Tending the infrastructure commons: ensuring the sustainability of our vital public systems

R. G. LITTLE

The Keston Institute for Infrastructure, University of Southern California, USA

(Received 17 December 2004; accepted in revised form 1 March 2005)

It is by now well recognized that civil infrastructure systems are essential in providing the range of services generally considered necessary to support a nation’s economic well-being and quality of life. Consequently, it is of the utmost importance to government, business, and the general populace that these services are sustained over the long term by periodic replenishment of the physical systems that deliver them. This has proven challenging for several reasons. First, civil infrastructures do not all lie within the public domain. Second, infrastructures are complex systems that are vulnerable to failures or service degradations in other systems because of their interconnected and interdependent natures. Third, despite the obvious importance of reliable and resilient systems to our collective social, economic, and political well-being, it has been difficult to fund the necessary maintenance and capital improvements. Deciding what levels of reliability should be provided and who should pay for it are not simple questions to be addressed solely by service providers, customers, or government regulators. This paper will present the concept of an infrastructure commons and recommend some initial actions that the federal government could take to ensure that the national ‘public good’ aspect of civil is preserved and enhanced. These actions include the beginning of an informed and serious dialogue between the public and private sectors, building coalitions for action among stakeholders in the governance process, and exploring alternative mechanisms for ensuring dependable and sustainable resource streams.

Keywords: Sustainable infrastructure; Supply chains; Infrastructure commons

1. Introduction

Throughout history, cities have generated high levels of economic, social, and political activity supported, in part, by extensive and effective infrastructure systems (NAE 1988). Modern economies rely on the ability to move goods, people, and information safely and reliably and it is now generally well accepted that civil infrastructures are vital to a nation’s economy, security, and quality of life. However, although it may be the hardware (i.e. the roads and railways, pipes, transmission lines, communication satellites, and network servers) that initially focuses discussions of infrastructure, it is actually the services that these systems provide that are of real value to the public (Little 2002). It is the need to ensure the long-term viability of these critical services that leads directly to the concept of sustainable infrastructure.

2. ‘Sustainable’ infrastructure

Sustainability has been defined in many ways by many groups. Of the many definitions available, perhaps the most applicable to infrastructure is that of the World Commission on Environment and Development (1987)
which states that sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs.

Inherent in this definition is the concept of renewal over consumption. Just as natural systems embody the process of generation, growth, maturation, decline, and death, the life cycle of physical systems follows a similar path. Unlike natural systems, however, physical systems cannot sustain themselves; they must be renewed from without. This implies exogenous actions in the form of maintenance, repair, renewal, and replacement on a more or less continuous basis. These sustaining actions require inputs in the form of capital, materials, labour, and other resources. Depriving a physical system of capital for maintenance and repair, for example, will have a similar effect to depriving a natural system of food or water—it will decline and, ultimately, cease to function.

Because it is of the utmost importance to government, business, and the general populace that the services provided by infrastructure be sustained over the long term, periodic replenishment of the physical systems that deliver them must also be ensured. However, this has proven challenging for several reasons. First, civil infrastructures do not all lie within the public domain; most, in fact, are private, for-profit enterprises with varying degrees of public scrutiny and control of their management and operating practices. Second, infrastructures are complex systems that are vulnerable to failures or service degradations in other systems because of their interconnected and interdependent natures. The reliability of the total system is less than the sum of its parts. Third, despite the obvious importance of reliable and resilient systems to our collective social, economic, and political well-being, it has been difficult to fund the necessary maintenance and capital improvements solely through revenues derived from direct consumers of services; no governance structure or funding source for this ‘public good’ aspect of infrastructure has emerged. It will be shown that all of these factors conspire against the implementation of a holistic strategy for sustaining our critical systems.

3. The nature of civil infrastructure

The President’s Commission on Critical Infrastructure Protection (PCCIP) was established to develop a national strategy for protecting and assuring the continuity of critical infrastructures from physical and cyber threats. The PCCIP defined critical infrastructures as those systems whose incapacity or destruction would have a debilitating impact on the defense or economic security of the nation. They included telecommunications, electrical power systems, gas and oil, banking and finance, transportation, water supply systems, and government and emergency services; with the exception of banking and finance and government and emergency services, these are typically considered civil infrastructures (PCCIP 1997). Obviously, not all of these systems are germane to the current discussion but of those that are (i.e. telecommunications, electrical power systems, gas and oil, transportation, water supply systems), only transportation (roadways) and water supply are predominantly in public ownership. With some exceptions (e.g. the Tennessee Valley Authority which produces and sells electric power in the south central US), the others are in private ownership. Although the focus of the PCCIP was on vulnerabilities and threats to infrastructure, it recognized the mixed nature of infrastructure services and how they are provided, and strongly supported the concept of public–private partnerships to address the existing and emerging threats to critical systems. In the proposed ‘National Policy for Infrastructure Protection’ the PCCIP noted that (PCCIP 1997):

Further the US recognizes that assuring infrastructure is not just a government or business responsibility, but is shared by those public and private interests that own and operate the infrastructures and the government agencies responsible for defense, law enforcement, and economic security of the nation.

The lack of a workable governance structure to address infrastructure issues holistically, and its impact on a sustainable infrastructure strategy, will be discussed later in this paper.

4. Infrastructure interdependency

Civil infrastructures are inherently interconnected and interdependent in ways that are not fully understood. Therefore, failures in one infrastructure system can cause disruptions in others that could ultimately spread to still other systems. For example, during the August 2003 power blackout in the northeastern US, the Cleveland (Ohio) Division of Water, which serves 1.5 million customers, was unable to pump treated water into the distribution system. This exacerbated the effects of the power outage and further impacted business and industry (WKYC 2003). Although in this case, the interdependency is straightforward (e.g. the role played by electric power in providing water service is obvious), the interdependencies of other systems are no less real if not as visible. Figure 1 depicts some of the interconnections that exist between our basic service infrastructures.

Interdependent failure effects occur when an infrastructure disruption spreads beyond itself to cause appreciable impact on other infrastructures, which in turn causes more effects on still other infrastructures. When an infrastructure system suffers an outage, it is often possible to estimate the impact of that outage on service delivery.
These are the ‘directly dependent effects’ of the outage. However, that outage may also diminish the ability of other infrastructures, through no malfunction of their own, to deliver the level of services that they normally provide. These indirect effects make up a first-order interdependent effect. The impact of the disruption may not stop at these first-order effects. They may go on to affect adversely still other critical infrastructure components, possibly the system that was the original source of the problem thus further aggravating the situation. These second-order impacts can propagate still further, causing yet another round of effects. How far these effects propagate, and how serious they become, depends on how tightly coupled the infrastructure components are, how potent the initiating cause, and whether or not countermeasures such as redundant capacity are in place. The outage effects will either die out as they move further away in time and space from the initiating event, limiting overall damage, or they will gather force in successively stronger cascading waves until part or all of the infrastructure network breaks down. In the latter case, losing a key component can create a much broader failure that is out of proportion to the initiating event.

Because of system interconnections, cascading failures can cross infrastructure boundaries as demonstrated by the 1998 Galaxy IV satellite failure (CNN 1998). When the PanAmSat Galaxy IV communication satellite rotated out of its orbital position in May 1998, over 80 per cent of the digital pagers in the US went off-line. Cable and broadcast transmissions were affected, as were credit card authorizations and ATM transactions. This event also could have had serious human impacts as many hospitals and healthcare providers in the US faced a crisis in emergency communications when they could not page doctors and other care givers. A failure of this nature is particularly critical in a healthcare system which, in the quest for increased efficiency and productivity like much of the economy, relies on just-in-time service delivery.
The potential for failures in one infrastructure system to cause disruptions in others that could ultimately cascade to still other systems with unanticipated consequences is very real. In truth, beyond a certain rudimentary level, the linkages between infrastructures, their interdependencies, and possible failure mechanisms are not well understood. Although recognized as a serious concern in certain industries, the issue of infrastructure interdependency has only recently begun to receive serious attention. For example, the director of the Office of Science and Technology Policy made the following points in testimony before Congress (Marburger 2002):

(a) the economy and national security of the US are becoming increasingly dependent on US and international infrastructures, which themselves are becoming increasingly interdependent;
(b) deregulation and growth of competition in key infrastructures have eroded spare infrastructure capacity that served as a useful shock absorber;
(c) mergers among infrastructure providers have led to further pressures to reduce spare capacity as management has sought to wring out excess costs;
(d) the issue of interdependent and cascading effects among infrastructures has received almost no attention.

5. The value of infrastructure

Over the past twenty years, increasing attention has been drawn to the deterioration of US infrastructure and the implications of not addressing it (Choate and Walter 1981, NCPWI 1988, ASCE 1988). The core premise underlying these discussions is that ‘infrastructure matters’ and that investment in infrastructure (primarily by the public sector) needs to be dramatically increased. Aschauer (1990) bases his arguments on quality of life and economic issues. He maintains that public expenditures to improve water quality, manage wastes, provide for leisure activities, and increase economic opportunity by means of improved mobility return far more in direct and indirect benefits than their direct cost. However, he also posits that public investment in infrastructure (‘public capital’) has a direct and positive impact on private sector output.

Given that public capital complements private capital, an increase in the public capital stock can be expected to stimulate private capital accumulation through its effect on the profitability of private capital.

Munnell (1990) also found positive linkages between physical infrastructure and productivity, although to a lesser degree. Taken together, their combined body of work offers a compelling argument that increased public sector investment in infrastructure has a positive impact on private sector output. However, their findings have been questioned, particularly for the magnitude of these effects (Aaron 1990, Musgrave 1990, Eberts 1990). Although there still remain many unresolved questions regarding the precise nature of the relationship between public investment in infrastructure and private output, and the size of the coefficients that describe it, a discussant of Munnell’s paper notes (Friedlaender 1990)

Alicia Munnell argues convincingly that public infrastructure investment has a positive impact on regional output and growth. I, for one, do not have to be convinced of this; on a purely intuitive and anecdotal level in both developed and less developed countries, one senses that regions with an extensive base of public infrastructure have stronger economic performance than those with a weak or decaying base.

The work of Aschauer, Munnell, and others was championed by many groups in the 1990s to support greatly increased levels of public sector investment in physical infrastructure. Unfortunately, the Keynesian perspective of these arguments often had the negative effect of casting needed infrastructure improvements as make-work jobs programmes. Similarly, others who called for increased public spending to address the problem of deteriorating infrastructure, such as contracting and engineering groups, were seen as self-serving special interests.

During this period, the generally passive and non-strategic national approach to infrastructure renewal continued to receive attention from economists (Ebenberg 1994, Aschauer 1998). Unfortunately, warnings that economic growth was outstripping the capacity of the physical systems on which it depends were met with either apathy or disbelief. For example, when funding for the replacement of the Woodrow Wilson Bridge that spans the Potomac River just south of Washington, DC was announced by President Clinton and Secretary of Transportation Rodney Slater in 1998, they felt compelled to assure their audience that the project was really needed and not merely part of the local ‘pork barrel’ (Little 1999). That such assurance was even thought necessary for a 40-year-old bridge with documented structural integrity issues that carries over 200,000 vehicles a day around the nation’s capital is a telling comment. It is logical to wonder why then, in the face of so much supporting economic and engineering evidence, that there is no initiative underway to renew the nation’s infrastructure and no effective constituency to demand it.

6. The infrastructure commons

The following brief discussion from the field of natural resources provides a context for sustainable infrastructure
systems that is the core concept of this paper. Common-pool resources have been defined as those to which a large number of people have access and whose combined overuse can lead to diminishment and ultimately, the collapse of the resource. Examples of common-pool resources are fisheries, pastures, or forests. If each user of the resource exercises restraint in what they take from the common pool, these resources can be sustained indefinitely. If they do not, the resource will collapse. However, if one or a few users exercise restraint and others do not, the resource still collapses and the persons acting in the common interest lose out on the short-term benefits of overuse. Thus, it is in no single user’s best interest to act for the common good. This dilemma is the ‘tragedy of the commons’ that Hardin (1968) described in his seminal paper that served as a rallying point for the environmental movement in the 1970s.

Unlike a common-pool resource, a public good is something to which everyone has access, but one person’s use of the resource does not necessarily diminish the potential for use by another. Public radio and television and scientific knowledge are examples of public goods in that they can be enjoyed by all without reducing the quantity or quality of the good (Dietz et al. 2002). In a large group, an individual can enjoy the benefits of such public goods without contributing to their production, and if the group is large enough, it can support a significant number of non-contributors. Much as in the commons dilemma, a person’s narrow self-interest and the common good are in conflict. One can initially reap the benefits of contributions made by others at no personal cost. Of course, if everyone followed this same ‘free rider’ strategy, the public good would ultimately cease to exist and all would lose out. Civil infrastructures exhibit characteristics of both common-pool resources and public goods and this could serve to explain, at least in part, why no one individually feels compelled to contribute to their long-term vitality. Infrastructure behaves as a common-pool resource in that the physical systems require periodic replenishment in terms of maintenance, repair, and rehabilitation. If the systems are not replenished, they will wear out prematurely and ultimately fail.

Although it is in everyone’s personal and collective best interest that this does not occur, the commons dilemma has shown that it is in no one’s individual best interest to contribute to ensuring that it does not. Similarly, although customers may be willing to pay for the direct benefits of infrastructure services, it is much more difficult for suppliers to recapture the cost of providing long-term reliability, i.e. the ‘public good’ aspect of the service. This conflict between the enterprise nature of providing infrastructure services and the public good aspect of long-term reliability has been explored for the US electric power industry (Kleindorfer 2004). In the absence of a model for investors to recover the cost of improved transmission reliability, it appears that the current climate of under-investment will continue and place the entire grid at continuing risk of outages such as occurred in August 2003. This is not surprising. Without some assurance that investments can be recovered in a timely manner, there is little incentive to make long-term capital improvements to improve reliability that will primarily benefit those who are not direct customers of the system. This attitude was reflected in the corporate policies of First Energy (the system where the August 2003 blackout initiated) which never considered voluntarily cutting off power to its customers to ease congestion on the power lines. Rightly or wrongly, the company placed its immediate business concerns above the security of the national grid and questioned why it should have interrupted power to its own customers to accommodate long-distance transmission through the system (Behr 2003)

The attacks on the World Trade Center also provide some interesting lessons for this discussion of sustainable infrastructure. New York City was able to recover relatively quickly (compared with how other cities might have fared) after 11 September primarily because of long-term investments in hardware and people that had produced highly resilient (i.e. reliable) systems (Wallace et al. 2003, Zimmerman 2002). Many of the New York service providers (e.g. Consolidated Edison, Verizon, AT&T, MTA) possessed considerable capacity in people who are considered international experts in their fields; state-of-the-art equipment and configuration management; as well as other physical and institutional resources necessary to affect recovery (O’Rourke et al. 2003) It is not apparent that leaner, less robust systems would have performed as well and as a result, it is likely that recovery would have been hampered and significantly delayed. The importance of this level of sustained strategic investment in infrastructure over a broad front that includes both physical systems and institutional capital cannot be overstated (Little 2004). Shortly after the terrorist attacks of 11 September, it was noted that the laxity of airport security had been an issue for years while calls to fund a professional, high-quality service had gone unheeded (Krugman 2001). This was ascribed largely to a national culture that was unwilling to pay the price of public safety and, which up until 11 September, was willing to depend on private companies to do a job that probably belonged in the public domain all along. A parallel was drawn to the decay of the public health infrastructure in the US and the concluding observation is germane to this discussion:

…if we continue to nickel-and-dime crucial public services, we may find—as we did last week—that we have nickel-and-dimed ourselves to death.
Although the case for sustainable infrastructure is intellectually compelling, implementing a strategy to achieve more reliable systems faces formidable challenges. The commons dilemma and significant free rider public good problems both conspire to starve infrastructure of the investment needed to improve overall reliability. So long as infrastructure is viewed primarily as local systems serving local customers, there will be little national incentive to address the problem holistically. A broader, unifying theme is required and a reconfigured manufacturing industry may provide it.

7. Infrastructure and supply chains

Manufacturing in the US is a $1.4 trillion activity and comprises over 12% of Gross Domestic Product (BEA 2004). For several years, manufacturing has been undergoing what amounts to a revolution and casts infrastructure in a far different light than that of just a necessary public service. Infrastructure is now seen as the enabler and backbone of supply chain logistics. As manufacturing has changed from a regional or sub-regional activity to one of globally networked enterprises, reliable transportation in all modes has become of paramount concern. Just-in-time deliveries from suppliers dispersed around the world to assembly plants that hold a few days or sometimes only a few hours of inventory requires ultra-dependable transportation networks. However, regardless of the length or complexity of these global supply chains, or the sophistication of the final assembly, they inevitably must depend on the reliability of the local infrastructure systems where components and subassemblies are manufactured. Additionally, in relentless efforts to drive down costs, major corporations are increasingly electing to deal with a single supplier of a critical component of the finished product. Any disruption in the production or delivery of these critical components can seriously disrupt an industry several states or even a continent away. Thus, the reliability of innumerable local infrastructures, and the combined reliability of the resultant ad hoc supply chain, takes on new significance for these industries. It is no longer sufficient to ensure quality services at the final assembly plant if the overall supply chain is breaking down because of local service disruptions to electricity, water, or transportation.

The automobile industry is an excellent example of such globally networked manufacturers. General Motors, for example, assembles over 100 different vehicles at plants located throughout the US and deals literally with thousands of suppliers, some of whom make critical components for which there are no readily available substitutes. Ensuring the reliability of these complex and extended supply chains is one of the industry’s major challenges in meeting production targets (David Skiven, Executive Director of Worldwide Facilities Group, General Motors Corporation, personal communication, 13 August 2004). For example, following the terrorist attacks on 11 September 2001 and the ensuing border closures and cessation of air traffic, several automobile manufacturers were within hours of halting production (Sheffi 2001). Similarly, during the 2003 power outage in the northeastern US, other industries in the southern and western states were shut down because critical parts were produced in the northeast and Midwest. Although many firms employed multiple sourcing strategies to avoid this scenario, they did not plan on multiple sources being lost owing to such a widespread infrastructure disruption (Reddy 2004).

At the present time, manufacturers are essentially on their own when it comes to ensuring the reliability of their supply chains. However, both the cost and complexity of guaranteeing dependable service is beyond the long-term capability of even major corporations. If acceptable levels of reliability cannot be maintained, US manufacturing will face another competitive disadvantage (labour and production costs already being high) when compared with nations where this is not an issue. The trickle-down effects of locally disrupted manufacturing productivity ripple through the national economy impacting everyone. As a result, a rationale exists that, since everyone benefits from disruptions avoided, everyone can reasonably be expected to contribute to the cost of avoiding those disruptions. How we might begin this task is discussed in the following section.

8. An approach to sustainable infrastructure

Deciding what levels of reliability should be provided and who should pay for it are not simple questions to be addressed solely by service providers, customers, or government regulators. All perspectives need to be considered in establishing goals and responsibilities for the public good component of civil infrastructures to ensure the sustainability of the joint ‘commons’ represented by these systems. In the case of natural resources, some kinds of common-pool resource problems are inherently difficult to perceive and assess particularly those that are very diffuse, mostly invisible or intangible, and not easily associated with particular consequences (McCay 2002). As we have seen, this is generally the case with infrastructure systems. Many cases where self-governance of a commons has not emerged can be traced to difficulties at the level of problem recognition and placement on an ‘agenda’. Groups may not be able to appreciate the magnitude of the problems confronting them because of the subtlety, novelty, or stochasticity of the system or imperfection in the monitoring system. They may be unaware or uninterested in the public goods associated with their private uses (McCay 2002). Self-governance, in the natural resources sense of fisheries or grazing land, is not a feasible option for
infrastructure, which suggests some form of external governance.

One of the criticisms levelled at Hardin’s pessimism at finding a solution to the commons dilemma was that he limited the options to centralized government or private property (Dietz et al. 2003). More recent thinking (and action) on commons issues has pointed towards multi-stakeholder governance structures which have been more effective at identifying and implementing approaches through their inclusionary nature. Accordingly, the emergence of institutions for governing the infrastructure commons should include not only rules and governance systems but should also be capable of effecting new and changed patterns of behaviour and inculcating different norms and values (McCay 2002). How people—individuals, organizations, communities, bodies of experts—are affected by and perceive the risks from deteriorating and increasingly unreliable systems is critical to whether and how they respond, including responses that affect the emergence of institutions for reducing or preventing those risks.

Since part of the solution to improving the public good component of infrastructure (i.e., the overall reliability of coupled, complex systems) will require considerable additional financing, a way to provide these funds by means other than direct user charges or general fund tax revenues must be found. The mechanism to accomplish this long-overdue task is beyond this exploratory paper but various excise and value-added revenues have been used for these purposes. In any event, this issue needs to be addressed and potential solutions explored in the short term. Funds from traditional sources will never be available to deal with this issue in the magnitude and for the timeframe required. Waiting until the time is ‘politically right’ will only forestall action and make any approach more challenging and costly. As a first step, the federal government could play the role of facilitator and initiate several necessary actions. Namely, a commission, appointed by the President or Congress, could proactively address this issue by beginning an informed and serious dialogue between the public and private sectors; building coalitions for action by assembling the stakeholders in the governance process; and exploring alternative mechanisms to provide dependable and sustainable resource streams for maintaining these vital systems. These are only beginning steps in a journey that will be long, arduous, and costly. However, failing to take action now to ensure the sustainability of our critical systems ultimately will have detrimental and perhaps irreversible, impacts on the nation’s long-term economic and physical security and overall quality of life.

Acknowledgement

Preparation and presentation of this paper was partially supported by the National Science Foundation under Grant No. CMS-0408577. This support is gratefully acknowledged.

References


American Society of Civil Engineers (ASCE). Report Card for America’s Infrastructure, 1998 (American Society of Civil Engineers: Reston).


Marburger, J., Testimony before the House Committee on Science. 14 June 2002.


