Cite as: Bergman, M. K. The Situation. in *A Knowledge Representation Practionary: Guidelines Based on Charles Sanders Peirce* (ed. Bergman, M. K.) 45–64 (Springer International Publishing, 2018). doi:10.1007/978-3-319-98092-8_3

Official site: https://link.springer.com/book/10.1007/978-3-319-98092-8

Full-text: http://www.mkbergman.com/publications/akrp/chapter-3.pdf

Abstract: Knowledge representation is a primary driver for using computers as a means to improve the economic well-being of all peoples. Solow, a student of Schumpeter, had the insight in two papers in the 1950s, for which he won a Nobel prize, that technological change is the 'residual' left over from empirical growth once we remove the traditional inputs of labor and capital, what we now call total-factor productivity . Romer's subsequent work internalized this factor as a function of information and knowledge, what became the endogenous growth model.

The need to organize better and present information came with its increased volume. We can define intangible assets as private expenditures on assets that are intangible and necessary to the creation and sale of new or improved products and processes, including brands, designs, software, blueprints, ideas, documents, know-how, artistic expressions, recipes, and the like. Some 25% of the annual trillions of dollars spent on information creation lend it to actionable improvements. If we are to improve our management and use of information, we need to understand how much value we routinely throw away.

3

THE SITUATION

Since Adam Smith, a focus of economics has been its attempt to explain the basis of growth. This emphasis is not surprising since the birth of the field of economics corresponded to a historically unprecedented inflection point in economic growth. Smith ascribed this growth to productivity resulting from the division of labor, using his famous example of the pin factory. However, it is only within the past fifty years or so that economists have begun unpacking growth from the other factors of production.[1] In this chapter, we talk specifically about the role of information in growth, and how it may contribute further.

Growth is a percent increase from a prior state. Economic growth compounded over a period has the virtuous reward of resulting in increased wealth. We measure economic growth through such means as revenues (for the individual firm) or GDP (for regions or countries). Net worth (for the firm or individuals) or GDP per capita measure the wealth associated with the current stock of economic goods at any given point in time. Such measures, while useful proxies, still do not account for other changes in comfort, convenience, freedom, choice, leisure, and mobility that may accompany growth and transcend the material. On the other hand, growth may also create 'externalities,' some of which may be negative such as pollution or traffic congestion. Wealthier societies have tended to regulate against such harmful effects over time. We should include all of these factors in the value equation.

Throughout history, we have seen discontinuities in growth (and then wealth) for individuals, families, firms, industries, cities, regions, and nations. Growth thus has immense importance across the entire economic spectrum. This chapter makes the argument that access to information — and impediments to that — are a significant determinant of wealth and economic growth. Better knowledge representation using computers is one means to improve the economic well-being of all peoples.

INFORMATION AND ECONOMIC WEALTH

If we toil, year by year, doing the same activity, like growing wheat, and we gain the same harvest for the same inputs of labor and land, we are not surprised. Sometimes, the weather or rainfall patterns may differ, or we may have more children helping us in the fields, or a mule to help plow. Money helps us buy more of the essential inputs, maybe more land, seed or mules, or the comfort to have more children. These are the traditional factors of production: that is, land, capital, and labor.

If we add more of these factors to the mix, we still understand we have merely tweaked the regular basis of our wheat production. Differences in the amount of these factors of production, throughout most of human history, are what accounted for the differences between rich and poor, landlord and serf. If by having more land or children, we are now able to feed more people, we are by definition more wealthy, and if we can accumulate more of this wealth, we can leverage these standard factors even more. Control and exploitation have been common paths to much wealth creation.

These factors are pretty easy to observe and track. We intuitively understand that more inputs of labor, land or capital can result in growth, but one that feels and appears somewhat fixed based on the change in these inputs. This kind of growth has a more-or-less trending return based on changes in these inputs. These types of inputs may also be subject to diminishing returns, wherein adding more of a given factor reduces payoff. For example, adding more fertilizer to the wheat crop produces less per unit output yield after some optimum, eventually lowering yields by chemically burning the crop. Alternatively, while a computer increases the productivity of an individual worker, giving her more computers may degrade her overall performance.

Historical Breakpoints

Still, a different kind of growth is not constrained to a fixed return based on inputs. Perhaps we have a neighbor who raises more wheat, possibly on drier, more marginal land, or with less water or fertilizer. His yield exceeds our own. These differences occur because our neighbor is doing something different and is producing more given his inputs.

Many of us (now) older people can recall grandparents talking about their first sight of a car or airplane. In my own life (born 1952) I can remember the first instance of color TVs, electronic calculators, personal computers, the Internet, and smartphones. The fact is, the pace of development and technological change is now so constant that its very existence seems unremarkable — part of the daily background noise. For 99.5% of human history, this has not always been so.[2]

In our daily lives we are bombarded by statistics: quarterly economic growth rates, sports scores, weather precipitation likelihoods and daily temperatures, in a constant and thus background stream of numeric immersion. It is interesting to note that *statistics* (originally derived from the concept of information about the *state*) only began in France in the 1700s. The first actual population census, as opposed to enumerations in biblical times or the land and tax recordings of the Domesday Book in England in 1086, occurred in Spain in that same century, with the United States being the first country to set forth a decennial census beginning around 1790.[3]

Because the state collected no data — indeed, the idea of data and statistics did not exist — attempts in our modern times to re-create economic and population assessments in earlier centuries are a heroic exercise, laden with estimation. Nonetheless, the renowned economic historian Angus Maddison and his team, written in some definitive OECD studies, prepared economic and population growth estimates for the world and various regions going back to AD 1.[3 4]

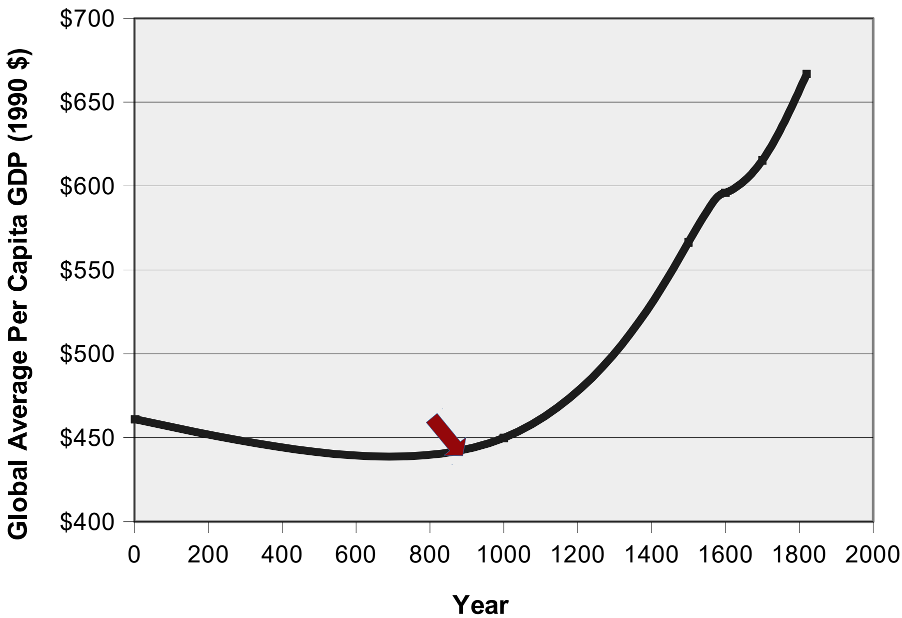

Figure 3-1: Global Average per Capita GDP, 1 - 1800 AD

Through at least 1000 AD global economic growth per capita (as well as population growth) was approximately flat. Maddison estimated that a doubling of economic well-being per capita only occurred every 3000 to 4000 years. A historical shift occurred about 1000 AD when flat or negative growth began to accelerate slightly.[5] The growth trend looks comparatively impressive in *Figure 3-1*, but growth over a period of about 800 years to 1750 AD only totals 45%, a doubling of per capita wealth that still requires 1000 to 2000 years. These are annual growth rates about *30 times lower* than today, which, with compounding, prove anemic indeed over such long historical periods.

By 1820 or so onward, this doubling accelerated at warp speed to every 50 years or so, as shown in *Figure 3-2*. Historically flat income averages skyrocketed, as this famous figure showing global changes in per capita (person) GDP from Maddison illustrates.[2 3 3]

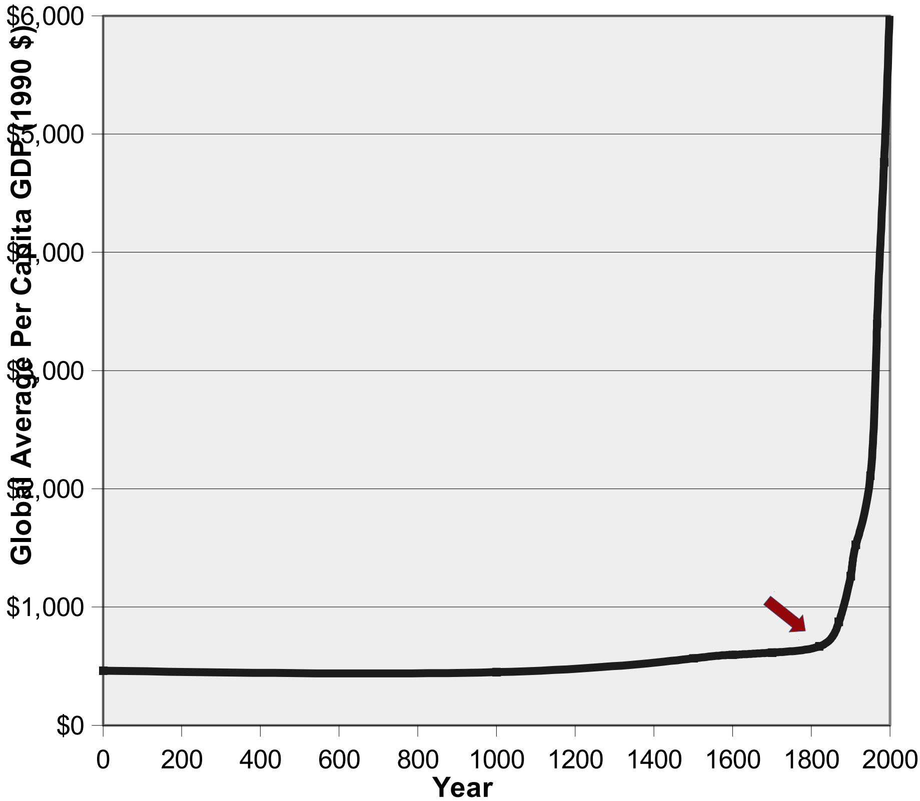

Figure 3-2: Growth Skyrockets About 1820 AD

William Nordhaus captured a similar discontinuity looking at the price of light, normalized according to the labor effort needed to obtain 1000 lumens. His study, too, shows an exponential decrease in the price of lighting beginning about 1800.[6] More recent trends show an additional upward blip in growth shortly after the turn of the 20th century, corresponding to electrification, but then a more massive discontinuity beginning after World War II, as next shown in *Figure 3-3*. Growth rates accelerated to a doubling of wealth every 40 to 45 years. These comparatively abrupt changes in growth rates and concomitant changes in wealth were more than two orders of magnitude higher than what had been experienced before in human history, and thus garnered the attention of economists and economic historians as never before. Something huge *did* happen in the early 1800s.

Since their occurrence, many have attributed the inflection points in growth rates of the 1820s and 1950s to 'technological change,' but the specific causes of this change lack consensus. The prior era of the Enlightenment suggested some fundamental shift in thinking. Had a notable transition occurred in the mid-1400s to 1500s it would have been obvious to ascribe more modern economic growth trends with the availability of the printing press. While the printing press had massive effects as Elizabeth Eisenstein has shown,[7] the empirical record of changes in economic growth is not coincident with its adoption. The closer concurrence with the Industrial Revolution lent credence to the adoption of machines, prime movers and the harnessing of energy as a likely explanation. Cultural and religious factors have also been posited to explain why Britain and then the United States were the original centers of growth. Earlier, I noted the invisible hand of the market and division of labor and specialization, as advocated by Adam Smith. Education, followed by literacy, and support for basic and applied research have their advocates. Financial and banking innovations and the rule of law and patents and other intellectual property rights are other possible causes.

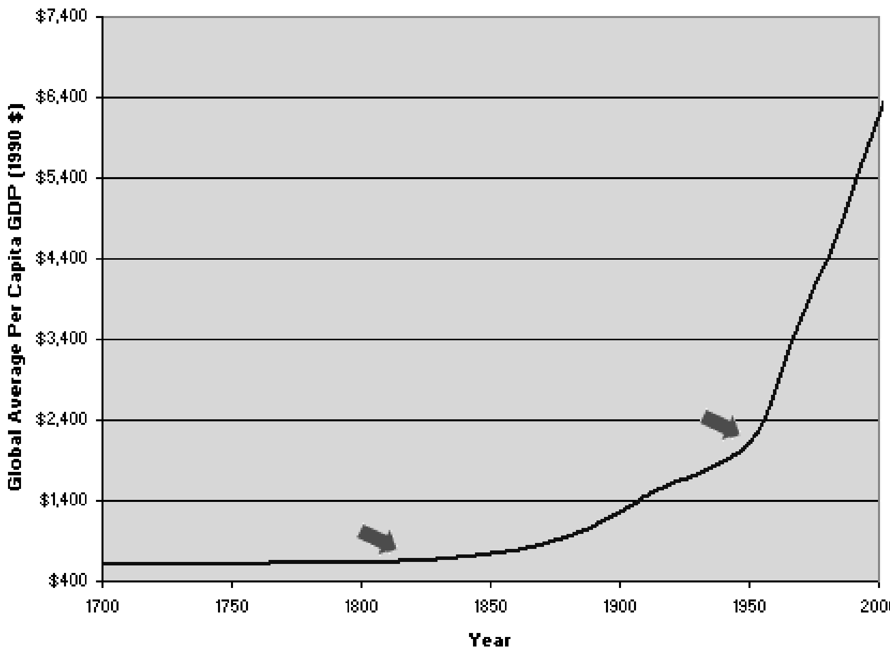

Figure 3-3: Growth Skyrockets Again in 1950s AD

Common sense tells us that all of these factors, and perhaps more, can all work as force multipliers to the traditional inputs to the economic function. However, I posit one element reigned supreme in these trends — information.

The X Factor of Information

Joel Mokyr provides a sweeping and comprehensive account of the period from 1760 (what he calls the 'Industrial Enlightenment') through the Industrial Revolution beginning roughly in 1820 and then continuing through the end of the 19th century.[8]

Mokyr centers his explanation for these growth changes on ‘useful knowledge,’ a phrase first coined by Simon Kuznets, as expanded upon by Michael Polanyi and others. Mokyr argues how *propositional knowledge,* the base of knowledge such as science, combines with *prescriptive knowledge*, the ‘recipes’ for applying knowledge, with *discovery*, to create the innovations that fueled the observed growth acceleration.

One of Mokyr’s key points is that both kinds of knowledge reinforce one another, with that time frame being a period of unprecedented growth in such knowledge. Another point, easily overlooked since ‘discoveries’ are the most visible, is that *techniques* and *practical applications* of knowledge provide a multiplier effect to knowledge growth. Mokyr notes that the inventions of writing, paper, and printing not only significantly reduced access costs but also materially affected human cognition, including the way people thought about their environment. Mokyr notes but does not adequately pursue, “In the decades after 1815, a veritable explosion of technical literature took place. Comprehensive technical compendia appeared in every industrial field.” Statements such as these in his outstanding book, *The Gifts of Athena*, hint at these fundamental drivers.

The industrialization that proceeded apace in the Americas and Europe is the engine that produced the wealth reflected in the earlier figures. However, from where did that mechanization and know-how come? It came from innovations and improved methods, for sure, but the more direct cause, I believe, was the broader dissemination of information. Our first inflection point in the 1820s roughly corresponds to the innovation of cheaper ‘pulp’ paper (and the genesis of ‘pulp’ fiction by serialized writers like Dickens or Hugo);[9] the second inflection point in the 1950s corresponds to the beginning use of the computer and digital information. Change was everywhere, and many factors were at work. It is hard to deny that information and greater access to it must surely have been central factors for increased innovation, literacy, and social and political change.

Knowledge and Innovation

Until the mid-1950s, economists ascribed the sources of this notable growth to ‘technological change’ and other vague factors, often argued in anecdotal ways. Empirical datasets were few and far between to test hypotheses, and quantitative means of reasoning over economic problems were only beginning. Growth theory was becoming an economic discipline in its own right.

Joseph Schumpeter, in *The Theory of Economic Development*, first published in 1911,[10] argued that innovation is central to economic growth and continuously disrupts the general equilibrium of market exchange. Innovation gains the firm a temporary monopoly status in which to charge higher rents, thereby providing an incentive for further innovation. Schumpeter’s emphasis on entrepreneurship and his popularization of ‘creative destruction’ recognized that new innovative market entrants might cause older firms to become obsolete. He tied these ideas into his basic views on business cycles, also driven by technological change. Innovation was central to Schumpeter’s economic worldview.

The theoretical story begins in earnest after World War II when the concept of *total factor productivity* came to the fore. Robert Solow is an American economist mainly known for his work on the theory of economic growth; we name the exogenous growth model after his work. Solow took courses from Schumpeter at Harvard and was influenced by his views on innovation and technological change, though Solow was also part of the generation of economists embracing the new discipline of mathematical or quantitative economics, which was foreign to Schumpeter.[11] Solow's insight in two papers in 1956 and 1957, for which he won a Nobel prize, was that technological change, what he called 'technological progress,' must be the 'residual' left over from empirical growth once we remove the traditional inputs of labor and capital.[12] Total-factor productivity (TFP) is the 'residual' in total output not credited to the traditional inputs of labor and capital.* Solow calculated that 87.5% of the growth in US output per worker was due to technical progress.[13] In 1954, Solomon Fabricant similarly estimated the amount as 90%.[14] But these were 'lumpy' measures; factors like a changing composition of the workforce (especially the growth of women and two-earner families) were also at play.

Fritz Machlup's seminal 1962 book, *The Production and Distribution of Knowledge in the United States*, was the first to coin the terms 'knowledge industry' and 'knowledge worker'.'[15] It noted that the knowledge industry generated 29 percent of the US GNP in 1958.[18] Marc Porat updated Machlup's efforts for 1967 using a different methodology based on national income accounts, an approach that is less comprehensive than Machlup's, but which has the advantage of relying on standard data collection. This effort, *The Information Economy*, authored with Michael Rubin in 1977, was also adapted as the methodology for cross-country comparisons by the OECD in the 1980s.[24] Another influential paper of this era was by Kenneth Arrow in 1962, in which he introduced the concept and evidence for what he called 'learning by doing,'[17] what is now more formally understood and accepted as the learning curve. Unlike a specific innovation, the idea of the learning curve captured that experience and practice led to efficiencies and productivity on their own as we master our tasks.

By the 1960s and 1970s, it was becoming clear that developed economies were becoming information economies, increasingly staffed by knowledge workers. These forces needed explicit attention within quantitative economic models. Robert Lucas, now a Nobel laureate from the University of Chicago, probed the questions of rational expectations and internal factors promoting growth. By the mid-1980s, a group of growth theorists had become increasingly dissatisfied with standard accounts of exogenous factors determining long-run growth. The focus shifted to the needs for quantitative models that made these 'technological' or 'information' factors explicit. In other words, these 'X' factors are not a lump, residual consideration as defined by TFP, but are an internal one within the models with multipliers and feedbacks. In short, these new growth factors needed an explicit and *endogenous* (internal) specification in the model, not left as some *exogenous* (external) residual.

Arguably, the field of information economics began with David Lawrence's book, *The Economic Value of Information*, in 1999.[18] A book by David Warsh in 2007, *Knowledge*

* By definition, TFP cannot be measured directly.

and the Wealth of Nations: A Story of Economic Discovery,[19] is an explicit account of the transition from TFP to an internal growth model. The book focuses on Paul Romer, then of Stanford University, a recent chief economist of the World Bank, but earlier a colleague of Lucas, pivoting on his seminal paper, "Endogenous Technological Change."[20] By bringing the consideration internal to the model, it could be inspected and broken into parts. The first Romer insight is that information and its artifacts are also products and outputs of the economic function. Romer's second insight is that once produced, information or knowledge assets may be provided or distributed at essentially zero marginal cost. Romer had added a new dimension of 'rival' and 'non-rival' goods to the growth theory lexicon. Information and knowledge were becoming both inputs and outputs to the economic function. Romer's papers provided the concepts to analyze further the role of information in growth.

For example, between 2000 and 2005, estimates at the industry level indicate that almost half of the aggregate productivity was due to productivity growth originating from information technology,[21] though the IT industries themselves only accounted for a little over 3% of nominal aggregate value.[22] Jorgenson, Ho, and Samuels[22] explicitly separated out innovation from the diffusion of prior innovations due to information. The study by Apte and Nath, mostly an update of the earlier analyses by Porat, found that by 1997 two-thirds of the US economy was an information one.[23]

By 2009, Romer and Jones were able to claim proof for the endogenous growth model, and they put forward six research questions to look for in the coming 25 years, including the role of human capital, differential growth rates between countries, and accelerated growth.[24] Innovation and its grounding in knowledge had finally assumed its central, internal role in economists' understanding of economic growth. What Schumpeter had referred to as 'innovation' is now understood as too broad; innovation is but a part of the overall growth effect due to information. What is helpful from these more recent studies is to separate out innovation from information dissemination. The next step, for which we have not yet developed useful datasets, would be to unpack the ideas of innovation and information into the categories from Mokyr,[8] namely, propositional and prescriptive knowledge.

Innovation is an individual affair in its discovery, but a communal one in its application, at which point we call it *knowledge*. We mimic innovations that produce real differences. Farming innovations may include better ways of planting or spacing the wheat, perhaps using a plow; selecting specific wheat strains for next year's plantings; irrigating the land; providing harnesses to the mules; or dividing and specializing the responsibilities between the children, Some of these innovations are new devices, such as harnesses or plows. Some of these innovations are new practices, such as tilling or irrigation methods or specializations in tasks or labor. Not every farmer must innovate on his own. Copying and imitation diffuse these changes across farms and workers.

Indeed, for millennia, this is how human progress took place. Some innovations, such as fire, the wheel, iron and bronze, the arch, alphabets, the plow and the yoke had material benefits to all who encountered them. These innovations were fundamental and diffused at the pace of human movement. However, one could argue, each arose as a flash of insight and not from a process of systemic information. Further, innovations tended to diffuse slowly, following the pace and concentration of trade routes. The innovative event was quite rare, and most practices had been stable for centuries. It is not at all surprising that early economic ideas tended to focus on the traditional factors of production of land, labor, and capital. These had been the steady constants for what had been flat growth for centuries.

If the nature of the biological organism is to contain within it genetic information from which adaptations arise that it can pass to offspring via reproduction — an information volume that is inherently limited and only transmittable by single organisms — then the nature of human cultural information is a massive breakpoint. With the fixity and permanence of printing and cheap paper — and now cheap electrons — all prior information across the entire species can be accumulated and passed on to subsequent generations. Our storehouse of available information is thus growing geometrically, and accessible to all, factors that make the fitness of our species indeed a shift from all prior biological beings, including early humans.

It is silly, of course, to point to single factors or offer simplistic slogans about why this growth occurred and when. Indeed, the scientific revolution, industrial revolution, increase in literacy, electrification, printing press, Reformation, rise in democracy, and many other plausible and worthy candidates have been brought forward to explain these historical inflections in accelerated growth. For my lights, I believe each of these factors had its role to play. Still, at the most fundamental level, I think the drivers for this growth came from prior human information. Undoubtedly, the printing press helped to increase total volumes, but it was declining paper costs that made information access affordable and (nearly) universal.

Information, specifically non-biological information passed on through cultural means, is what truly distinguishes us humans from other animals. We have been easily distracted looking at the tangible when it is the information artifacts ('symbols') that make us human. So, the confluence of cheaper machines (steam printing presses) with cheaper paper (pulp) brought information to the masses. In that process, more people learned, more people shared, and more people could innovate. Now, with computers and the Internet, we can also digitize and place nearly all of the accumulated human knowledge into anyone's hands. What will that bring?

UNTAPPED INFORMATION ASSETS

Today, in the advanced knowledge economy of the United States, the information contained within documents represents about a third of total gross domestic product. Some 25% of the annual trillions of dollars spent on document creation lends itself to actionable improvements.[25] If we are to improve our management and use of information, we need to understand how much value we routinely throw away.

Valuing Information as an Asset

For an enterprise, we can define intangible assets as private expenditures on assets that are intangible and necessary to the creation and sale of new or improved products and processes, including designs, software, blueprints, ideas, documents, know-how, artistic expressions, recipes, branding, and the like. Nakamura made one of the first economy-wide investigations of intangible assets in 2000.[29] He presented direct and indirect empirical evidence that US private firms invested at least $1 trillion annually in intangible assets in that year. This amount was nearly equal to the amount spent on plant and equipment. Firms also held a capital stock of intangibles with a market value of at least $5 trillion, representing a third of the amount of US corporate assets.

Another group — Carol Corrado, Charles Hulten, and Daniel Sichel, known as 'CHS' across their many studies — also began systematically to evaluate the extent and basis of intangible assets.[26] They estimated that spending on long-lasting knowledge capital — not just intangibles broadly — grew relative to other major components of aggregate demand during the 1990s. CHS was the first to show that by the turn of the millennium that fixed US investment in intangibles was at least as large as business investment in traditional, tangible capital. Surveys of more than 5,000 companies in 25 countries confirmed these trends and showed that most of these assets did not get reflected in financial statements. A large portion of this value was due to 'brands' and other market intangibles.[27] The total 'undisclosed' portion appeared to equal or exceed total reported assets. In 2009 the National Academies in the US reported on their investigation into policy questions related to intangible assets,[28] with much relevant information. The study contained an update by CHS confirming and extending their prior findings. In 2010 Nakamura also re-visited his earlier analysis and found that intangible values had finally exceeded expenditures on plant and equipment, with intangible investments now being on the order of 8% to 10% of GDP annually in the US.[29]

In parallel, these groups and others began to decompose the intangible asset growth by country, sector, or asset type. The specific component of 'information' received a great deal of attention. Apte, Karmarkar, and Nath, in particular, conducted a couple of important studies during the 2000 decade.[23] [30] [31] They found nearly two-thirds of recent US GDP was due to information or knowledge industry contributions, a percentage that had been growing over time.[23] They found that a secondary sector of information internal to firms constituted well over 40% of the information economy or some 28% of the entire economy. So the information activities that are internal to organizations represent a considerable part of the economy.

Today, intangibles now equal or exceed the value of tangible assets in advanced economies. The methodological and conceptual issues of how to explicitly account for information on a company's books are, of course, matters best left to economists and professional accountants. However, with the growing share of information related to intangible assets, this is a matter of great importance to national policy. For example, accounting for R&D efforts, one possible component of intangible assets, as an asset versus a cost, has been estimated to add on the order of 11 percent to US national GDP estimates.[28]

The mere generation of information is not necessarily an asset. Some of the information has no value, and some indeed represent a net sunk cost. What we can say, however, is that valuable information that is created by the enterprise but remains unused or created anew means that what was an asset has now been turned into a cost* — sometimes a cost repeated many times over. Information that *is* used *is* an asset, intangible or not. The value of this information depends on its nature and use. We may value the information by *cost* (historical cost or what it cost to develop it), *market value* (what others will pay for it), or *utility* (what is its present value as benefits, broadly accounted for, accrue into the future). Traditionally the historical cost method has been applied to information. However, since information can both be sold and still retained by the organization, it may have both market value and utility value, with its total value being the sum or a portion thereof.†

Researchers estimated in the early 2000s that enterprises adequately use only five to seven percent of existing information and the total value of information in enterprises is in the range of 10% to 33% of US GDP.[25] [23] Amongst all enterprise resources and assets, information is the least understood and the least managed. Managers are overlooking the value of their information.

More than a decade ago Moody and Walsh put forward a seminal paper on the seven 'laws' of information.[32] Unlike other assets, information has some unique characteristics that make understanding and valuing it more difficult, which leads to lower perceived importance. I have taken some liberty with the Moody and Walsh 'laws' to reflect my experience:

1. Information is (infinitely) shareable, it is not necessarily a depletable resource (though sharing may reduce proprietary advantage);
2. The benefit of information often increases with use, such as through the learning curve;‡
3. The value of information increases with accuracy;
4. The value of information increases in combination;§
5. The value of information is situational and perishable, with varying shelf life;
6. More is not necessarily better; the question is one of relevancy; and
7. Information builds upon prior information, the combinations of which often stimulate new insights.**

* That is because time and effort is required to generate unique information.

† Of course, information can also have a multiplicative effect, especially in those areas Mokyr calls prescriptive knowledge; but, that is not applicable to this specific point, since we are talking about *re-use*.

‡ A corollary is that it is an asset only if it provides future economic value, another is that awareness of the information's existence is an essential requirement in order to obtain this value, and a third corollary is that information requires an understanding of where it fits and how to take advantage of it.

§ Network effects are particularly important here; see discussion of the Viking algorithm in *Chapter 10*.

** This propagation results from summations, analysis, unique combinations and other ways that basic datum get recombined into new information. Thus, while the first law noted that information can not be consumed (or depleted) by virtue of its use, we can also say that information tends to reproduce and expand itself via use and inspection.

Lost Value in Information

Information — more specifically, knowledge management — has bedeviled enterprises for decades. As the prior section indicates, information has enormous importance to most organizations and the overall economy. Why do these disheartening statistics keep cropping up concerning information management?

- 65% of data integration or KM projects 'fail';
- A typical organization only uses 5 to 7 percent of the information it already has on hand;
- 20% to 25% of a knowledge worker's time is spent trying to find information;
- We waste 25% of all document creation costs; or
- IT now consumes 4% of all enterprise expenditures and employs 6% of enterprise workers.

These are statistics I have encountered, or about which I have researched and written.[1] As rough figures or averages, they say nothing about what an individual enterprise or project may experience — there are, after all, good managers out there — but they do provide a pretty fair metric for the typical experience.

About a decade ago I began a series of analyses looking at how we spend money on preparing documents within US companies, and how much of that investment was being wasted or not re-used.[25] The total benefit from improved information access and use to the U.S economy may be on the order of 8% of GDP. For the 1,000 largest U.S. firms, benefits from these improvements can approach nearly $250 million annually per firm (2002 basis). About three-quarters of these benefits arise from *not* re-creating the intellectual capital already invested in prior document creation. About one-quarter of the benefits are due to reduced regulatory non-compliance or paperwork, or better competitiveness in obtaining solicited grants and contracts. Finding and re-using information for compliance purposes as well as avoiding duplicate content creation are areas amenable to waste reductions. Note that new initiatives, as discussed in the next *Chapter 4*, are *not* included in this analysis.

This overall value of document use and creation is in line with the analyses of intangible and information assets noted above, and which arose from entirely different analytical bases and data. This triangulation brings some confidence that the estimates are approximately accurate. In any case, the potential benefits to the better use of existing information assets likely exceed what most managers currently believe, otherwise we would see better performance trends.

IT departments seem to have particular difficulty with information and knowledge management projects. Transaction and relational data systems require a different set of skills and viewpoint than for information sharing and the open nature of knowledge.* Relational database systems, which embody a closed-world design, work well for environments where the information domain is known and bounded, but do

* See *Chapter 8*.

not work well with knowledge and changing information. Moreover, the schema that governs closed-world designs is brittle and hard to improve and manage. It is this fact that has often led to IT delays and frustrations. Re-architecting or adding new schema views to an existing closed-world system, as knowledge systems demand, can be fiendishly difficult. Here are some other areas of frustration with IT regarding information and knowledge management:

IT Problems for KM	*Comments*
Inflexible Reports	• reports are rarely 'self-service' • new requests need to be placed in a queue • 90% of stored report templates are never used • unlimited 'slicing and dicing' not available
Inflexible Analysis	• the analysis is rarely 'self-service' • new requests need to be placed in a queue • many requests not accepted due to schema rigidities, cascading changes needed • analysis options are 'pre-canned,' inflexible
Schema Bottlenecks	• brittleness of relational data model and typical star schema • crossing across schema or databases difficult • load and re-indexing cycles can limit access, impose expensive back-end requirements • cannot (often) accommodate new data, structures
ETL Bottlenecks	• getting data into the system needs to be placed in queue • new external data requires extract, transform and load (ETL) routines to be written • schedule and update cycles can be a mismatch to access needs
Reliance on Intermediaries	• all problems above work through intermediaries • there is a disconnect between those with need and decision-makers and those who implement the solutions • inherent issues in communicating requirements to implementers • related time delays to implementation exacerbate the communication of requirements
Specialized Expertise Required	• expertise and skill sets needed to implement solutions different from those of the knowledge consumer • inherent issues in communicating requirements to implementers • high costs for attracting necessary expertise • expertise is inherently an overhead function
Slow Response Time	• all problems above lead to delays, slow response • timely communications, analysis, decisions suffer • delays mean knowledge management is not an active 'contact sport,' becomes mired and unresponsive • some needs are just not requested because of these problems

IT Problems for KM	*Comments*
Dependence on External Apps	• new apps need to be identified, procured • design and configuration of apps requires external expertise, programming skills • multiple sourcing of apps leads to frequent incompatibilities, high costs for integration, poor interoperability
Unmet Needs	• many KM needs are not requested • by the time responses are forthcoming, needs and imperatives have moved on • communications, analysis, and decisions become hassles • the 'contact sport' of active discovery and learning is unmet
High Opportunity Costs	• many KM insights are not discovered • delays and frustration adds to costs, friction, inefficiencies • no way to know the opportunity costs of what is not learned — but, surely is high
High Failure Rates	• the net impact of all of the problems above is to lead to high failure rates (~60% to 70%) and unacceptable costs • reliance on IT for KM has generally failed

Table 3-1: Enterprise IT Weaknesses in Relation to KM

The problems raised in *Table 3-1* show that losses in information and its poor organization and handling lead to a decline in business value. Removing unnecessary mediation roles by IT and placing the knowledge management function directly into the hands of the knowledge worker presents a huge opportunity to recapture that lost value. Much of what I discuss throughout the remainder of this book is geared directly to this aim.

The Information Enterprise

One can probably clock the start of enterprise information technology (IT) to the first use of mainframe computers in the early 1950s, or nearly seventy years ago.[33] The earliest mainframes were massive and expensive machines that required their own specially air-conditioned rooms because of the heat they generated. The first use of 'information technology' as a term occurred in a Harvard Business Review article from 1958.[34] Architectures progressed from mainframes to minis and then personal computers with networks, leading to today's dominance of the Internet. Relational database designs won out for the enterprise in the 1970s and 80s, continuing into today's dominance, but with the recent adoption of graph and NoSQL datastores.

The apogee for enterprise software and apps occurred in the 1990s. Whole classes of new applications (most denoted by three-letter acronyms) such as enterprise resource planning (ERP), business intelligence (BI), customer relationship management (CRM), enterprise information systems (EIS) and the like came to the fore. These systems also began as proprietary software, which resulted in the 'stovepiping' or creating of information silos. In reaction and with market acceptance, vendors such as SAP arose to provide comprehensive, enterprise-wide solutions, though often at high cost, with disappointing results not uncommon. Software revenues as a percent of IT vendor revenues peaked in about the mid-1990s. The plateau for IT expenditures as a portion of GDP appears to have occurred somewhat around 2000.

More significantly, the 1990s also saw the innovation of the World Wide Web with its basis in hypertext links on the Internet. Greatly facilitated by the Mosaic Web browser, the basis of the Netscape browser (ultimately Firefox), and the HTML markup language and HTTP transport protocol, millions began experiencing the benefit of creating Web pages and interconnecting. By the mid-1990s, enterprises were on the Web in force, bringing with them larger content volumes, dynamic databases, and enterprise portals. The ability for anyone to become a publisher led to a focus and attention on the new medium that led to still further innovations in e-commerce and online advertising, creating entirely new categories of business. New languages and uses of Web pages and applications emerged, creating a convergence of design, media, content, and interactivity. Venture capital and new startups with valuations independent of revenues led to a frenzy of hype and eventually the dot-com crash of 2000. The growth companies of the past 15 years have not had the traditional focus on enterprises but the use and development of the Web. From search (Google) to social interactions (Facebook and Twitter) to media and video (Flickr, YouTube) and information (Wikipedia), the engines of growth have shifted away from the enterprise.

Meanwhile, the challenges of data integration and interoperability that were such a keen focus going back to initial enterprise computerization remain. Now, however, these challenges are even higher, as we see images, documents (unstructured data) and Web pages, markup and metadata (semi-structured data) become first-class information citizens. What was a challenge in integrating structured data in the 1980s and 1990s via data warehousing, remains daunting for the enterprise today in the face of unprecendented scale and scope. Services have drifted to the largest IT vendors, and open source is now a primary source of innovation and challenge. We have climbed the data federation pyramid sufficient to overcome most obstacles of hardware, protocols, and data formats, but are stuck at the levels of semantics and trust (provenance).

Roughly in 1997 or so, the number of public enterprise software vendors peaked as did venture funding.[35] There was an uptick in preparing for Y2K and a significant downtick due to the dot-com bubble, and then later the financial crisis. However, change is coming about from the shift of expenditures from license and maintenance fees to services. Some software vendors began to see revenues from services overcome that from licensing in the 1990s. By the early 2000s, this was true for the enterprise software sector as a whole.[35] Today, service revenues account for 70% or so of aggregate sector revenues. Combined with the emergence of open source and other alternatives such as software as a service (SaaS) and cloud computing in general, I think it fair to say that the era of proprietary software with exceedingly high margins from monopoly rents has come to an end.[36] According to Gartner, in the US, more than 70% of IT expenditures are devoted to running existing systems, with only about 11% of budgets dedicated to innovation. This relative lack of support for innovation and high percentages for running existing systems has held true for a decade. Meanwhile, IT's contribution to US productivity has been declining since 2001.[37]

Arguably the emphasis on consumer and Internet technologies means that is where the best developers gravitate. Developing apps for smartphones or working at one of the cool Internet companies or joining a passionate community of open source developers are now attracting the best developers. Open source and Web-based systems also lead to faster development cycles. The very best developers are often the founders of the next generation startups and Web and software companies, as startup costs plunge.[38]

The shift in innovation away from the enterprise has been structural, not cyclical. That means that very fundamental forces are at work to cause this change in innovation focus. Every knowledge-oriented organization must learn to support and nurture its information enterprise. These structural shifts need to affect priorities, mindsets, budgets, and staffing. In an environment of cost pressures and the need for quantifiable results, we need to make pragmatic choices, Peirce's dominant message. The rest of this book, in part, talks about machine learning and various other aspects of artificial intelligence. These are all exciting topics, the shiny new thing. Still, the pragmatic viewpoint insists that in the process of making expenditures for these purposes, that we should include in our design more fundamental and, perhaps, useful applications in information sharing and knowledge management. We will try to weave this practical viewpoint through our narrative as well.

IMPEDIMENTS TO INFORMATION SHARING

Our survey of the current situation suggests a few things. Better use of information will be a significant factor in future economic growth. Growth is vital to wealth creation. Leveraging our existing information assets through re-use and connections is one immediate source of growth, with surprisingly large upsides. Innovations using artificial intelligence will continue the virtuous cycle to help support healthy growth. Individuals and enterprises need to grasp the challenge of knowledge management and need to place those functions into the hands of the knowledge worker. In an information-driven economy, education and access to information and knowledge management resources are essential foundations.

Cultural Factors

Since the widespread adoption of the Internet, which marked a passage beyond hardware, protocols, and formats as limits to interoperate data, the main impediments to information sharing have become cultural. Awareness of the importance of information as an asset has been lacking. Knowing how to interoperate across information stores is not a sought skill. Rewards are geared to information hoarding and gatekeeping over openness and sharing, which are rarely a formal part of performance evaluations. A lack of enlightened understanding about the importance of information leads to a lack of vision and an absence of a knowledge strategy. Since many bosses don't know what to do, they make second-rate decisions about information and knowledge assets. Budget and operational fiefdoms continue a climate that guarantees inefficiencies, with a high opportunity cost, in how to manage information and knowledge. This book tries to provide a consistent viewpoint and logic for why information sharing is so important to enterprises. A change in perspective is required to unleash new growth, one that focuses on management and mindset.

The fact that an overlay of semantic technologies is required, as I discuss in *Chapter 5*, is good news from a cultural standpoint. A critical aspect of shared knowledge schema within an enterprise is the need for relevant stakeholders to have a role in bounding and defining the terminology of the domain. The first dictum of effective messages and reasoning is to communicate with a shared grammar and semantics. The absolute wrong strategy is to try to find or impose 'official' terminologies. Rather, we need to capture language as we use it daily in our tasks and find ways to relate these uses to a shared knowledge graph. Relevant stakeholders need to interact to document current terms and tasks, using the language of their daily work. With sufficient top management commitment, not often easy, such first steps can help set a new cultural tone for sharing. Given the potential incremental nature of deploying semantic technologies, early efforts should focus on prototypes at the level of workgroups or departments. The open nature of semantic technologies means we can readily expand the vocabularies and relate them to what already exists as we bring new stakeholders into the process. One can start small and grow as results become evident. I have called this the 'pay as you benefit' strategy.*

No fundamental technical roadblocks are preventing any enterprise from moving to a vision of shared information, providing useful knowledge support. Despite decades of trying, enterprises have still not broken down their data stovepipes. Rather, they continue to proliferate. In the process, the enterprise has failed to unlock 80% of its information value in documents (unstructured data) and has continuously wasted money in unneeded duplication and lost opportunities.

Tooling and Technology

The semantic technologies recommended in this book are open standards with years of implementation experience; still, their state of tooling is weak. Knowledge graph (ontology) editors and development environments exist, but all of the create, edit, manage, update, delete, map, and validate tools could be improved. Each operation would benefit from being streamlined from the standpoint of the user, starting with subject matter experts (SMEs). Rather than comprehensive IDEs (integrated development environments), many of these functions are better separated out as options embedded within current workflows. Such a function, say, might be to add a new synonym for a concept in the knowledge base when encountered in a relevant

* See *Chapter 13*.

document. Many individual functions would benefit from being split apart and incorporated in a distributed way across multiple steps in information handling, from creating to using, validating, updating, or archiving.

Another weak area is in user permissions and authorizations. Optimally, a single sign-on should be sufficient to grant access or not to various datasets, records, and applications. Methods and protocols exist, mostly IP-based, with some working installations from which to draw lessons, but more robust and secure options are needed, likely using third-party applications. When relying on Internet protocols, we need to manage unauthorized access and hacking.

Perspectives and Priorities

I think it is fair to say, in general, that we do not have a broad and informed view on the value of information. We do not know what information we have and can not find it, and we waste much time looking for it. We misallocate resources for generating, capturing and storing information because we do not understand its value and potential and don't know what we already have. We do not manage the use of information or its re-use. We do a lousy job of using information to bridge communication differences across our stakeholders. We inadequately leverage what information we have and often miss valuable insights. What we have we do not connect. We do not know how to turn our information into knowledge.

Fundamentally, because we do not understand information in our bones as central to the well-being of our enterprises, we continue to view the generation of information as a 'cost' and not an 'asset.' Perhaps, akin to the perspective of Thirdness in Peirce's universal categories, which we discuss in *Chapter 6*, we need to bring new perspectives to our understanding and appreciation of information.

Peirce defended and is known as a realist. Within that realism and subject to his pragmatism, I believe he can also be called an idealist. Real and practical ways exist to achieve meaningful visions of information sharing, which can release hidden value within any information enterprise. This foundation can then be extended with knowledge bases and artificial intelligence to mine further still the value contained in that information.

Chapter Notes

1. Some material in this chapter was drawn from the author's prior articles at the *AI3:::Adaptive Information* blog: "Untapped Assets: The $3 Trillion Value of US Enterprise Documents" (Jul 2005); "A Trillion is a Large Number (12 zeros)" (Jul 2005); "Why Are $800 Billion in Document Assets Wasted Annually? V Summary" (Mar 2006); "Climbing the Data Federation Pyramid" (May 2006); "Knowledge: Unravelling the X Factor in Growth and Wealth" (Jun 2006); "Historical Origins of the Knowledge Economy" (Jul 2006); "Information is the Basis for Economic Growth" (Aug 2007); "When Linked Data Rules Fail" (Nov 2009); "Declining IT Innovation in the Enterprise" (Jan 2011); "Spring Dawns on Artificial Intelligence" (Jun 2014); "Innovation, Information, Growth and Wealth" (Jun 2014); "Big Structure and Data Interoperability" (Aug 2014); "Logical Implications of Interoperability" (Jun 2015); "Hidden Expenses Underneath Machine Learning" (Feb 2016).
2. If we take the first growth breakpoint as 1820, and a nominal estimate of 40,000 years for human history

(fudging a bit on the estimated occurrence of cave paintings to account for oral histories), we end up with exactly 99.5%.

3. Maddison, A., *The World Economy: Historical Statistics*, Paris: OECD Development Centre, 2003.
4. The historical data were originally developed in three books by Angus Maddison: Monitoring the World Economy 1820-1992, OECD, Paris 1995; The World Economy: A Millennial Perspective, OECD Development Centre, Paris 2001; and The World Economy: Historical Statistics, OECD Development Centre, Paris 2003. All these contain detailed source notes. Figures for 1820 onwards are annual, wherever possible. For earlier years, benchmark figures are shown for 1 AD, 1000 AD, 1500, 1600 and 1700. These figures have been updated to 2003 and may be downloaded by spreadsheet from the Groningen Growth and Development Centre (GGDC), a research group of economists and economic historians at the Economics Department of the University of Groningen headed by Maddison. See http://www.ggdc.net/.
5. This inflection point in approx. 1000 AD corresponds somewhat to the adoption of raw linen paper v. skins and vellum, among other correlations that might be drawn.
6. Nordhaus, W. D., "Do Real-Output and Real-Wage Measures Capture Reality? The History of Lighting Suggests Not," *The Economics of New Goods*, T. F.Bresnahan and R.J. Gordon, eds., University of Chicago Press, 1996, pp. 27–70.
7. Eisenstein, E. L., *The Printing Press as an Agent of Change*, Cambridge University Press, 1980.
8. Mokyr, J., *The Gifts of Athena: Historical Origins of the Knowledge Economy*, 2002.
9. In Paul Starr's *Creation of the Media: Political Origins of Modern Communications*, Basic Books, 1-2005, he notes how in 15 years from 1835 to 1850 the cost of setting up a mass-circulation paper increased from $10,000 to over $2 million (in 2005 dollars). True, mechanization was increasing costs, but from the standpoint of consumers, the cost of information content was dropping to zero and approaching a near-time immediacy. The concept of "news" was coined, delivered by the "press" for a now-emerging "mass media."
10. Schumpeter, J. A., *The Theory of Economic Development: An Inquiry Into Profits, Capital, Credit, Interest, and the Business Cycle*, Harvard University Press, 1955.
11. Solow, R., "Economist's View: Robert Solow on Joseph Schumpeter" Available: http://economistsview.typepad.com/economistsview/2007/05/robert_solow_on.html.
12. Solow's endogenous model of economic growth is also known as the Solow-Swan neoclassical growth model as the model was independently discovered by Trevor W. Swan and published in "The Economic Record" in 1956, allows the determinants of economic growth to be separated out into increases in inputs (labor and capital) and technical progress.
13. Solow, R. M., "Technical Change and the Aggregate Production Function," *The Review of Economics and Statistics*, 1957, pp. 312–320.
14. Fabricant, S., *Economic Progress and Economic Change, a part of the 34th annual report of the National Bureau of Economic Research.*, New York: 1954.
15. Machlup, F., *The Production and Distribution of Knowledge in the United States*, Princeton University Press, 1962.
16. Porat, M. U., *The Information Economy: Definition and Measurement*, Office of Telecommunications (Department of Commerce), 1977.
17. Arrow, K., "Economic Welfare and the Allocation of Resources for Invention," *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Universities-National Bureau, Committee for, Economic Research, Committee on Economic Growth of the Social, and Science Research Council, eds., Princeton University Press, 1962, pp. 609–626.
18. Lawrence, D. B., *The Economic Value of Information*, Springer-Verlag New York Inc, 1999.
19. Warsh, D., *Knowledge and the Wealth of Nations: A Story of Economic Discovery*, WW Norton & Company, 2007.
20. Romer, P. M., "Endogenous Technological Change," *Journal of Political Economy*, vol. 98, 1990, pp. S71–S102.
21. Rosenthal, S., Russell, M., Samuels, J. D., Strassner, E. H., Usher, L., and others, "Integrated Industry-Level Production Account for the United States: Intellectual Property Products and the 2007 NAICS," *3rd World KLEMS conference (May), Tokyo, Japan*, 2014.
22. Jorgenson, D., Ho, M., and Samuels, J., "Long-Term Estimates of US Productivity and Growth," *Third World*

KLEMS Conference, 2014, pp. 19–20.

23. Apte, U. M., and Nath, H. K., *Size, Structure and Growth of the US Information Economy*, UCLA Anderson School of Management on Business and Information Technologies, 2004.
24. Jones, C. I., and Romer, P. M., "The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital," *American Economic Journal: Macroeconomics*, vol. 2, 2010, pp. 224–245.
25. Bergman, M. K., "Untapped Assets: The $3 Trillion Value of U.S. Enterprise Documents," *AI3:::Adaptive Information*, Jul. 2005.
26. Corrado, C. A., and Hulten, C. R., "How Do You Measure a 'Technological Revolution'?," *The American Economic Review*, vol. 100, 2010, pp. 99–104.
27. Jarboe, K. P., and Furrow, R., *Intangible Asset Monetization: The Promise and the Reality*, Athena Alliance, 2008.
28. Mackie, C., *Intangible Assets: Measuring and Enhancing Their Contribution to Corporate Value and Economic Growth*, The Board on Science, Technology, and Economic Policy (STEP) Committee on National Statistics (CNSTAT), ISBN, 2009.
29. Nakamura, L. I., "Intangible Assets and National Income Accounting," *Review of Income and Wealth*, vol. 56, 2010.
30. Apte, U. M., Karmarkar, U. S., and Nath, H. K., "Information Services in the US Economy: Value, Jobs and Management Implications," *California Management Review*, vol. 50, 2008, pp. 12–30.
31. Apte, U., "The U.S. Information Economy: Value, Employment, Industry Structure, and Trade," *Foundations and Trends® in Technology, Information and Operations Management*, vol. 6, 2012, pp. 1–87.
32. Moody, D., and Walsh, P., "Measuring The Value Of Information: An Asset Valuation Approach," *Seventh European Conference on Information Systems (ECIS'99), Copenhagen Business School, Frederiksberg*, Denmark: 1999.
33. Punch cards were an early form of automatin, and extend back to the 1700s and then famously to the Jacquard loom. Though calculating machines and others extend back to Charles Babbage and more relevant efforts during World War II, the first UNIVAC was delivered to the US Census Bureau in 1951, and the first IBM to the US Defense Department in 1953. Many installations followed thereafter. See, for example, Lectures in the History of Computing: Mainframes.
34. As provided by "information technology" (subscription required), Oxford English Dictionary (2 ed.), Oxford University Press, 1989, http://dictionary.oed.com/, retrieved 12 January 2011.
35. Cusumano, M. A., "The Changing Software Business: Moving from Products to Services," *Computer*, vol. 41, Jan. 2008, pp. 20–27.
36. While we may occasionally see some vendors successfully buck this trend, I suspect these will only occur for established vendors with established platform advantages or for isolated applications where the innovating vendors have a significant first-mover advantage.
37. PricewaterhouseCoopers, "Why Isn't IT Spending Creating More Value?" Available: http://www.pwc.com/en_US/us/increasing-it-effectiveness/assets/it_spending_creating_value.pdf.
38. Open source and Internet-based systems have reduced the capital necessary for a new startup by an order of magnitude or so over the past decade. It is now possible to get a new startup up and running for tens to hundreds of thousands of dollars, as opposed to the millions of years past.