Introduction

Lifting the Veil

This book is about network infrastructures. We consider network infrastructures as socio-technological systems characterized by the interdependence and complementarity of their two dimensions: institutions and technology. Relying on a combination of nodes and links, these infrastructures require coordination along both dimensions, in order to fulfill functions hereafter identified as "critical." Critical functions determine the capacity of a network to deliver expected services in line with societal values. Thus understood, network infrastructures cover a wide range of sectors, from energy, water and sanitation, urban transportation, to telecoms and internet. These networks provide the backbone of economic as well as social activities. The key argument underlying our analysis is that alignment between the two dimensions, institutions and technology, is central to the fulfillment of performance expected from these networks. Misalignment can generate discrepancies or gaps challenging the integrity of a network and its capacity to meet its goal. The so-called New York blackout is a good illustration of this.

I.1 In the Dark

At 16:00 EDT on August 14, 2003, the most widespread electricity blackout in North American history paralyzed the city of New York and a wide corridor all the way up to Detroit (United States) and Toronto (Canada). All in all, about 50 million inhabitants were deprived of electricity and thousands of businesses were shut down, some of them for over a week, with a total loss estimated to range between USD 4 billion and USD 10 billion.

The causes of the outage are now relatively well known, thanks among other contributions to the detailed report from the US-Canada Power System Outage Task Force created after the event

https://doi.org/10.1017/9781108962292.002 Published online by Cambridge University Press

(US Secretary of Energy, 2004).[1] According to this report, the Cleveland-Akron area was highly vulnerable to voltage instability problems, largely because a private operator "had not conducted the long-term and operational planning studies needed to understand those vulnerabilities and their operational implications" (US Secretary of Energy, 2004: 23). This was the case because the company did not fulfill the standards and practices codified by the electricity industry through the North American Electric Reliability Corporation (NERC),[2] a behavior warranting dismissal apparently shared by other parties to the network. Indeed, according to the task force report, several operators in the Midwest consistently under-forecasted load levels the days before the blackout, with the institutional mechanisms of control (in this case delegated to the NERC) not perceiving the problem early enough and/or not responding adequately to the problem. So the cause of the outage would be "the inadequate understanding" of how the system worked and/or the inadequate performance of the tasks assigned.

Following the public outcry the outage caused, along with much associated media coverage, the American Congress promptly reacted in adjusting the regulation and in 2005 adopted the Energy Policy Act, which delegates to the NERC the responsibilities for drawing, implementing, and enforcing reliable standards throughout the US bulk power system. However, this rapid adjustment may not have fully taken into account the far-reaching technological changes that had already transformed the conditions of production and transportation of electricity and the significant institutional changes following the deregulation and liberalization of the sector since the creation of the NERC (in 1968). The new legislation, adopted in 2005, might well have underestimated the impact of these changes as a source of misalignment between institution and technology that weakened the system by challenging its coherence.

I.2 This Book in a Nutshell

Notwithstanding its specificity, the complex combination of technological as well as institutional flaws that created the greatest ever

[1] *Final Report on the August 14 Blackout in the United States and Canada: Causes and Recommendations*. US Secretary of Energy, April 2004. Available at www.energy.gov/sites/prod/files/oeprod/DocumentsandMedia/BlackoutFinal-Web.pdf

[2] The NERC is a nonprofit organization created by the electrical industry in 1968 to coordinate norms and standards in distribution.

disruption in the distribution of energy in North America points to the much more general problem of the potential misalignment between these two dimensions and its consequences.

Indeed, the New York blackout, more appropriately identified as the "Northeast blackout," provides an excellent illustration of the problem explored in this book: what are the sources of alignment (or misalignment) between the technological requirements and the institutional rules that shape network infrastructures and determine the success or failure of their organization? Blackouts in electricity systems, repeated train accidents, disruptions and delays in underground transportation, and faults in communication systems of self-driving vehicles all provide examples of failures in technological coordination or institutional coordination or both, and flaws in the alignment of these two dimensions.

Our analysis of this interdependence between institutions and technology, and of the resulting successes and failures in network infrastructures, is framed around the concept of *criticality*. "Critical" in this book is understood as pinpointing factors that can provide indispensable support, but which can also obstruct or even derail the fulfillment of functions we refer to as "critical functions." These functions, to be specified and discussed extensively in Part I of this book, are at the core of network infrastructures. The challenge they pose comes from the need to simultaneously coordinate and secure complementary technological and institutional entities and devices. On the institutional side, meeting this goal requires to define, allocate, implement, and enforce rights. On the technological side, it requires to identify, design, implement, and monitor the physical artefacts instrumental in providing expected services.

In the coming chapters, we explore how these two dimensions, the technological one and the institutional one, are interdependent and have to be in alignment in order for a network infrastructure to meet its assigned role. More precisely, we shall argue that the performance of network infrastructures depends on their capacity to fulfill four fundamental functions: capacity allocation, system control, interoperability among components, and interconnection between segments of the network. Figure I.1 summarizes this first insight into our approach.

I.3 Our Core Question and Hypotheses

Our driving motivation in developing this framework can be expressed as a quest to answer the following question:

https://doi.org/10.1017/9781108962292.002 Published online by Cambridge University Press

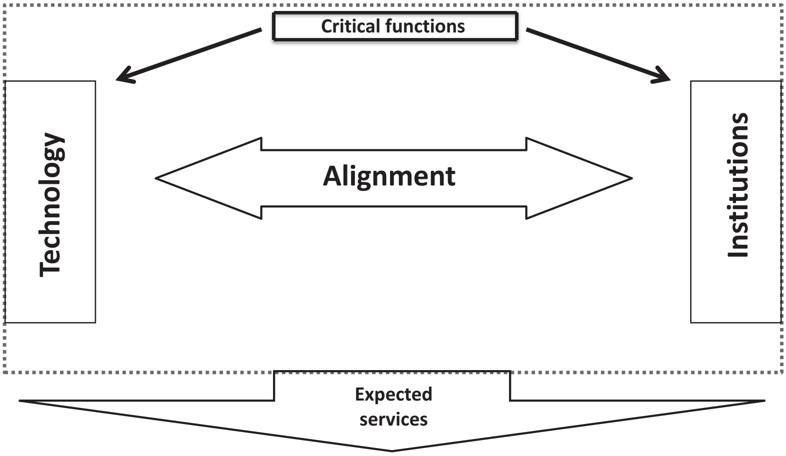

Figure I.1 Our alignment framework: first insight

What features can align the institutional and technological dimensions of network infrastructures, in order to obtain performance that meets societal expectations?

In pursuing this goal, the challenge we face comes from connecting the complex engineering systems of network infrastructures to their institutional embedding. What we are looking for is those institutional features needed to support specific network infrastructures; and, symmetrically, those technological requirements needed to make a specific network operational. We assume, and will substantiate in the coming chapters, that both the technological dimension and the institutional dimension need to be aligned in order to fulfill the requirements imposed by the critical functions, thus giving a specific network the capacity to deliver expected services. In exploring these issues, and notwithstanding differences across network infrastructures, we assume the existence of certain regularities; that is, the presence of reproducible and transposable patterns, which we aim to unravel. While elaborating on our theoretical framework in Part I of this book, and in order to better substantiate its empirical significance, we refer to examples from different network infrastructures, mainly the energy, transportation, water, and information sectors. In Part II, we explore the relevance of our theoretical approach through in-depth analysis of specific cases from a variety of network infrastructures.

More specifically, our analysis relies on the hypothesis that alignment operates differently but in complementarity along three interdependent layers: (1) a layer that defines the "structure" of a network, connecting its technological architecture to general norms and rules embedded in laws, customs, and other institutional settings; (2) a layer that models the "governance" of a network infrastructure, relating its specific technological design with the specific institutions that transform general norms and rules into customized ones and delineate the conditions of their implementation; and (3) a layer that corresponds to the planning and actual implementation of "transactions," providing support to the transfer and usage of rights and related resources in a way that makes technical features operational.[3] The central characteristics associated with these layers can be summarized in the following four hypotheses, and are substantiated in the theoretical analyses and empirical explorations developed in the coming chapters.

H1: *The structural alignment (or misalignment) between the technological architecture and the general institutional set of rules and norms in which network infrastructures are embedded conditions their existence and frames their properties.*

We identify this structural level as the macro layer. For instance, networks can have a centralized or decentralized mode of operation. Centralization (or decentralization) is heavily dependent on the technological architecture of the system and the general rules framing its usage. To illustrate, in most countries electricity networks have for a long time been built and developed through an architecture that imposes very tight constraints on the production, transportation, and usage of power, thus resulting in a strictly centralized system. At the opposite end of the spectrum, the internet owes its success to the widely decentralized characteristics of its architecture and conditions of access.

[3] With respect to this distinction between layers, we found initial inspiration in the scheme proposed by Williamson (2000: 597), who made a distinction between "Institutional Embeddedness" (Level 1 in his terminology), the "institutional environment" (Level 2), the level of "governance" (Level 3), and the level of "Resource Allocation and Employment" (Level 4). However, we transformed Williamson's scheme to deal with aspects his typology did not capture, particularly his "benign neglect" of the technological dimension.

H2: *General norms and rules must be translated and embedded into context-specific ones linking specific protocols and guidelines with context-specific technological features through modalities of governance.*

We identify this intermediate level as the meso layer. It involves entities and devices developed at the intersection between the structural features (rules and architecture) and the level of transactions at which the network becomes operational. For instance, following the Northeast blackout, the Energy Act of 2005 delegated to the NERC, a nonprofit corporation, the role of securing the critical functions for the North American electricity transmission networks. To meet this goal, the NERC operates as a private institution in charge of the definition of "reliable" standards and procedures, their implementation, and their monitoring.

H3: *Ultimately, rules and norms and their technological counterparts must be implemented through the organization of transactions that interconnect specific institutional arrangements and specific technical requirements, allowing a network to operate.*

We identify this operational level as the micro layer. This is the layer at which transactions must be carried out through the coordination of relevant agents within organizational arrangements that must meet and satisfy the technical requirements of the technology/technologies adopted to deliver the expected services. For instance, once a producer of electricity has been allocated rights of access to the grid under conditions established at the macro layer and its capacity to meet the technical procedures of access agreed by the entities and devices through which the governance operates at the meso layer, this producer must still organize transactions, for instance through contracts with providers of coal or uranium, in order to meet the standards of the technology chosen.

H4: *The alignment (or misalignment) along the three interdependent layers of structure, governance, and transactions determines the capacity (or failure) of a network infrastructure to deliver expected services.*

We shall argue in the coming chapters that the capacity of a network infrastructure to meet these requirements along the three layers, and to secure the alignment between the technological side and the institutional side, is central to the delivery of expected services by way of modalities that are economically sustainable and socially acceptable.

I.4 Underlying Premises

Our structuring question as well as our working hypotheses are built on the premise that network infrastructures share common features that make them different from other economic activities; for instance, the centrality of interconnection through nodes and links that impose challenging requirements with respect to coordination. Although it is the purpose of this book to substantiate this statement through the development of a conceptual framework and empirical tests, some initial insights might provide a useful guideline throughout the book.

First, we consider network infrastructures as socio-technological systems. Infrastructures are engineering systems that function in a specific social context. They perform intended functions, for instance, the safe and reliable provision of energy to households. Human agents purposefully design these systems; they monitor and adjust them in order to meet expectations that encompass values. To be sure, not everything in infrastructures is purposefully planned. Technology develops according to path-dependent trajectories. That the Netherlands has one of the most developed gas networks in the world is an important condition for the smooth transition to the use of biogas in residential areas, following the political decision to close down the Groningen gas field. However, the interest of farmers in providing biogas can certainly not be planned in every detail, and the willingness of consumers to adjust their behavior or of environmentalists to accept the fundamental changes in agricultural practices that biogas requires cannot be predetermined.

Second, coordination within each dimension, technology on the one hand, institutions on the other, is essential to the existence and running of network infrastructures. The different components of an infrastructure are not operating in isolation, they are interdependent. The local production of energy needs to be attuned with the technological properties and the capacities of the grid, the required energy quality, the limited possibility of storage, and of course the needs of final users. There is the necessity to implement institutional entities that will allow the coordination of the required technical features, in order to produce and deliver the expected services. Access to safe, drinkable water is a case in point: "safety" refers to parameters of quality that evolve over time and vary according to societies, the implementation of which also depends on available technologies or innovation. Coordination of each dimension is required and provides the alignment (or misalignment) that conditions the performance of the network.

Third, critical functions that characterize network infrastructures must be fulfilled for expected services to be delivered. System control must be implemented to make the network operational. For gas, this would be a certain calorific value and chemical composition, for electricity voltage (230 volts) and frequency (60 hertz), and so on. Capacity must be allocated and managed in a way that balances production and delivery with the actual demand. For electricity systems, this physical balance (called "load balancing") is a central feature, otherwise the system might collapse. In the railroad or airline industry, scheduling and monitoring the allocation of "slots" is a determinant factor for a secure system. Interconnection among segments of a network is needed to improve the technical functioning of the system, in order to benefit from externalities associated with the delivery of expected services. For instance, the reliability of rail transportation depends on the existence and quality of interconnection between the local, regional, national, and even international networks. Last, interoperability among different parts of a network requires that they are technically equipped and institutionally monitored in a way that fits the technical needs of the system. For instance, solar panels need to fulfill certain technical requirements in order to be connected to the electric grid; signal systems on board trains need to be tightly coordinated in order to secure passengers' transportation.

Ultimately, our main concern when exploring these properties of network infrastructures and the interdependence between technology and institutions that characterize them is about the outcome. How can network infrastructures be technically reliable, institutionally feasible, and perform satisfactorily thanks to the proper alignment between these two foundational dimensions? At this stage, our approach remains static. We consider the alignment issue at a certain point in time. Although we may refer occasionally to dynamic aspects, for example, the impact of innovation on the organization of electricity grids, we do not develop a dynamic model. In that respect, we are aware that criteria other than alignment could be introduced to assess performance, for example, sustainability (which refers to co-evolution) or affordability (which refers to pricing strategies and financial considerations).

What we do explore in this book is how the factors commanding the interdependence between the macro, meso, and micro layers identified above shape the capacity of specific networks to meet expectations and fulfill the objectives they are assigned. For instance, in the railroad industry passengers expect trains to be on time, to be safe, to provide a certain level of comfort, and to make connections easy. These expectations translate into substantial technical requirements as well as rules implemented and enforced in a way that allows transactions to be accomplished efficiently.

I.5 Outline of the Book

Based on this introductory discussion of our core concepts, which will be extensively developed in the coming chapters, we can reformulate our initial figure, in order to better capture the hard core of our framework.

Figure I.2 summarizes and somewhat anticipates developments explicated in the coming chapters. These chapters are organized in two parts.

Part I ("Conceptual Framework") is analytical. It develops our core concepts and explores in some detail how they are interconnected, thus defining an integrated and coherent approach to network infrastructures. Chapter 1 examines the features identified as typical in the literature on network infrastructures and points out substantial flaws, particularly when it comes to taking into account the societal values involved in the determination of institutional rules and in the choice of specific technologies. As well as the introduction of key concepts, we emphasize how value-loaded is the domain of network infrastructures when understood as socio-technological systems. Chapter 2 focuses on the institutional dimension of our framework. Building on contributions coming mainly – although not exclusively – from the new institutional approach, which has its own flaws, particularly when it comes to the analysis of technologies, the chapter provides an in-depth study of the characterization of institutions as composed of interdependent layers identified through specific concepts. The chapter emphasizes the different institutional modalities through which value-oriented network infrastructures are ruled. Chapter 3 turns to the technological

https://doi.org/10.1017/9781108962292.002 Published online by Cambridge University Press

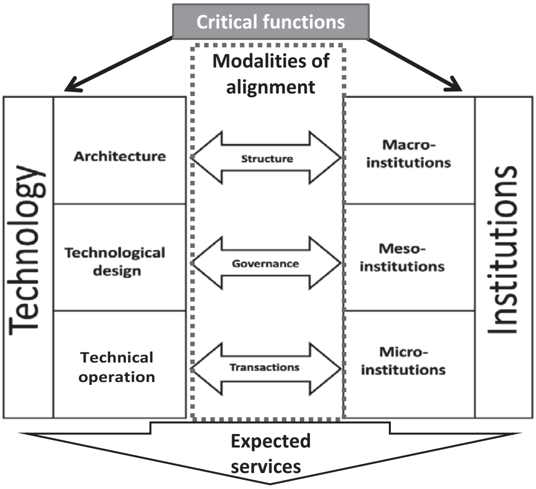

Figure I.2 Our framework with its key components

dimension. Paralleling the previous chapter, it identifies distinct but interdependent layers through which the technological architecture of a network infrastructure is translated into technological features that need to be made operational. Building on engineering approaches, it emphasizes the modalities through which the technological dimension of the critical functions can actually be fulfilled. Chapter 4 puts together these elements to deal with the central question of our book: what conditions regarding the alignment of institutions and technology of network infrastructures must be fulfilled, in order to meet the critical functions that determine the capacity of these infrastructures to deliver expected services. The chapter emphasizes the sources of alignment (or misalignment) between the two interdependent dimensions of our framework.

Part II ("Empirical Exploration") delivers analyses through which our framework is mobilized to assess factors of alignment or misalignment of specific network infrastructures. Chapter 5 takes stock of the complex transition toward sustainable energy systems to show the structural factors that make macro-institutional rules and norms and the architecture of energy networks interdependent. The emphasis is on the tension between the existing centralized arrangements and the emergence of alternative technologies that support a radically different structure. Chapter 6 focuses on the critical provision of water services to a major city, and draws lessons from what is considered to be a success story to illustrate how modalities of governance have been implemented that allow the appropriate alignment between meso-institutions and the technological design of a system based on four substantially different technologies. The emphasis is on conditions that made the alignment appropriate, but also on factors that may plague the governance of an otherwise successful network. Chapter 7 discusses the ongoing technological changes in urban transportation through the case of the introduction of self-driving vehicles. It shows how the modalities through which transactions are established and monitored impact all three layers of our framework because of the societal values that such emerging technologies challenge. The emphasis is on the interdependence of layers exhibited by technological changes that require the transformation of existing social norms.

Chapter 8 concludes by pointing out the novelty of our approach, its relevance for the study of specific empirical cases, and also the directions in which new developments could enrich the conceptual framework. The chapter also comes back to policy issues raised by our framework and the different cases explored. In doing so, we also discuss some limitations of our analysis, thus opening up room for future research.

Throughout the entire book, our focus is on identifying and characterizing the different and interdependent layers that shape the institutional as well as the technological dimensions of network infrastructures. Our goal is to provide conceptual tools for assessing the conditions under which the alignment or misalignment of the two dimensions and their internal coordination determine the performance of specific network infrastructures.

https://doi.org/10.1017/9781108962292.002 Published online by Cambridge University Press