15.00	$308.8	2062	2035	2032	2029	0.948
10.00	$205.9	2061	2033	2033	2029	0.948
5.00	$102.9	2058	2031	2044	2031	0.948
1.00	$20.6	2052	2025	-	-	0.949
0.50	$10.3	2050	2023	-	-	0.949
0.15	$3.1	2045	2018	-	-	0.950

Thomas, Douglas. 2013. Economics of the U.S. Additive Manufacturing Industry. NIST Special Publication 1163. Gaithersburg, MD: U.S. Dept. of Commerce, National Institute of Standards and Technology.

more likely scenario to the authors seemed to be that additive manufacturing would be between 5 % and 35 % market saturation. At these levels, additive manufacturing would reach 50 % of market potential between 2031 and 2038 while reaching 100 % between 2058 and 2065, as seen in the table. The industry would reach $50 billion between 2029 and 2031 while reaching $100 billion between 2031 and 2044. As illustrated in Fig. 2.5 and Table 2.2, it was likely that additive manufacturing was at the far-left tail of the diffusion curve, making it difficult to forecast the future trends. The figure illustrates the diffusion at each market saturation level presented in Table 2.2 with the exception of the 0.50 % and 0.15 % levels, as they are too small to be included in this graph. This example illustrates how diffusion modeling can provide insight into adoption, providing reasonable boundaries for diffusion expectations.

To address the issue of having long maturity time spans (e.g., 10+ years), a change agent can create a hypothesized rate of adoption and diffusion for their products (e.g., standards and technologies). To do this, an **adoption index** can be used as a proxy for diffusion progress. Rather than attempting to measure the exact number of firms using a standard or technology,



Thomas, Douglas. 2013. Economics of the U.S. Additive Manufacturing Industry. NIST Special Publication 1163. Gaithersburg, MD: U.S. Dept. of Commerce, National Institute of Standards and Technology.

Fig. 2.5: Forecasts of U.S. Additive Manufacturing Shipments, by Varying Market Saturation Levels

an adoption index combines multiple observable indicators—such as awareness signals (e.g., conference sessions or trade mentions), ecosystem embedding (e.g., software vendor integration, training programs, and standards citations), and direct implementation signals (e.g., procurement requirements or pilot deployments)—into a single normalized score. By weighting these components according to their predictive value and aggregating them into a 0–100 scale, the index provides a consistent, repeatable measure of where an industry sits on the diffusion curve for a standard or technology. Over time, the index can be calibrated against known adoption milestones and used to estimate the proportion of potential adopters reached, identify bottlenecks, and update diffusion parameters, enabling a more dynamic and evidence-based forecast of long-term adoption and economic impact. In the event that the products of a project are not being adopted, various factors of adoption and diffusion (see factors in Fig. 2.6) can be altered to address barriers that were previously unidentified.

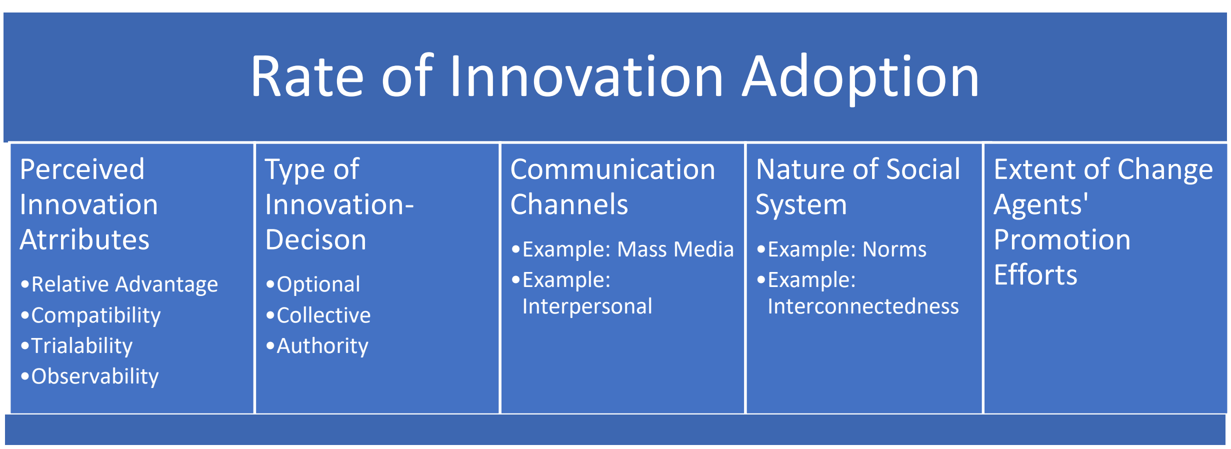


Fig. 2.6: Rogers Variables Determining the Rate of Innovation Adoption

2.6. Metrics and Units of Observation

Metrics: In identifying metrics, it is important to understand the holistic perspective. As illustrated in Fig. 2.7, the manufacturing industry change agent path to increasing long-run economic growth (i.e., real per capita gross domestic product or GDP) and increasing consumer utility (e.g., saving consumers money and/or time) starts with conducting research and development activities. These activities lead to efficiency and/or productivity innovations that are then adopted and disseminated among manufacturers. By definition, efficiency and productivity increases result in making the same things with less resources; thus, resources are freed up. Then, something new is done with these resources; for instance, a manufacturer might produce more goods or a consumer buys a new product. The new activity results in improving the quality of life and/or increasing economic security by making things better and/or cheaper.

Recall that there are three primary mechanisms through which economics achieves impact within change agents, as illustrated in Fig. 2.7:

- **Motivating** manufacturers to or not to adopt innovations by estimating their cost effectiveness;
- **Guiding** change agent R&D to high impact projects by evaluating the costs and benefits for each potential project; and
- **Justifying** change agent R&D budgets by evaluating if or when impact occurs.

The gold standard metric for “guiding” and “justifying” is commonly considered to be benefit cost analysis (BCA) and economic rate of return (ERR) (Boardman et al. 2018; Office of Management and Budget 2023). Conceptually, BCA is often calculated, in part, as the freed-up resources from Fig. 2.7 less the resources used to free them up (i.e., benefits less costs). The ERR is the rate of return calculated using the value of freed-up resources and resources used to free them. Note that projects may also create other types of social benefits; thus, the measures here focus on monetized costs/benefits, but other metrics could be appropriate.

For “motivating” manufacturers, net present value (NPV) and internal rate of return (IRR) are the gold standard metrics (Thomas 2017), which 75 % of firms report always or almost always using for deciding which projects or acquisitions to pursue (Graham and Harvey 2001). By forecasting gold standard metrics or reasonable proxies (discussed below), change agents can prioritize projects that are most likely to generate high impact relative to their cost or resource requirements. Note that metrics should be interpreted alongside assessments of uncertainty and risk, as high predicted impact does not guarantee success.

In estimating impacts, there are both first order and second order effects. First order effects typically include the direct or immediate impact of an innovation, such as reduced electricity costs for manufacturing, increased quality, or increased throughput. Second order effects include indirect consequences that arise in response to first order effects, such as increased

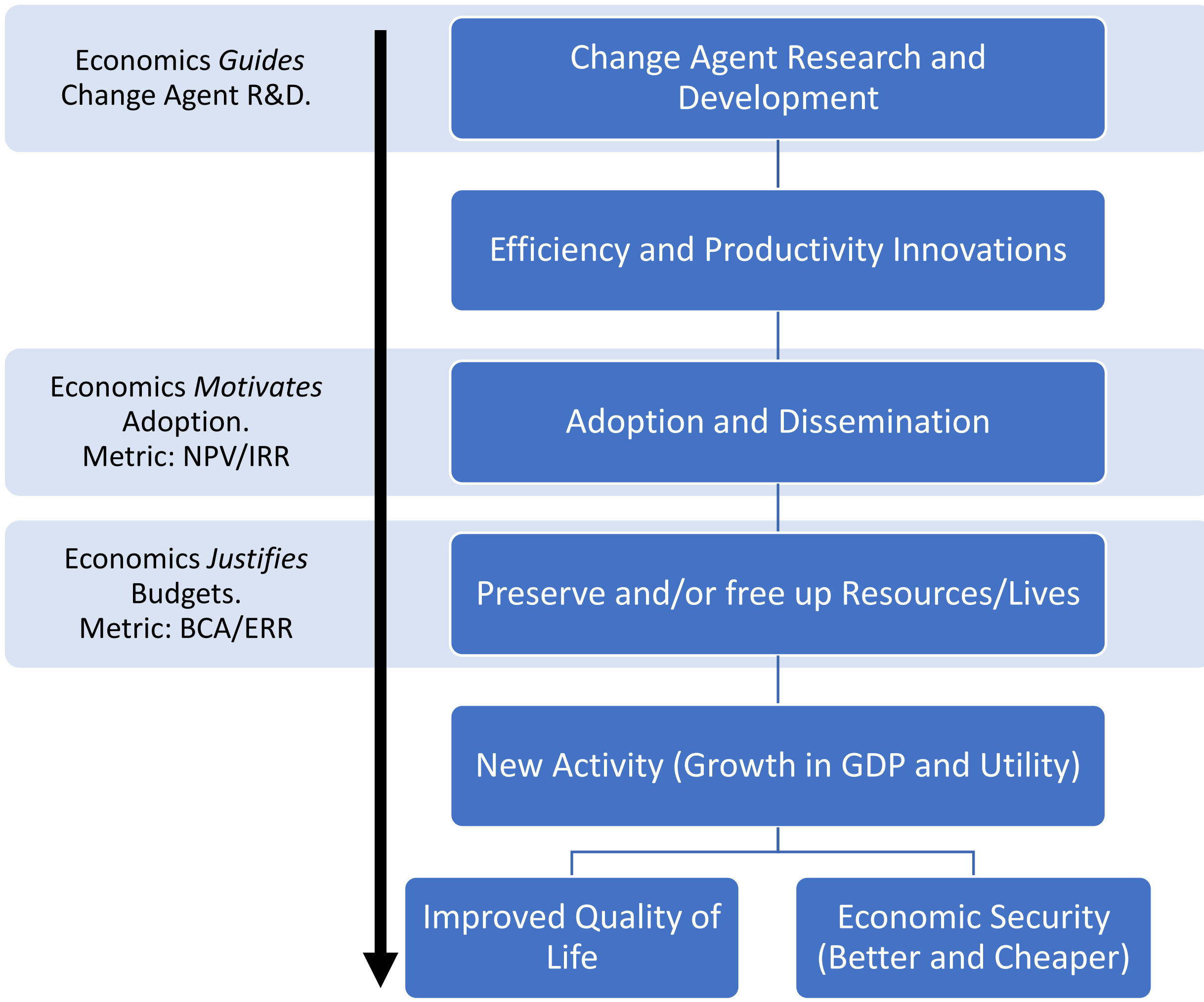


Fig. 2.7: Change Agent Path to Growing Both Real Per Capita GDP and Consumer Utility

adoption of a particular technology or facilitating further innovation. Second order effects can exceed that of first order effects but are more difficult to predict. Technologies that alter decisions throughout a larger system tend to have larger second-order effects. Thresholds and tipping points, complementary technology, network effects (i.e., value increases nonlinearly as adoption increases), large downstream consequences, high elasticity, learning curves, and rare event prevention can all result in high second order effects. Second order effects might be mapped in order to determine if there are potential high levels of second order effects. If there is the potential, then they might be further investigated in order to approximate the potential magnitude. The data cube would tend to be able to capture first order and second order effects that occur within the manufacturing industry itself (e.g., advances in robotics that decrease automation costs); however, second order effects that are external to the industry (e.g., lives saved due advances in medical equipment) would need to be explored through other means and data.

As a result of the data requirements and the required technical knowledge, the gold standard metrics are often difficult to calculate in practice. Therefore, proxies could potentially be used as a substitute for prediction purposes. As discussed previously in Section 1.3, there is a tradeoff between costs and accuracy. One can reduce the cost burden by allowing more error into the prediction. An example of a proxy might be to apply the manageable subcomponents approach discussed previously. Another example of a proxy would be to estimate the annualized benefits rather than the BCA or NPV for a study period. These approaches make generating predictions more accessible to practitioners but lead to reduced accuracy in prediction. Moreover, there is a challenge in achieving the optimal level of prediction accuracy.

If gold standard metrics or gold standard proxies are not feasible, the next metric priority would be intermediate metrics. These include items that lead to impact. For instance, for a researcher this might be publication counts or standards created. This is not a measure of impact, but they tend to lead to impact, making them an intermediate metric. In addition, qualitative evidence can offer complementary insights, enriching or validating quantitative metrics when direct measures are unavailable or incomplete.

One approach for using proxies and gold standard metrics together is to create a layered data system that connects proxies to measured impact. This system would start with hypotheses and move through deeper economic impact analyses:

- Layer 0: Industry data for hypothesis generation and project development
- Layer 1: Tracking of hypotheses and decision-making process (enables a decision system with continuous improvement)
- Layer 2: Proxy layer with fast scalable leading indicators (e.g., Technology Readiness Level, Adoption/Utilization, Standards, Surveys)
- Layer 3: Occasional deep impact studies (ground truth calibration)
    - Independent third-party assessments for increased credibility
- Layer 4: Aggregated macro impact
    - Estimate indirect effects using IO or CGE models

Layer 0 is the data cube discussed previously. As illustrated in Fig. 2.8, it guides project development and is used to develop predictions of impact. This creates a standardized source of data for prediction.

Layer 1 tracks the decision-making process and the predictions on which they are based, creating a referenceable trail for improvement. Predictions are to be testable and at a granular level to facilitate improvement in prediction and decision accuracy.

Layer 2 includes a proxy layer of fast scalable leading indicators. Accurate impact assessments are costly and time consuming; thus, there is a limitation on the extent to which they can be made at a granular level and used for decision making. As a result, it is necessary to utilize lower-cost higher-frequency metrics with the tradeoff being that there is a loss in accuracy for estimating impact.

Layer 3 addresses the loss in accuracy associated with proxy metrics by implementing impact studies that link proxies with measured impact. This enables translating proxy metrics into

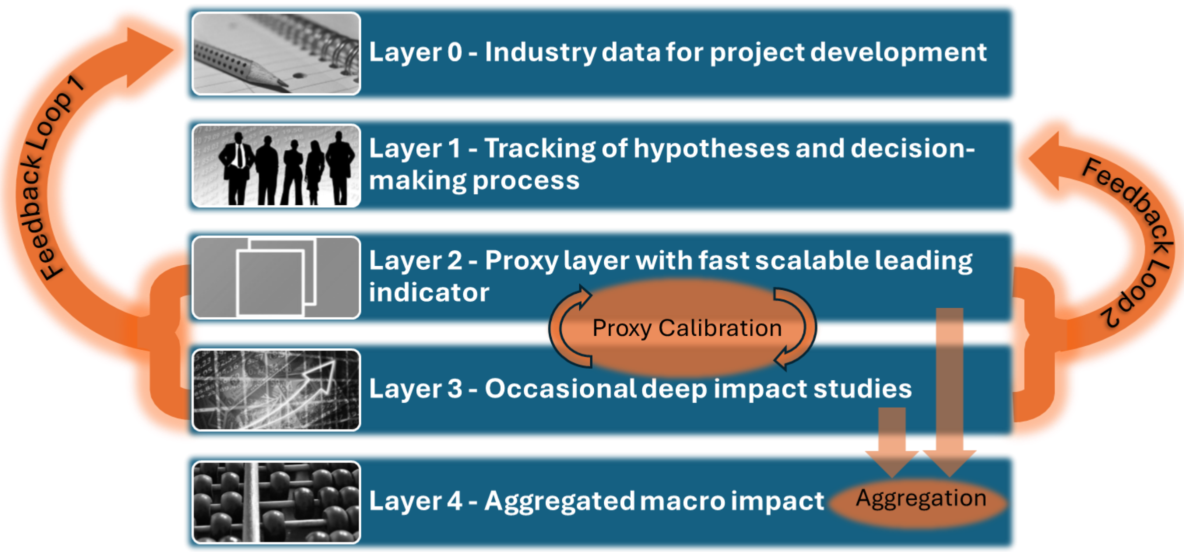


Fig. 2.8: Illustration of Data Layers and Data Flow

approximated impact estimates in a common unit. Over time, this approach supports increasingly accurate, timely, lower-cost, and decision-relevant impact estimates at a sufficiently granular level.

Layer 4 aggregates impact estimates at a high-level view of performance to facilitate overall accountability.

The system enables near real-time aggregatable impact estimates, forming a growing layered body of evidence that supports drilling down to underlying data. This system can be implemented to create an interoperable impact accounting system at the Layer 3 and Layer 4 levels. The system supports improvements in impact and performance by tracking components of the decision-making process, formalizing the assumptions underlying decisions (i.e., the predictions of impact), and generating testable impact hypotheses. These characteristics facilitate identifying when alterations in decision-making components increase performance.

Units of Observation: In addition to considering metrics, it is important to develop effective units of observation. This report has discussed developing high-impact projects, but a project can encompass a wide range of activities. When some activities are grouped together, they may obscure the ability to study the specific configurations that lead to high impact. For example, combining the effort to develop a standard for data interoperability in product design with a standard for data interoperability in production machinery blends impacts from different sources, making it harder to isolate the effects of each.

To solve this, a **project** is—for the purposes of this report—defined based on three criteria:

1. It influences the same industry or adoption community.

2. It uses the same mechanism to improve productivity/efficiency (e.g., the cost item reduced for the manufacturer).
3. It groups related products and activities (e.g., a standard plus its supporting workshops and articles).

By this definition, a project is an impact-oriented grouping where, for instance, a standard for AI in manufacturing processes and a separate standard for AI in quality assurance are treated as two distinct projects. Even though both involve AI, their mechanisms of impact are different. The goal is to ensure that the grouping of activities never mask the factors that contribute to optimal results. An organization might call this a subproject or some other term; however, for simplicity this report uses the term "project."

A project should be defined so that all its activities contribute to impact primarily through the same causal mechanism or closely aligned mechanisms. Activities with substantially different mechanisms should be considered separate projects for analytical purposes. For instance, an effort in digital twins might be parsed into four separate projects based on impact:

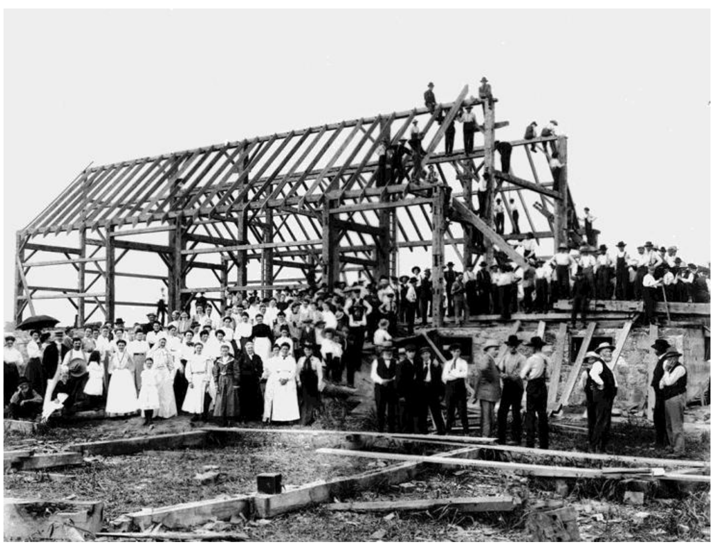
- Digital twin standards development (increasing digital twin performance), which reduces
  - Costs of integrating equipment
  - Rework/redesign costs due to incompatible data formats
  - Switching costs
- Digital twin implementation methods (decreasing installation costs), which reduces
  - Engineering design costs for digital twin setup
  - Training costs for staff
  - Implementation errors and delays
  - Setup costs
- Digital twin demonstrations (reducing the cost of investment analysis), which reduces
  - Prototype testing costs
  - Risk mitigation from untested automation
  - Downtime costs from virtual testing
- Digital twin validation (ensuring digital twin performance), which reduces
  - Quality assurance cost
  - Costs of errors or deviations

These projects collectively encompass the full impact of the broader digital twin initiative, but by parsing them, each unit can be analyzed individually to determine which specific mechanisms and configurations drive the greatest cost savings and efficiency gains.

3. Platforms

The framework involves economic decision making, generating hypotheses, and collection and analysis of economic data; however, there are not enough economists to make hypotheses and measure impact for every project. Thus, many of these activities will be executed by non-specialists, who have limited understanding about investment analysis methods. To increase the impact of investments in manufacturing competitiveness, one might aim to increase the accuracy of these non-specialist hypotheses (i.e., impact predictions), data collection, and economic decision making by developing and adopting platforms or what might be called standardized frameworks; that is, a standard method or approach, which can reduce the cost/burden of, for example, making a prediction and reduce the learning curve of making predictions while maintaining/creating an acceptable level of quality, performance, and accuracy in the predictions. This facilitates widespread quality hypothesis generation among other things.

An example of a platform or standard framework might be found in past methods for barn raising, which the Amish still use today as illustrated in Fig. 3.1 (Bronner 2019). It involved skilled workers handling joinery and structure while others lifted beams, prepared materials, and provided food. When barns are raised this way there is a method or standard approach that is used. Not every worker is a carpenter, but everyone is able to contribute because of the standard approach, which can allow people to literally work shoulder-to-shoulder. Another example might be found in NIST EL, which developed a framework for safety evaluation. Some years ago, the Engineering Laboratory at NIST aimed to reduce the risk of safety incidents; however, there is a general challenge for NIST and all organizations in that there are not enough resources to hire a safety professional for every project or program. The solution is to develop a platform or standardized approach that non-safety experts can use to evaluate risk and systematically select solutions that pose higher levels of safety. The program then has a safety expert to assist or guide this effort. Part of the standardized approach utilizes a risk matrix, such as the one in Table 3.1, for evaluating hazards.



Source: Wikimedia. Barn raising in Lansing.
https://commons.wikimedia.org/wiki/File:Barn_raising_in_Lansing.jpg

Fig. 3.1: A 20th Century Barn Raising, Toronto, Canada

Table 3.1. Example of a Framework: Risk Matrix for Evaluating Hazards

	Potential Severity of Hazard			
Likelihood of Occurrence	Catastrophic: Death or permanent disability, system or facility loss, lasting environmental or public health impact	Severe: Serious injury; temporary disability; subsystem loss or facility damage; temporary environmental or public health impact	Moderate: Medical treatment; lost-work day(s), minor facility damage, external reporting/cleanup requirements	Minor: First aid only, negligible or slight damage, routine cleanup
Frequent: Likely to occur repeatedly	Critical: Not Permitted	Critical: Not Permitted	Serious	Medium
Probable: Likely to occur multiple but infrequent times	Critical: Not Permitted	Critical: Not Permitted	Serious	Medium
Occasional: Likely to occur at some time	Critical: Not Permitted	Serious	Medium	Low
Remote: Possible, but not likely to occur	Serious	Medium	Medium	Low
Improbable: Very unlikely; can reasonably assume it will not occur	Medium	Low	Low	Minimal

There are two sets of data that facilitate the framework. The first is observational data on the manufacturing industry, which is referred to in the framework as an opportunity map. This includes an extensible dataset characterizing the economics of the industry (e.g., machinery cost, maintenance cost, and labor cost). This data is utilized to guide the development and selection of projects to advance competitiveness, and it is used to generate hypotheses/predictions for economic impact. The second dataset includes tracking the hypotheses for projects, tracking the estimated impacts of projects, and tracking characteristics of the project that affect impact. All the characteristics that need to be controlled for and studied to further the change agent's understanding of how greater impact can be achieved. Methods need to facilitate data collection, storage, utilization, and analysis; thus, there are needs for the following:

- Dataset 1: Opportunity map
    - Standardized data categories
    - Standardized metrics
- Dataset 2: Hypotheses and Impact Estimates
    - Standardized methods for generating hypotheses
    - Standardized methods for impact assessment
    - Standardized data categories
    - Standardized metrics

Successfully implementing the framework across an organization will hinge on creating these platforms in such a way that metrics can be estimated and stored in a consistent way.

3.1. Platforms for Hypotheses and Economic Decision making

A platform or standardized approach for increasing prediction accuracy for manufacturing industry R&D investments is likely to harness existing methods for economic decision making; however, it will likely need to be simplified or automated for use by non-specialists. Such a method would not be intended for economists or other financial experts and should not be seen as having the highest level of accuracy. Rather, it is to facilitate a step increase in the accuracy of predictions made by non-specialists. A simple example might harness the style of the previously discussed risk matrix from NIST's EL. The approach could put costs and benefits on the axes of a matrix such as that in Table 3.1. Rather than having costs and benefits that are precise, as an economist might estimate, they might be estimated with higher levels of error. An additional option is to graph the range of cost and benefits for each project. This will give a visual aid for identifying relative rankings, as illustrated in Fig. 3.2. The benefit cost ratio for projects can be visually examined by drawing a line from the origin through the top right corner of the projects graph and another through the bottom right corner, as illustrated for Project A in Fig. 3.3. Because the slope of the line represents a constant benefit cost ratio at the intersecting point of the project, any projects, or portions thereof, that lie above and to the left of the upper line outranks the project. So, Project E and Project B outrank Project A. Any projects that lie below and to the right of the lower line are outranked. So, Project A outranks Project D. Projects that lie between the upper and lower lines have overlapping benefit cost ratios with Project A. Net present value relationships can be revealed by drawing a 45-degree line through the top left and bottom right points of the project graph, as shown in Fig. 3.4. Because this line represents a 1 to 1 change in costs and benefits, the line represents constant net present value; so, those projects that lie above the line have higher net present value and

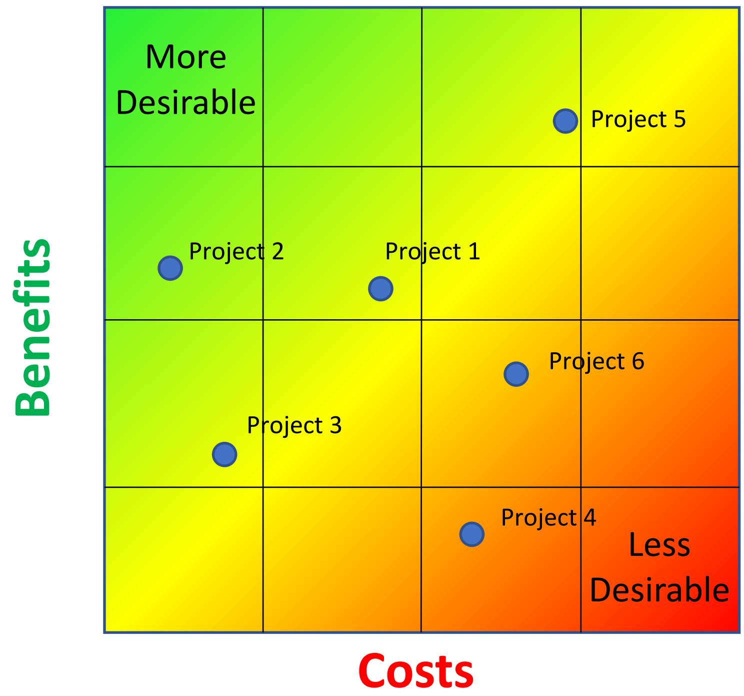


Fig. 3.2. Example of a Framework for Evaluating Projects or Investments

those that lie below the lower line have lower net present value. While the precise nature of the approach is flexible, it is important to have a standardized approach that can be applied by non-specialists across organizations to facilitate data collection and study of the investments. The platform for economic decision making needs to be developed in collaboration with decision science experts and those who would be utilizing the method.

Estimating the benefits and costs of a proposed R&D project would likely involve quantifying annual dollar values. For instance, suppose a proposed R&D collaboration with a plastic bottling plant manufacturer develops a means for reducing energy consumption for a medium sized stretch-blow molder; however, we do not know exactly how much savings will be produced. To estimate the benefits of this R&D project, we need to approximate the savings. Using a simple internet search or even a search on an artificial intelligence platform, one might find that such a machine uses about 3 000 kWh per day. Further, we could search and find that the average cost of electricity is around $0.12 per kWh. If the machine runs 260 days per year, the estimated annual electricity cost is $93 600. This estimate provides a ceiling for the savings that might be had for the R&D for this piece of machinery. Let's say that we do not know the exact percent reduction in energy that should be expected, but given the nature of the R&D, we might surmise it to be between 1 % and 5 % reduction, based on individual insight. The annual savings would be estimated to be between $936 and $4680 annually. This estimate has a moderate level of error; however, when the costs/benefits of multiple projects are estimated and

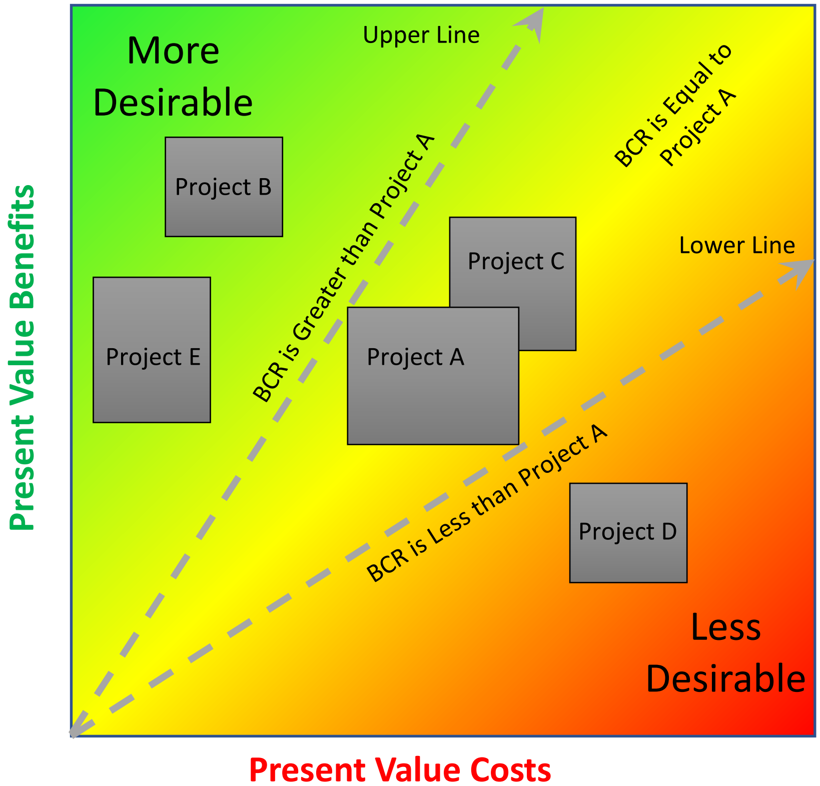


Fig. 3.3. Graphing Costs and Benefits: Examining Benefit-Cost Ratio (BCR)

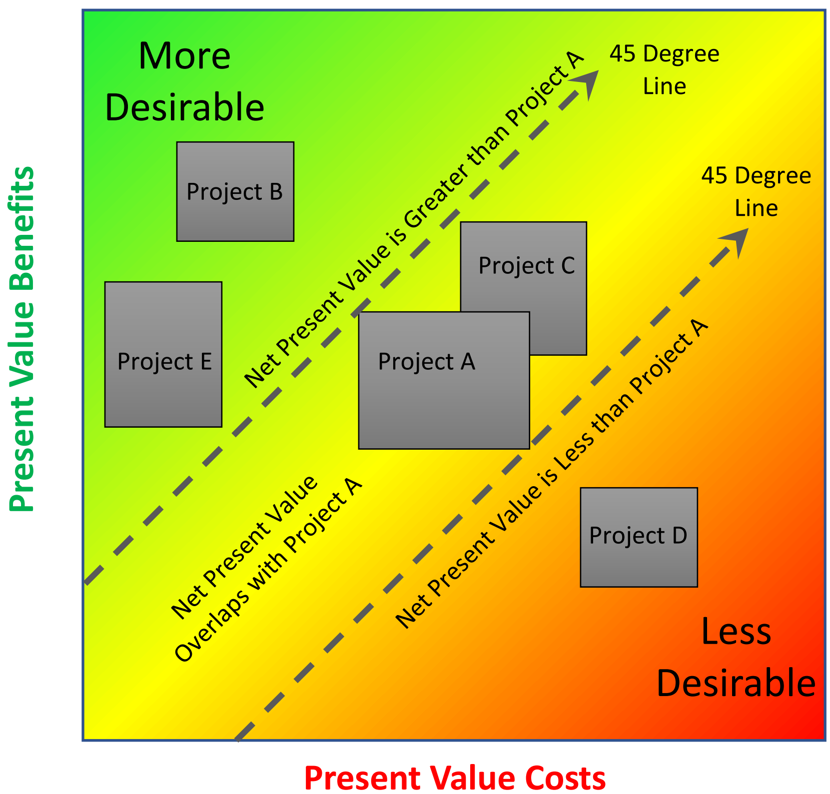


Fig. 3.4. Graphing Costs and Benefits: Examining Net Present Value

compared, this approach likely results in selecting/designing higher impact projects than if one used individual insight without calculating the savings. If the method for energy savings were intended to be disseminated among other bottle manufacturers or other facilities, we might use similarly simplified methods to extrapolate the impact.

An estimate from an economist or other decision science expert might employ additional rigor for these estimates, including using net present value, internal rate of return, or modified internal rate of return. These metrics provide additional accuracy in predictions for economic decision making, but a non-specialist might struggle to employ them. Platforms might be utilized to automate such calculations from basic information from the decision maker and can standardize quantifying more challenging situations such as estimating the benefits of training for employees, increased product quality, or that of increased safety, which often utilizes a more complex concept of the statistical value of life.

3.2. Data/Information Distribution Platforms

In addition to platforms for economic evaluation, it is also important to have platforms or standardized platforms for distributing data/information for non-specialists to make economic evaluations. For instance, tools that distribute industry cost data that evaluators can use for their assessment. Decision makers will need to assess the benefits of various projects, which will require understanding the costs and other factors involved in manufacturing. The tools that provide data need to provide standardized data that are easily accessible and understandable. NIST's MCG tool for supply chain statistics (Thomas 2019) is a tool that aims to fill this type of role. Estimates that would typically require expertise in economic input-output analysis are provided such that no expertise is needed. Similarly, platforms could be designed to disseminate data on items such as previous R&D investments, manufacturing industry costs, and/or losses experienced by manufacturers.

3.3. Data Collection Platforms

As discussed previously, there is a need to track predictions and impacts to learn from the past. Data classifications facilitate the systematic collection of information such that it facilitates easy utilization. Currently, there are limited classifications for collecting data on investments made for advancing manufacturing competitiveness. However, some insight into data classification for investments might be gained from the Department of Energy (DOE) Industrial Training and Assessment Center program. This program has a publicly available database of 148 000 recommendations for 20 000 facilities, as of October 2021. The data is the result of DOE technical assessments of facilities conducted by university engineering students and staff from 26 ITACs made up of 31 universities (Industrial Assessment Center 2021; U.S. Department of Energy 2011). Each observation in the ITAC database is a recommendation for an investment. It includes an Assessment Recommendation Code (ARC) (a system of classification developed by the program), the cost to implement the recommendation, estimated annual savings, year, whether the recommendation was implemented, and some characteristics of the establishment including sales, various energy expenditures, and number of employees (U.S. Department of Energy 2011).

Data on the recommendations are collected using their ARCs, which have between one and five digits and a total of approximately 424 codes. More digits represent additional detail. There are three single digit codes: ARC 2 - energy management; ARC 3 – waste minimization / pollution prevention; ARC 4 – direct productivity enhancements. Additional detail is indicated in numbers to the right of the decimal. For instance, ARC 2.4157, which is within ARC 2, is “Establish a predictive maintenance program.” ASME’s Investment Analysis subcommittee within the Manufacturing and Advanced Manufacturing Committee is utilizing the ARC codes to develop an Industrial Investment Classification system. These types of classifications facilitate data collection and the examination of investments.

Classification of projects or R&D investments is critical for generating actionable information. Impact estimates must be disaggregated by source to enable meaningful attribution of outcomes. When data is aggregated across projects, it becomes difficult to determine whether increases or decreases in impact are driven by specific project characteristics. This can be illustrated through the analogy of a factory containing two types of machinery, such as an additive manufacturing system and a stamping press. Each system has distinct inputs and outputs. If the inputs and outputs of both systems are combined into a single dataset, it becomes difficult to optimize the efficiency and productivity of either machine because the effects of configuration changes are obscured by the performance of the other system. Therefore, data must be categorically separated at a level and specification that enables measurement of the impact of specific configuration changes. It is not sufficient for data to be granular; it must also be properly classified according to the source of impact. Detailed data that combines fundamentally different systems still obscures causal relationships and limits attribution. Accordingly, project categorization must distinguish between different mechanisms through which impact is achieved.

4. System of Continuous Improvement Across Change Agents

The platforms discussed in the previous section can be brought together to create a system of continuous improvement that implements the framework. Change agents can adopt platforms for economic decision making by non-specialists, resulting in a likely increase in accuracy in their assessment of projects/investments. This platform can be accompanied by platforms for data disbursement and data collection.

As illustrated in Fig. 4.1, the process for continuous improvement would start with change agents' project proposals (see *step 1: Change Agent Project Proposals*). The next step (*step 2: Estimate Potential Economic Impact*) in the cycle is predicting or estimating the potential economic impact or return. Both steps 1 and 2 utilize outside economic data that has been previously assembled. The impact prediction (i.e., hypothesis) information is collected in a centralized location for comparison to actual impacts and to guide future investments. The estimated economic impacts are then used to select projects which are then executed (i.e., *Step 3: Project Selection and Execution*). Following the completion of projects, there is an estimate made of the actual or realized impacts, which is seen in *Step 4: Estimate Actual Economic Impact*. In *step 5: Compare and Adjust* of the cycle, the predicted impacts are compared to the

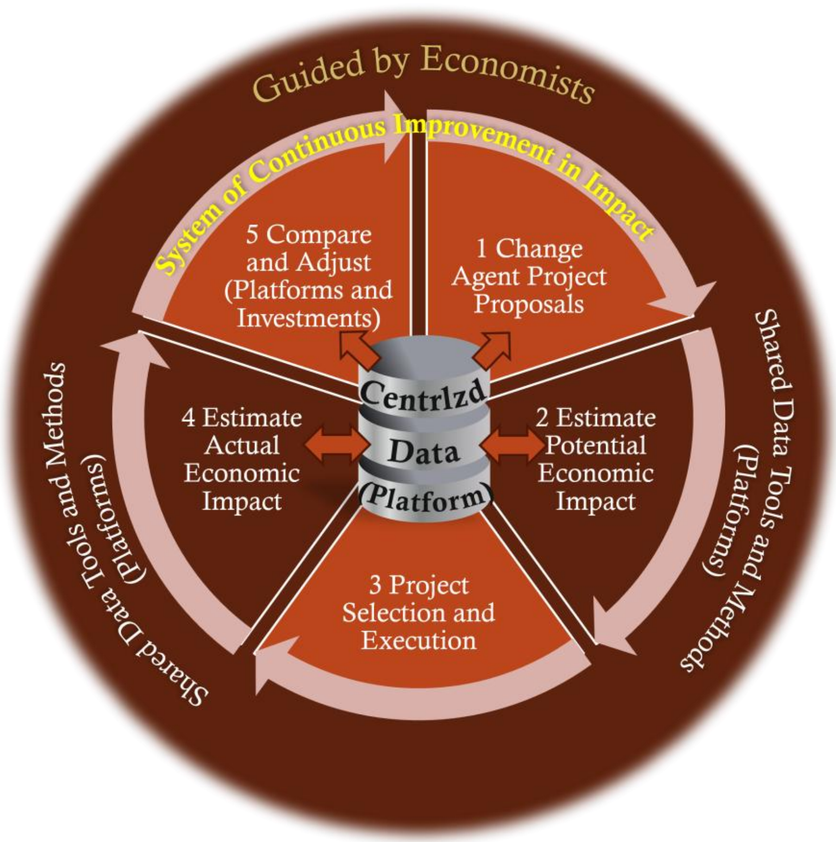


Fig. 4.1. System of Continuous Improvement in Impact Implementing the Economic Framework Process

measured realized impacts to adjust the platforms for economic assessment, a process supported by findings made by the Intelligence Advanced Research Projects Activity (i.e., testing predictions) (Block and Kellerman 2016). The information from the previous projects is then used to guide the development of the next round of proposals in Step 1 of the cycle.

This report proposed an organization-wide system of continuous improvement to increase impact in advancing U.S. manufacturing competitiveness. To maximize the impact of each dollar invested in research and development (R&D), change agents will likely need to transition from intuition-based selection to rigorous measurement science. By treating each R&D project as an experiment with a clear hypothesis (ex ante prediction) and measurable outcome (ex post estimate), organizations can create a feedback loop that systematically improves future investment accuracy. The methods proposed fulfill the three functions discussed in Section 2: Motivate, Guide, and Justify. The method is composed of three parts: Opportunity Map and Hypotheses, Validation and Recalibration, and Enterprise-Wide Utilization. Achieving enterprise-wide utilization is likely to require a centralized hub that maintains data (opportunity map and hypotheses data) and maintains the standard classifications and methods with the result being a hub-and-spoke model. Because of its wide-ranging effects on R&D, this approach can start an engine of sustained growth for change agent impact.

References

[1] Advanced Manufacturing Office, Department of Energy. "AMO Introspective Performance Assessment Methodology with Verification and Validation of R&D Projects (IPA/V&V)." https://docs.nrel.gov/docs/fy19osti/74076.pdf
[2] Ariely, D. (2008). Predictably irrational: The hidden forces that shape our decisions. Harper.
[3] Block, F. L., & Keller, M. R. (2016). State of innovation: The U.S. government's role in technology development. Taylor & Francis.
[4] Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2018). Cost-Benefit Analysis: Concepts and Practice (5th ed.). Cambridge: Cambridge University Press.
[5] Bolger, F., & Önkal-Atay, D. (2004). The effects of feedback on judgmental interval predictions. International Journal of Forecasting, 20(1), 29–39. https://doi.org/10.1016/S0169-2070(03)00009-8
[6] Bronner, Simon J. (2019). "Barn Raising." https://www.encyclopedia.com/humanities/encyclopedias-almanacs-transcripts-and-maps/barn-raising
[7] Burgman, M. A., McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., et al. (2011). Expert status and performance. PLoS ONE, 6(7), e22998. https://doi.org/10.1371/journal.pone.0022998
[8] Chapman, R. (2001). Benefits and costs of research: A case study of construction systems integration and automation technologies in commercial buildings (NISTIR 6763). National Institute of Standards and Technology.
[9] DARPA. (2024). "Proposer Instructions: General Terms and Conditions." https://www.darpa.mil/about/offices/contracts-management/proposer-general-terms/
[10] Endeavor Business Media. (2024). Reshoring manufacturing: Progress and challenges. https://www.industryweek.com/members/ebooks-digital-editions/whitepaper/55129690/reshoring-takes-hold-across-multiple-industries
[11] Google. (2026). "Gemini." https://gemini.google.com
[12] Graham, John and Campbell Harvey. (2001). "The Theory and Practice of Corporate Finance: Evidence from the Field." Journal of Financial Economics 60 (2001): 187-243.
[13] Industrial Assessment Center. (2021). Saving energy and reducing costs at small and medium-sized U.S. manufacturers. https://iac.university/#resources
[14] Industrial Training and Assessment Centers. (2026). "ITAC Download Database." https://itac.university/download
[15] Institute of Medicine (US) Committee for Assessment of NIH Centers of Excellence Programs, Manning, F. J., McGeary, M., & Estabrook, R. (Eds.). (2004). NIH extramural center programs: Criteria for initiation and evaluation. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK24675/
[16] J.S. Held. (2025). 2025 J.S. Held Global Risk Report. J.S. Held. https://www.jsheld.com/insights/articles/2025-js-held-global-risk-report
[17] Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.
[18] Lewis, M. (2004). Moneyball: The art of winning an unfair game. W. W. Norton & Company.

[19] Liker, J. K. (2004). The Toyota way: 14 management principles from the world's greatest manufacturer. McGraw-Hill.
[20] Mansfield, E. (1995). Innovation, technology and the economy: Selected essays of Edwin Mansfield (Economists of the Twentieth Century Series). Edward Elgar Publishing.
[21] NASA. (1969). Apollo manned lunar landing: GOSS mission profile. U.S. Government Printing Office.
[22] NASA. (2024). Apollo 11 mission overview. https://www.nasa.gov/history/apollo-11-mission-overview/
[23] Nichols, Conrad. (2025). Critical Material Recovery 2026–2046: Technologies, Markets, Players. IDTechEx. https://www.idtechex.com/en/research-report/critical-material-recovery/1124
[24] Office of Management and Budget. (2023). Circular No. A-94: Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs. Washington, DC: Executive Office of the President. https://www.whitehouse.gov/wp-content/uploads/2023/11/CircularA-94.pdf
[25] OpenAI. (2026). ChatGPT (May 20 version) [Large language model]. https://chat.openai.com
[26] Panel on Assessment of the National Institute of Standards and Technology Engineering Laboratory. (2025). An assessment of selected research programs and goals of the Engineering Laboratory at the National Institute of Standards and Technology: Fiscal year 2024. National Academies Press. http://nap.nationalacademies.org/27444
[27] Panel on Assessment of the National Institute of Standards and Technology Information Technology Laboratory. (n.d.). An assessment of selected divisions of the National Institute of Standards and Technology Information Technology Laboratory. National Academies Press. http://nap.nationalacademies.org/27430
[28] Sandor, Debbie. 2019. "AMO Introspective Performance Assessment Methodology with Verification and Validation of R&D Projects." https://www.energy.gov/documents/038-poster19-26-introspective-performance-assessment-methodology-verification-and
[29] Shao, G. (2021). "Use case scenarios for digital twin implementation based on ISO 23247." (NIST Advanced Manufacturing Series 400-2). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AMS.400-2
[30] Staal, J., Katarya, K., Speelman, M., Brand, R., Alsma, J., Sloane, J., Van den Broek, W. W., & Zwaan, L. (2024). Impact of performance and information feedback on medical interns' confidence–accuracy calibration. Advances in Health Sciences Education, 29(1), 129–145. https://doi.org/10.1007/s10459-023-10252-9
[31] Stanford. (2026). "Program Evaluation." https://med.stanford.edu/oaa-mentoring/new-program-toolkit/program-evaluation.html
[32] Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? Acta Psychologica, 47(2), 143–148. https://doi.org/10.1016/0001-6918(81)90005-6
[33] Team Ascend. (2025). How UPS's ORION system slashed delivery costs with route optimization. https://www.ascendanalytics.co/post/how-upss-orion-system-slashed-delivery-costs-with-route-optimization
[34] Tetlock, P., & Gardner, D. (2015). Superforecasting: The art and science of prediction. Penguin Random House.

[35] Thomas, D. S., (2013). Economics of the U.S. additive manufacturing industry (NIST Special Publication 1163). U.S. Department of Commerce, National Institute of Standards and Technology.

[36] Thomas, D. S., & Gilbert, S. W. (2014). Costs and cost effectiveness of additive manufacturing: A literature review and discussion (NIST Special Publication 1176). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.1176

[37] Thomas, D. S. (2017). Investment analysis methods: A practitioner's guide to understanding the basic principles for investment decisions in manufacturing (NIST Advanced Manufacturing Series 200-5). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AMS.200-5

[38] Thomas, D. S. (2019). MCG for supply chain statistics. https://www.nist.gov/services-resources/software/mcg-supply-chain-statistics

[39] Thomas, D. S. (2022). "Cost-Effective Environmental Sustainability: A Focus on a Circular Economy." NIST AMS 100-48. https://doi.org/10.6028/NIST.AMS.100-48

[40] Thomas, D. S. (2024). "Economics of Digital Twins: Costs, Benefits, and Economic Decision Making." NIST Advanced Manufacturing Series 100-61. https://doi.org/10.6028/NIST.AMS.100-61

[41] University of Manchester. (2026). "Evaluation Cycle Framework." https://www.staffnet.manchester.ac.uk/umitl/resources/evaluation-toolkit/guidance-resources-and-case-studies/

[42] U.S. Census Bureau. (2024). All Sectors: Summary Statistics for the U.S., States, and Selected Geographies: 2022. *Economic Census, ECN Core Statistics Summary Statistics for the U.S., States, and Selected Geographies: 2022, Table EC2200BASIC*. Retrieved February 18, 2026, from https://data.census.gov/table/ECNBASIC2022.EC2200BASIC?q=manufacturing+shipments+by+establishments+by+size.

[43] U.S. Census Bureau. (2025). All Sectors: County Business Patterns, including ZIP Code Business Patterns, by Legal Form of Organization and Employment Size Class for the U.S., States, and Selected Geographies: 2023. Economic Surveys, ECNSVY Business Patterns County Business Patterns, Table CB2300CBP. Retrieved February 18, 2026, from https://data.census.gov/table/CBP2023.CB2300CBP?q=manufacturing+shipments+by+establishments+by+size.

[44] U.S. Department of Energy. (2011). IAC assessment database manual. https://iac.university/technicalDocs/IAC_DatabaseManualv10.2.pdf

[45] U.S. Department of Energy. (2026a). "Office of Science National Laboratories." https://www.energy.gov/science/office-science-national-laboratories

[46] U.S. Department of Energy. (2026b). "Office of Science Lab Appraisal Process." https://www.energy.gov/science/office-science-lab-appraisal-process

[47] Van den Steen, E. (2004). Rational overoptimism (and other biases). American Economic Review, 94(4), 1141–1151. https://doi.org/10.1257/0002828042002697

[48] Wiggins, Richard H III. (2004). "Personal Digital Assistants." Journal of Digital Imaging. Feb 17, 2004. 17 (1): 5-17. https://pmc.ncbi.nlm.nih.gov/articles/PMC3043961/

[49] Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. Journal of Experimental Psychology: General, 125(4), 387–402. https://doi.org/10.1037/0096-3445.125.4.387

Appendix A. Opportunity Map (Example)

Table A. 1 below presents an assemblage of data to illustrate an opportunity map with the base data being from the Annual Survey of Manufactures (ASM) - note that the ASM is transitioning into the Annual Integrated Economic Survey (AIES). The ASM data in the table is largely available for all six digit North American Industry Classification (NAICS) code. NAICS is a two-to-six-digit classification system for industries. The example in the table is for NAICS 336100, which is the code for motor vehicle manufacturing. The data in the table is four datasets stitched together with the ASM data: wage data from the Bureau of Labor Statistics (BLS) categorized by NAICS and the Standard Occupational Classification (SOC) system; energy use data from the Manufacturing Energy Consumption Survey (MECS) categorized by NAICS; data on materials, packaging, etc used from the Economic Census (EC) categorized by NAICS and by the North American Product Classification system; and data from the Survey of Plant Capacity Utilization (SPCU) categorized by NAICS.

The combining of these datasets illustrates the extensibility of industry data. Additional data can be collected by those who manage the framework hub in a way that allows it to match with the existing classification standards. This might include estimating defect rates, downtime, scrap, or other costs/losses. Other existing datasets can be stitched on as well. For instance, the O*NET database, which provides an extensive set of data —hundreds of variables— that describe work and worker characteristics, including:

- Worker Characteristics
- Worker Requirements
- Experience Requirements
- Occupational Requirements
- Occupation-Specific Information
- Workforce Characteristics

It is organized around the SOC classification system, allowing it to be stitched to the wage data in Table A. 1. The usefulness in this data lies in listing tasks and work activities. Economic Input-Output data, which reveals supply chain relationships, can be stitched on, as it is organized by NAICS codes. All the data in Table A. 1 has structure that utilizes standardized classifications, and it has a hierarchical cost structure that sums up to the value of shipments. This essentially fits the data cube previously described.

The data can be used to guide the development of change agent projects allowing innovators to examine where manufacturing costs amass. It also facilitates approximating the industry level impact of proposed projects by providing data on c_j in Equation 1. For instance, consider a proposed project that reduces machinery maintenance costs in motor vehicle manufacturing. We can estimate the costs of maintenance labor using the BLS data (i.e., 49-0000 Installation, Maintenance, and Repair Occupations) and purchased maintenance (i.e., Repair and maintenance services of buildings and/or machinery). This gives a standardized base cost estimated using standardized classifications and approaches.

Table A. 1: Opportunity Map: NAICS 336100 Motor vehicle Manufacturing

Cost and Revenue Components	$Million
Machinery, structures, and compensation expenditures	27,461
Payroll and benefits (ASM)	22,439
Annual payroll (ASM)	15,477
Wages	
11-0000 Management Occupations (BLS)	134
⋮	⋮
49-0000 Installation, Maintenance, and Repair Occupations (BLS)	67
49-1000 Supervisors of Installation, Maintenance, and Repair Workers (BLS)	92
⋮	⋮
[1447 Categories]	
Total fringe benefits (ASM)	6,962
Employer's cost for health insurance (ASM)	2,900
Employer's cost for defined benefit pension plans (ASM)	589
Employer's cost for defined contribution plans (ASM)	587
Employer's cost for other fringe benefits (ASM)	2,886
Capital Expenditures and Rental: Buildings & Other Structures (new and used) (ASM)	928
Capital expenditures for buildings and other structures (ASM)	818
Rental payments or lease payments for buildings and other structures (ASM)	110
Capital Expenditures and Rental: Machinery and Equipment (ASM)	4,094
Capital expenditures for machinery and equipment (ASM)	3,983
Capital expenditures for automobiles, trucks, etc. for highway use (ASM)	41
Capital expenditures for computers and peripheral data processing equipment (ASM)	98
Capital expenditures for all other machinery and equipment (ASM)	3,844
Rental payments or lease payments for machinery and equipment (ASM)	111
Repair and maintenance services of buildings and/or machinery (ASM)	960
Refuse removal (including hazardous waste) services (ASM)	141
Materials, fuels, contract work, and energy expenditures	232,567
Cost of materials, packaging, etc. used (ASM)	231,434
772000: Cost of materials, packaging, etc. used (ASM)	231,434
970099: Cost of all other materials and components, p...(EC)	78,320
971000: Materials, ingredients, containers, and suppl...(EC)	507
32610001: Plastics products (including packaging, foam ...(EC)	3,029
32620100: Fabricated rubber products, excluding tires, ...(EC)	1,974
32721000: Glass and glass products...(EC)	7,661
33200091: Miscellaneous fabricated metal products (excl...(EC)	2,131
33341500: Air-conditioning and warm air heating equipme...(EC)	1,705
33361810: Gasoline and gas-gasoline engines (excluding ...(EC)	2,584
33431000: Audio and video equipment, including radios, ...(EC)	2,640
33441310: Semiconductors, including transistors, diodes...(EC)	4,278
33451400: Motor vehicle clusters, meters, and gauges, e...(EC)	957
33631000: Motor vehicle gasoline engines and engine par...(EC)	16,133
33632000: Motor vehicle electrical and electronic equip...(EC)	6,676
33633000: Shocks, struts, and other motor vehicle suspe...(EC)	5,113
33635000: Motor vehicle transmission and power train pa...(EC)	17,736
33636000: Motor vehicle seating and interior trim...(EC)	12,134
33637000: Automotive stampings (including body parts, h...(EC)	57,678
33639000: Other motor vehicle parts, including air-cond...(EC)	10,181

Cost of resales (ASM)	243
Purchased fuels and electricity (ASM + MECS)	716
Cost of purchased fuels consumed (ASM) (ASM + MECS)	192
Indirect Uses-Boiler Fuel (ASM + MECS)	31
Conventional Boiler Use (ASM + MECS)	19
CHP and/or Cogeneration Process (ASM + MECS)	14
Direct Uses-Total Process (ASM + MECS)	82
Process Heating (ASM + MECS)	73
Process Cooling and Refrigeration (ASM + MECS)	2
Machine Drive (ASM + MECS)	3
Electro-Chemical Processes (ASM + MECS)	0
Other Process Use (ASM + MECS)	3
Direct Uses-Total Nonprocess (ASM + MECS)	76
Facility HVAC (f) (ASM + MECS)	67
Facility Lighting (ASM + MECS)	0
Other Facility Support (ASM + MECS)	2
Onsite Transportation (ASM + MECS)	3
Conventional Electricity Generation (ASM + MECS)	0
Other Nonprocess Use (ASM + MECS)	1
End Use Not Reported (ASM + MECS)	1
Cost of purchased electricity (ASM) (ASM + MECS)	524
Indirect Uses-Boiler Fuel (ASM + MECS)	3
Conventional Boiler Use (ASM + MECS)	3
CHP and/or Cogeneration Process (ASM + MECS)	0
Direct Uses-Total Process (ASM + MECS)	334
Process Heating (ASM + MECS)	48
Process Cooling and Refrigeration (ASM + MECS)	36
Machine Drive (ASM + MECS)	229
Electro-Chemical Processes (ASM + MECS)	3
Other Process Use (ASM + MECS)	21
Direct Uses-Total Nonprocess (ASM + MECS)	181
Facility HVAC (f) (ASM + MECS)	93
Facility Lighting (ASM + MECS)	60
Other Facility Support (ASM + MECS)	15
Onsite Transportation (ASM + MECS)	6
Conventional Electricity Generation (ASM + MECS)	0
Other Nonprocess Use (ASM + MECS)	3
End Use Not Reported (ASM + MECS)	0
Cost of contract work (ASM)	174
Services, computer hardware, software, and other expenditures	13,345
Communication services (ASM)	52
Computer hardware, software, and other equipment	245
Expensed computer hardware and other equipment (ASM)	136
Expensed purchases of software (ASM)	109
Professional, technical, and data services	1,235
Data processing and other purchased computer services (ASM)	53
Purchased professional and technical services (ASM)	1,182
Temporary staff and leased employee expenses (ASM)	956
Advertising and promotional services (ASM)	52
Taxes and license fees (ASM)	290
All other operating expenses (ASM)	10,515
Shipments	301,958
Expenditures	274,473

Net inventories shipped (see Inventory and Flow Time tab)	364	
Depreciation	9,642	
Net income	17,479	
Value Added (ASM)	68,898	
Flows, Stocks, and Capacity		
Inventories (ASM)		
Total inventories, beginning of year (ASM)	12,491	
Finished goods inventories, beginning of year (ASM)	1,885	
Work-in-process inventories, beginning of year (ASM)	3,131	
Materials and supplies inventories, beginning of year (ASM)	7,475	
Total inventories, end of year (ASM)	11,731	
Finished goods inventories, end of year (ASM)	1,489	
Work-in-process inventories, end of year (ASM)	3,035	
Materials and supplies inventories, end of year (ASM)	7,208	
Plant Capacity Utilization (SPCU)	62	Percent
Plant Hours Per Week (SPCU)	77	Hours per Week
Flow time		
Materials and supplies inventories (ASM)	9.6	Days
Work-in-process inventories (ASM)	4.0	Days
Work-in-process inventories - Downtime (ASM + SPCU)	2.2	Days
Work-in-process inventories - Operation time (ASM + SPCU)	1.8	Days
Finished goods inventories (ASM)	2.2	Days

- Suppliers of services, computer hardware, software, and other costs
- Refuse removal, intermediate goods, and recycling
- Design, production, inventory, shipments, and net income
- Machinery, structures, and compensation
- Suppliers of materials
- Repair of the machinery and structures

Appendix B. Hypothesis and Impact Data Variables (Example)

The hypothesis and impact data does not have the established standardization of classifications that exist for the opportunity map; thus, developing it will require variable identification and negotiation. However, there is one relevant classification, the Industrial Investment Classification (IIC) System being developed at ASME. It is a classification of investments for advancing manufacturing efficiency and productivity and facilitates data collection and analysis of investments to further understand the types of investments that have the highest returns and where there might be opportunities. It is a five-digit classification system with 484 categories of investments for efficiency improvements. At the highest level of the hierarchy there are eight categories (see Table B. 1). The IIC is one of the variables that could be included in the hypothesis and impact data.

The other variables that need to be collected are those that facilitate learning and advancing knowledge for increasing impact. This likely includes variables that characterize a project (nature of work), the industry that it might affect, the costs/losses (user or manufacturer) that are affected, and other factors that affect the impact of a project. Table B. 2 provides an illustrative set of variables for the hypothesis and impact data collection, including those collected at the initial phases, mid-progress, and post-project. Because projects can have different types of products, the variables might also break the project into products or product groups. Table B. 2 has a set of subcategories for products.

Table B. 1: Industrial Investment Classification (IIC) System (single digit level)

10000 (2.7000): Buildings and Grounds - Alterations to buildings and grounds, including lighting, space conditioning, ventilation, building envelope, and other items.
20000: Machinery and Facility Systems - Alterations to machinery and facility systems, including combustion systems, cooling/drying/thermal systems, electrical power systems, motor systems, industrial design, operations, and alternative energy usage for these systems.
30000: Product Design and Data Management - Alterations to product design and data management, including data collection, storing, utilization, modeling, and design data.
40000: Labor Optimization - Alterations in labor, including distribution, utilization, training, management, and scheduling.
50000: Material Management - Alterations in the management of materials, including transport, storage, material selection, and waste.
60000: Utilities and Administrative Activities Management - Alterations to utilities and administrative activities, including changing providers and servicers.
70000: Natural and Man-Made Hazard Risk Management - Alterations that mitigate the risks of natural and man-made hazards.
90000: Other - Other investments or alterations not elsewhere specified.

Table B. 2: Example Illustrative Variable Set for Hypothesis and Impact Data

Variable	Time of Assessment	Data Type
Project Title	Initial	Narrative
Project description	Initial	Narrative
Industrial Investment Classification code	Initial	Discrete Categorical
Relevant NAICS codes	Initial	Discrete Categorical
NIST Topic Taxonomy	Initial	Discrete Categorical
Description of potential impact at full maturity	Initial	Narrative
Description of potential impact at 3-5 years	Initial	Narrative
Manufacturing Cost Classification (to be developed)	Initial	Discrete Categorical
Project type (e.g., standard, collaboration, education)	Initial	Discrete Categorical
Hypothesized dollar Impact Estimate (developed with opportunity map when possible): Mature adoption	Initial	Numerical
Estimated time to reach full impact maturity (years):	Initial	Numerical
Hypothesized non-dollar impacts	Initial	Numerical
Hypothesized non-dollar impact units	Initial	Narrative
Impact assessment approach	Initial	Narrative
Dissemination, Adoption, and Implementation plan	Initial	Narrative
Leading Indicators (e.g., downloads, page visits, citations)	During execution	Narrative
Lagging Indicators (e.g., impact analysis)	During execution	Narrative
Probability of reaching hypothesized impact at maturity		
Estimated Impact at Maturity		
Lessons Learned		
Product Assessments		
Product type	As developed	Discrete Categorical
Description	As developed	Narrative
Industrial Investment Classification code	As developed	Discrete Categorical
Relevant NAICS codes	As developed	Discrete Categorical
Impact description	As developed	Narrative
Hypothesized diffusion path (products)	As developed	Narrative
Hypothesized time to first adoption (products) – Example	As developed	Numerical
Total potential adopters	As developed	Numerical
Hypothesized average impact per adopter (product)	As developed	Numerical
Hypothesized annual dollar Impact Estimate (developed with opportunity map when possible): Mature adoption	As developed	Numerical
Dissemination, Adoption, and Implementation plan	As developed	Narrative

	Estimated impact per adopter (via case studies)	During execution	Numerical
	Estimated time to reach full impact maturity (years):		
	Hypothesized non-dollar impacts		
	Hypothesized non-dollar impact units		
	Extent that dissemination/impact are on track (periodic assessment)	During execution	Narrative
	Probability of reaching hypothesized impact at maturity	During execution	Numerical
	Estimated Impact at Maturity	Post project	Numerical
	Lessons Learned	As developed	Narrative

Appendix C. Project Entry Example: Data Infrastructure for Critical Material Recovery

Below is an example entry for the "Data Infrastructure for Critical Material Recovery" project from NIST's Engineering Laboratory that would go in a Hypothesis and Impact database. Several items are approximated based on individual insight; however, these estimates are often supported by data. For example, the variable "Hypothesized Dollar Impact Estimate (developed with opportunity map when possible): Mature adoption" is calculated by estimating raw material costs using economic Input-Output (IO) analysis. This data is then combined with expert judgment, assuming that approximately 1 % to 3 % of the raw material cost could be reduced through this project.

By breaking the estimate into manageable subcomponents and combining data with intuition, the accuracy of the hypothesized impact improves compared to relying on intuition alone. The IO data provides a maximum and minimum range, within which the intuition-based estimate is bounded (in this case, 1 % to 3 %). A similar approach is applied to supply chain disruption: estimated losses from an external report are combined with a judgment that 0.10 % to 0.25 % of the losses could be mitigated by the project. Waste recovery was calculated in a similar fashion.

This method not only sets bounds on the impact but is also testable. Unlike qualitative assessments, these estimates can be systematically evaluated and improved over time. Furthermore, by documenting a testable approach, others can learn from the estimates, enabling compounding growth in impact across projects.

Table C. 1: Project Entry Example: Data Infrastructure for Critical Material Recovery

Variable	Value
Project Title	Data Infrastructure for Critical Material Recovery
Project description	Development of datasets, metrics, indicators, predictive models, and technical standards to strengthen materials recovery data infrastructure and critical material supply chains, enabling regenerative manufacturing systems and improved U.S. manufacturing competitiveness.
Industrial Investment Classification code	53300: Material Recovery - Alterations in how or the extent to which waste materials are reused. 31100: Data Standards - Alteration in the standards for data, including formatting and organization. 34300: Supply Chain Modeling - Alterations in the modeling of supply chains.
Relevant NAICS codes	334 – Computer and Electronic Product Manufacturing 325 – Chemical Manufacturing (including plastics) 331 – Primary Metal Manufacturing 335 – Electrical Equipment, Appliance, and Component Manufacturing 562920 – Materials Recovery Facilities 423930 – Recyclable Material Merchant Wholesalers
NIST Topic Taxonomy	Manufacturing: Product Data Manufacturing: Sustainable Manufacturing
Description of potential impact at full maturity	Reduced Critical Material Procurement Costs Reduced Supply Chain Disruption Costs Increased Secondary Material Value Capture Reduced Waste Disposal and Environmental Impact Decreased capital costs (increased manufacturing system efficiency) Reduced Transaction Costs in Material Exchanges Loss Reduction thru Increased Resilience of Critical Material Supply Chains
Description of potential impact at 3-5 years	Within 3–5 years, this project will enable measurable reductions in critical material procurement and supply chain disruption costs by improving data infrastructure, predictive modeling, and standards for material recovery and traceability. Early adopters are expected to realize increased use of recovered materials, higher recovery rates, and lower waste disposal and coordination costs. By strengthening the data foundations for regenerative manufacturing, the project lays the groundwork for U.S. manufacturers to be more competitive in resilient supply chains and critical-material-intensive industries.

Manufacturing Cost Classification (to be developed)	Raw material Losses due to supply chain disruption (materials, downtime, delays) Inventory and carrying costs Disposal costs Production scrap and rework Regulatory costs Transaction costs R&D and product design costs
Project type (e.g., standard, collaboration, education)	Standards Development Collaborative (Academia, Industry, Government, SDOs) Infrastructure / Data Platform Project
Hypothesized dollar Impact Estimate (developed with opportunity map when possible): Mature adoption	Raw material cost reduction: 1% - 3 % of $77Billion cost (est. using IO) Supply chain disruption risk: 0.1% - 0.25% of supply chain losses $184Billion (J.S. Held 2025) Waste value recovery: 2% to 5% of $66 Billion (Nichols 2025) Total: $2.4Billion to $6.4 Billion Annually
Estimated time to reach full impact maturity (years):	10 to 15 years
Hypothesized non-dollar impacts	-
Hypothesized non-dollar impact units	-
Impact assessment approach	Early-Mid Stage: Product-level impact measured via proxies (e.g., downloads, citations, and page visits) coupled with value assessment and case studies Maturity stage: Survey and industry assessment
Dissemination, Adoption, and Implementation plan	ISO TC/323 participation IEEE PV3513 contributions ASTM E60 standards engagement Peer-reviewed publications Public datasets and technical reports Workshops and industry engagement events Collaboration with DOE and academic partners Conference presentations Digital dissemination via NIST platforms
Leading Indicators (e.g., downloads, page visits, citations)	Downloads of datasets, reports, and tools Citations in standards drafts References in policy documents GitHub/tool repository activity (if applicable) Mentions in industry guidance documents Adoption inquiries from OEMs/SMEs

Lagging Indicators (e.g., impact analysis)	Measured reductions in critical material acquisition costs Increased recovery rates (%) Reduction in supply disruption costs for raw materials Adoption of developed standards Case-study ROI from adopters Trade volume increases in standardized recovered materials
Product Assessments	
Probability of reaching hypothesized impact at maturity	60%
Estimated Impact at Maturity	TBD
Lessons Learned	TBD

Appendix D. Project Entry Example: Digital Twins for Advanced Manufacturing with the ISO 23247 Product

Below is an example entry for the "Digital Twins for Advanced Manufacturing" project from NIST's Engineering Laboratory that would be included in a Hypothesis and Impact database. Several items are approximated based on expert judgment, but these estimates are often supported by data.

The "Hypothesized Dollar Impact Estimate (developed with opportunity map when possible): Mature adoption" relies on two intuition-based variables. The first is AdoptionInfluence, representing the proportion of digital twin impact applicable to this project. The second is DependencyFactor, representing the proportion of impact that depends on standards or technologies provided by this project. These two factors are combined with the estimated impact of digital twins from NIST AMS 100-61 (Thomas 2024). One might argue that this approach introduces estimation error; however, such critiques often fail to compare it with prior methods that rely entirely on unbounded intuition and therefore likely introduce greater—and less transparent—error. More importantly, the approach presented here enables progressively greater accuracy where purely qualitative assessments cannot.

For the product assessment (ISO 23247), the total potential adopters are assessed using manufacturing industry data from the Census Bureau, adjusted by the proportion of adopters who could realistically adopt digital twins and the share of those adoptions that depend on ISO 23247 (see item ID 29 in Table D. 1). Per-adopter impact is estimated by applying the 0.73% potential revenue reduction from digital twins (NIST AMS 100-61) to the average revenue of medium and large manufacturers, calculated from payroll data and an industry-standard payroll-to-revenue ratio, divided by the number of establishments (see item ID 30 in Table D. 1). Combining these two (adopters and per adopter impact) results in an estimate of $1.2 billion in impact annually (see Approach 2 in item ID 31 in Table D. 1).

To gauge how realistic the magnitude of the $1.2 billion estimate might be, we can examine the estimated potential impact of another program such as NIST's STEP program, an international suite of standards for representing and exchanging digital product data across different CAD/CAM systems. In 2001, it was estimated that the potential impact of step was $928 million (2001$) annually; however, at the time, only 17 % of the potential benefits were being realized. Moreover, this suggests that the $1.2 billion estimated impact from digital twins is within the range of that estimated for other projects in manufacturing.

By breaking the estimate of impact into manageable subcomponents and combining data with expert judgment, similar to that of Appendix C, the accuracy of the hypothesized impact improves compared to relying on intuition alone. Setting bounds on variables and using a testable approach allows estimates to be systematically evaluated and improved over time. Documenting this methodology also enables others to learn from the estimates, facilitating compounding growth in impact across projects.

Figure D. 1 estimates the diffusion path for ISO 23247. The current path is linear, as shown in the lower right-hand corner of the figure. This suggests steady, predictable growth, where downloads (used as a proxy for adoption) increase by roughly the same amount each period, resulting in gradual expansion, which seems common for foundational standards.

Ideally, however, the path transitions to an S-curve diffusion pattern, where adoption generates additional adoption through network effects and increased awareness, as shown in the figure. It should be noted that Figure D. 1 is illustrative. A more rigorous estimate of true adoption would likely require a composite index incorporating multiple indicators, such as publications, downloads, citations, mentions, page views, Google search rankings, and related metrics. To accelerate adoption, NIST might consider publishing implementation guides, offer certification or validation, fund high-visibility pilot programs, showcase success stories, and document return-on-investment.

To increase accuracy in predictions, an organization likely needs to investigate realized impact. Additionally, an organization would likely prefer not to wait until the impact of a project is fully realized before evaluating impact. In this instance, bibliometric analysis can again be used as a proxy for adoptions, providing data supported conclusions about adoption. Case studies can be used to not only demonstrate the value of the standard to manufacturers, but also to estimate the average impact. Combining these two things (i.e., adoptions and average impact) can reveal the real-time impact of the products/project. Once diffusion is believed to have matured, selected projects or products can be evaluated more rigorously to estimate realized impact. These estimates can then be used to calibrate and improve the less rigorous real-time analyses. Thus, a proposed method might include the following:

- Bibliometrics → proxy for adoption
- Case studies → proxy for average impact
- Combine → real-time impact estimate
- Later rigorous evaluation → recalibration

In this way, the framework is applied to integrate prediction, measurement, and recalibration into a continuous system for maximizing and demonstrating manufacturing impact.

Table D. 1: Project Entry Example: Digital Twins for Advanced Manufacturing and ISO 24247

ID	Variable	Value
1	Project Title	Digital Twins for Advanced Manufacturing
2	Project description	This project develops measurement science, technical guidance, reference implementations, and standards to enable the reliable, interoperable, and trustworthy use of digital twins in advanced manufacturing. The project focuses on verification, validation, and uncertainty quantification (VVUQ), lifecycle integration through digital threads, and interoperability standards such as ISO 23247 and MTConnect. Through standards development, testbeds, and reference implementations, the project reduces barriers to digital twin adoption—particularly for small and medium-sized manufacturers—and supports the development of a competitive digital twin marketplace in the United States.
3	Industrial Investment Classification code	34000: Modeling - Alterations in the modeling, including that for products, production, and supply chains. 34100: Product Modeling - Alterations in the modeling of products, including adoption of three-dimensional modeling, digital twins, and other methods of modeling. 34200: Production Process Modeling - Alterations in the modeling of production processes. 34300: Supply Chain Modeling - Alterations in the modeling of supply chains. 34900: Other Modeling - Alterations in other modeling.
4	Relevant NAICS codes	31–33 – Manufacturing (cross-sector applicability)
5	NIST Topic Taxonomy	Digital Twins Internet of Things (IoT)
6	Description of potential impact at full maturity	-System-level impacts include increased U.S. manufacturing productivity, improved global competitiveness, and strengthened domestic supply chain resilience. -Reduced manufacturing costs through improved process optimization, predictive maintenance, and reduced downtime -Improved product quality and yield, resulting in reduced manufacturer losses and reduced customer burden (e.g., longer lasting products) -Faster product development cycles via lifecycle integration, resulting in product benefits being realized sooner -Reduced integration costs due to interoperability standards -Increased trust in simulation-driven decision-making through quantified uncertainty -Lower barriers to SME participation in digital manufacturing ecosystems -Emergence of a robust digital twin marketplace (technology providers + adopters)

7	Description of potential impact at 3-5 years	-Adoption of ISO 23247 parts and VVUQ guidelines by leading manufacturers-Increased interoperability among early adopters-Measurable reductions in integration time for digital twin deployment-Increased SME engagement via reference implementations and open tools-Expanded standards participation by U.S. industryGrowth in pilot digital twin use cases validated via the NIST testbed
8	Manufacturing Cost Classification (classification to be developed)	Primary cost impact channels likely include: Capital utilization efficiency Downtime reduction Engineering labor productivity Quality cost reduction (scrap, rework, warranty) Integration and IT implementation cost reduction Maintenance cost reduction Inventory and lifecycle management cost efficiency
9	Project type (e.g., standard, collaboration, education)	Standards Development Research Publications
10	Hypothesized dollar Impact Estimate (developed with opportunity map when possible): Mature adoption	Attributable Impact=DT Impact * AdoptionInfluence * DependencyFactor where: DT Impact = : Estimated potential annual savings from digital twins from NIST AMS 100-61. AdoptionInfluence = Proportion of digital twins that apply to the issues addressed by this project. DependencyFactor = The proportion of digital twins that is facilitated by a standard/technology for implementation. Low: 37.9 * 0.15 * 0.60 = $3.4 billion Annually High: 37.9 * 0.35 * 0.75 = $9.9 billion Annually
11	Estimated time to reach full impact maturity (years):	10-15 Years
12	Hypothesized non-dollar impacts	Not assessed
13	Hypothesized non-dollar impact units	Not assessed
14	Impact assessment approach	
15	Dissemination, Adoption, and Implementation plan	Active leadership in ISO 23247 and ASME VVUQ standardsNIST Digital Twin Testbed demonstrationsReference implementations (robot arm composition, digital thread pipeline)GitHub release of MTConnect adapters and toolsIndustry workshops and ecosystem conveningsCollaboration with OEMs and standards development organizationsPublications (NIST AMS, journal articles, book chapters)
16	Leading Indicators (e.g., downloads, page visits, citations)	Adoption index coupled with case studies
17	Lagging Indicators (e.g., impact analysis)	Impact and adoption analysis

18	Probability of reaching hypothesized impact at maturity	60%
19	Estimated Impact at Maturity	TBD
20	Lessons Learned	TBD
21	Product Assessments (single or group)	
22	Product type	ISO 23247: Automation systems and integration - Digital twin framework for manufacturing along with associated suite of standards/publications
23	Description	ISO 23247 is an international standard series that defines a comprehensive framework for implementing digital twins in manufacturing, enabling real-time, bidirectional synchronization between physical manufacturing assets (personnel, equipment, materials) and their virtual counterparts.
24	Industrial Investment Classification code	34000: Modeling - Alterations in the modeling, including that for products, production, and supply chains. 34100: Product Modeling - Alterations in the modeling of products, including adoption of three-dimensional modeling, digital twins, and other methods of modeling. 34200: Production Process Modeling - Alterations in the modeling of production processes. 34300: Supply Chain Modeling - Alterations in the modeling of supply chains. 34900: Other Modeling - Alterations in other modeling.
25	Relevant NAICS codes	31–33 – Manufacturing (cross-sector applicability)
26	Impact description	Enables faster, lower-risk implementation and seamless interoperability across machines, systems, and vendors. This accelerates adoption and allows digital twin solutions to scale across plants, improving uptime, throughput, quality, and energy efficiency. Over time, the standard potentially generates multi-billion-dollar annual productivity gains by reducing integration costs, increasing innovation, and strengthening manufacturing competitiveness and resilience.
27	Hypothesized diffusion path (products)	S-Curve
28	Hypothesized time to first adoption (products)	1-3 years

29	Total potential adopters	DT Compatible * StandardsDependency * Med and Lg Manufacturers Where: DT Compatible = Proportion of manufacturers where digital twins are cost effective. Estimated as 50% of medium manufacturers and 75% of large manufacturers. StandardsDependency = The proportion of digital twins that is facilitated by a standard/technology for implementation. Med and Lg Manufacturers = The number of medium manufacturing facilities or the number of large manufacturing facilities. Medium manufacturers (50-499 employees): 0.5 * 0.05 * 44 008 = 1100 Large manufacturers (500+ employees): 0.75 * 0.05 * 3384 = 127 Total: 1100 + 127 = 1227
30	Hypothesized average impact per adopter per year (product)	(Payroll * PayrollPerShipment) * (1/Establishments) * DTSavingRatio Where: Payroll = Annual dollar value for payroll from the County Business Patterns Survey. PayrollPerShipment = The annual dollar value of shipments divided by payroll from the Economic Census Establishments = The number of establishments from the County Business Patterns Survey DTSavingRatio: Estimated potential annual savings from digital twins divided by shipments, estimated in Table 5.10 of NIST AMS 100-61. $758.9B * 8.67 * (1/47 400) * 0.0073 = $1.0 million per adopter per year
31	Hypothesized annual dollar Impact Estimate (developed with opportunity map when possible): Mature adoption	Approach 1: DT Impact * AutomationShare of DT * StandardsDependency Where: DT Impact = Estimated potential annual savings from digital twins from NIST AMS 100-61. AutomationShare of DT = Share of digital twins savings that is related to automation. StandardsDependency = The proportion of digital twins that is facilitated by a standard/technology for implementation. $37.9 billion * 0.25 * 0.05 = $474 million annually Approach 2: #Adopters[item ID 29] * AvgImpact[item ID 30]: 1227 * $1.0 million = $1.2 billion annually Estimate: Between $474 million and $1.2 billion
32	Dissemination, Adoption, and Implementation plan	Case studies demonstrating value How-to-adopt guides Website Adoption Index to model adoptions
33	Estimated impact per adopter (via case studies)	TBD

34	Estimated time to reach full impact maturity (years):	10-15 years
35	Hypothesized non-dollar impacts	Not assessed
36	Hypothesized non-dollar impact units	Not assessed
37	Hypothesized extent that dissemination/impact is on track (periodic assessment)	1048 downloads with 11 citations: https://doi.org/10.1145/3652620.3688250 Assuming 1 adoption per 400 downloads = 2-3 est. adopters 5103 Downloads with 123 citations ASM 400-2 Assuming 1 adoption per 400 downloads = 12-13 est. adopters Average: 7-8 currently implementing or adopted. Early-stage diffusion. Currently, evidence suggests adoption to be linear.
38	Probability of reaching hypothesized impact at maturity	70%
39	Estimated Impact at Maturity	TBD
40	Lessons Learned	TBD

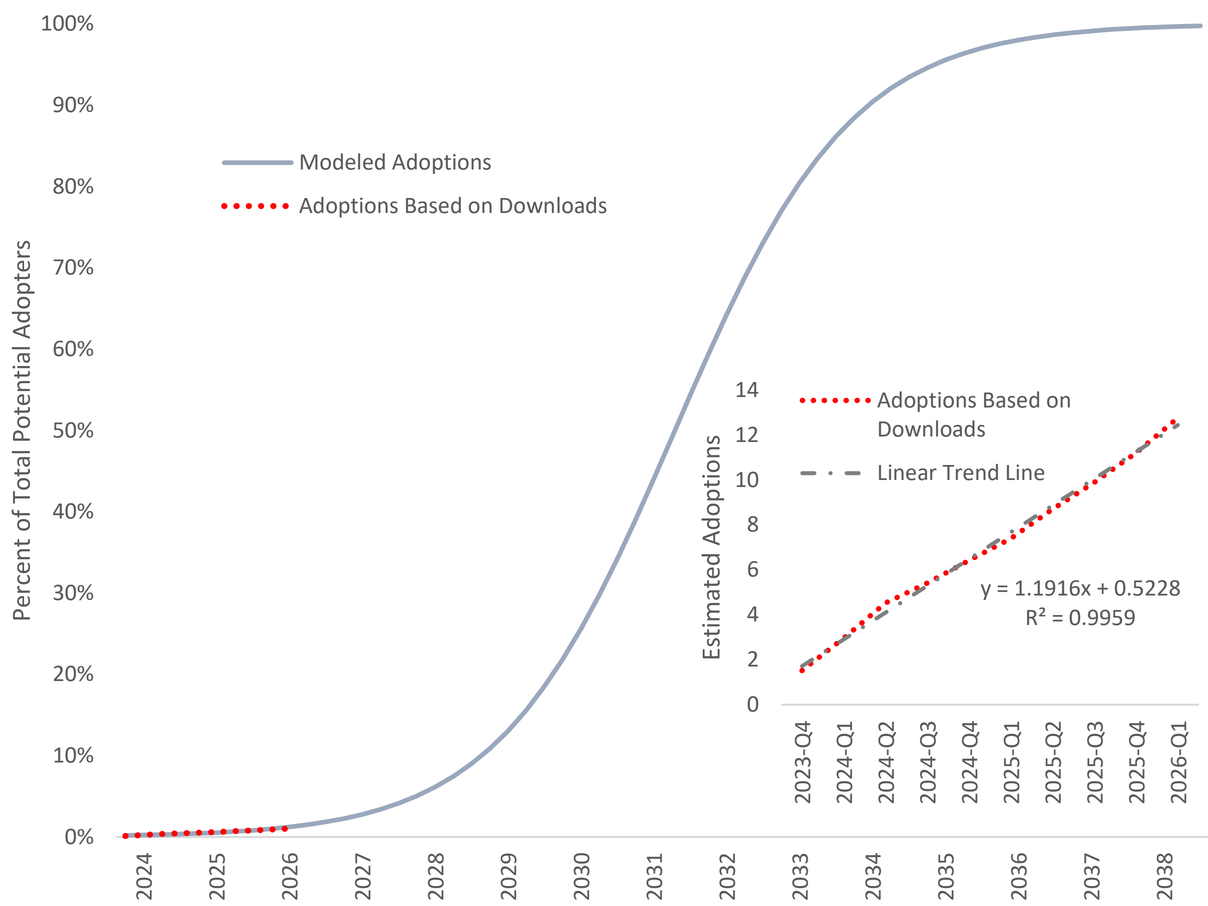


Figure D. 1: Illustrated Diffusion of ISO 23247 using Downloads of ASM 400-2 (Shao 2021) as a Proxy