design and management of infrastructures

The Infrastructure Playing Field in 2030

Editors
Prof. dr. ir. M.P.C. Weijnen
Prof. mr. dr. E.F. ten Heuvelhof

The Infrastructure Playing Field in 2030

The Infrastructure Playing Field in 2030

Proceedings of the
First Annual Symposium

Delft Interfaculty Research Centre
Design and Management of Infrastructures

Noordwijk, November 19, 1998

Editors:

Prof. dr. ir. M.P.C. Weijnen
Prof. mr. dr. E.F. ten Heuvelhof

Published and distributed by:

Delft University Press
P.O. Box 98
2600 MG Delft
The Netherlands
Telephone: +31 15 278 3254
Telefax: +31 15 278 1661
E-mail: DUP@DUP.TUDelft.NL

ISBN 90-407-1884-9

Printed in the Netherlands

Contents

Foreword

Seen van der Plas

Former Secretary-General of Ministry of Transport, Public Works and Water Management

Chairman Management Advisory Board of DIOC Design and Management of Infrastructures

A new economic order is quickly taking shape, due to the development of technology and logistics and fuelled by the seemingly unlimited possibilities in the field of information and communications. Distance will become less relevant in this new order and managing time and knowledge will be determining factors. The world will become one market. The international companies will focus even more on their core-competences, and companies will form world wide networks. Mass-individualisation, reversal of the economy from supply to demand oriented and opportunities like electronic commerce are flagpoles of this new trend. The value put to nature and environment will contribute to direct a sustainable economy. This development shows the relevance and urgency of the Delft Interfaculty Research Centre "Design and Management of Infrastructures".

"Infrastructure is the backbone of society and of the economy." That is the foundation of this important and ambitious infrastructures research programme.

The programme is important because efficient and effective infrastructures are prerequisites for the availability and qualities of vital goods and services for society and the economy, brought about by the energy, water, transportation, telecommunication and waste removal sectors. It is important for the position of our country in production, commerce (trade) and transport. It is important for the Delft University of Technology as well to take a frontline position in particularly this research area.

The programme is ambitious because of its strong interfaculty and interdisciplinary character. It is ambitious especially for the very concrete objective being a generic approach for the development of approaches, methods and tools for policy, design and management of infrastructures.

This book contains the proceedings of the First Annual Symposium, which marks the beginning of a research programme that will gradually and increasingly show its relevance. The synthesis of science and practice will determine its value. The management advisory board finds its mission in the establishment of this synthesis, with appreciation and trust in the way the programme is set up.

Foreword

John R. Ehrenfeld
Director
Massachusetts Institute of Technology
Technology, Business & Environment Program
Center for Technology, Policy & Industrial Development

Chairman Scientific Advisory Board of DIOC Design and Management of Infrastructures

This Symposium, dedicated to explore The Infrastructure Playing Field in 2030, marks an important turn in the research community of the Netherlands and beyond. Infrastructure is the essential framework in which all societal everyday activities takes place. The infrastructure that permeates our societies today both shapes and is shaped by the dominant cultural structures that characterize our modern, industrial world. As those underlying cultural beliefs and norms shift, so must the infrastructure. Today we are all facing a new world in which many of our beliefs must change. The idea of sustainability, which is my own personal area of research, represents a growing awareness that current cultures and their tangible elements such as infrastructure are no longer capable of reproducing life, human and otherwise, in a satisfactory way. Nowhere in the world more than here in the Netherlands has a society explicitly begun to address these concerns.

Among other signs of a need for change, the constraints of sustainable production and consumption are forcing those who design technological infrastructure and its institutional and policy framework to rethink their basic assumptions and their methods. This Symposium marks a commitment by the Dutch research community to move the rethinking forward. The presence of so many from all the sectors involved in the research – academia, industry and government – signals a commonality of interest and expectation to work together that bodes very well.

The speakers addressed the problems in two important dimensions – as sectoral or topical and according to crosscutting themes. Both are important elements in a comprehensive research program. Transportation issues, for example, demand a focus of their own. Transportation planning and design processes are unique to the sector. The players and the data they use are specific. But transportation, as an example, is also intimately connected to other sectors and interests. Successful development of new, effective infrastructure in this or any area demands close coordination with other sectors. The speakers have also discussed the need for and plans to develop crosscutting elements in this new project. Such dual research that follows the two axes of the matrix that Professor Margot Weijnen presented is rare. The project coordinators, funding agencies and other participants deserve much credit for the very substantial breadth and depth of this program. I can only say that, from my vantage point across the Atlantic at MIT, I am a bit envious. I look forward to a continuing role in this project and commend all those involved for the ambitious start they are making.

Introduction

Margot P.C. Weijnen

Programme leader
Delft Interfaculty Research Center for the Design and Management of Infrastructures,
P.O. Box 5069, 2600 GA Delft, The Netherlands
e-mail: M.P.C.Weijnen@infrastructures.tudelft.nl

This first volume of papers produced by the research team of the Delft Interfaculty Research Center for the Design and Management of Infrastructures is based on the papers presented at the First Annual Symposium of the center, held in Noordwijk, November 19, 1998. It is a volume of 'work in progress' rather than a collection of fully crystallized scientific papers. As such, it is aimed at informing a wide and varied audience of the progress being made by the research center. Many are showing a clear interest in our work, ranging from national to local authorities, technological research institutes and infrastructure sector-based research institutes, individual utility companies and branch organizations, engineering consultants and management consultants, et cetera. This overwhelming interest is not surprising in a time when utility markets are being liberalized and re-regulated, and utility companies and their partners are re-defining their mission and strategies.

The subject area of the research center encompasses a variety of infrastructure sectors: energy, telecommunications, water, waste and transport. Some of these sectors have almost completed the transition to fully liberalized market conditions, others have only taken their first hesitant steps. As each infrastructure sector is more or less facing the same research questions, and considering the commonalities and analogies between the infrastructure sectors, the Delft Interfaculty Research Center for the Design and Management of Infrastructures decided not to allocate all its research capacity to sector-specific research, but to make substantial capacity available to study the design, operation and management of infrastructures from a generic, supra-sectoral, perspective. The full collection of sector-specific and generic research projects making up the program is depicted in Figure 1.

The mission of the Delft Interfaculty Research Center on the Design and Management of Infrastructures is to *perform comparative analyses of the technological, economic and administrative developments in different infrastructure sectors, with a view to identifying commonalities and* *interrelationships. The commonalities and interrelationships identified shall serve as the basis for the development of generic approaches, methods and tools to support the future design and management of infrastructures and the policy making on infrastructure development and management.*

The program is truly multi-disciplinary, as it involves research staff from seven sub-faculties of the TU Delft:

- Applied Earth Sciences
- Civil Engineering
- Chemical Technology and Materials Science
- Electrical Engineering
- Mechanical Engineering and Marine Technology
- Philosophy and Technical/Social Sciences
- Systems Engineering, Policy Analysis and Management

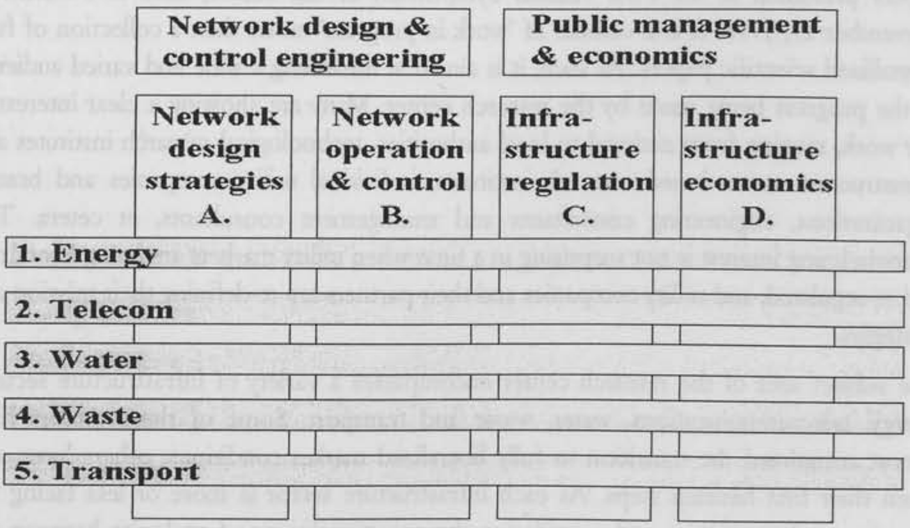


***Figure 1.** The collection of research projects comprising the research programme of the Interfaculty Research Center for the Design and Management of Infrastructures, at the Delft University of Technology. The sector-specific research projects are depicted in the rows of the matrix, the generic research projects in the columns.*

In view of the wide range of disciplines involved in the research program, each with its own culture and research traditions, the research team felt it needed to invest in the building of a shared vision and vocabulary. As this was recognized as a *conditio sine qua non* for effective multi-disciplinary research, a scenario workshop was organized in May 1998, where all members of the team contributed their views and expertise to a scenario analysis of each of the five infrastructure sectors by the year 2030. The method of scenario analysis used in the workshop is described in the paper by Wil Thissen, who made all members of the research team familiar with the method, the uses and limitations of scenario analysis, and who chaired the workshop.

Most of the sector-specific papers in this volume are built on the scenario analysis for the specific infrastructure concerned, with a time horizon stretching to the year 2030. The authors used scenario analysis to identify *robust* research questions for the infrastructure sector in their area of research, in other words, those research questions that will need to be answered in most, if not all, possible futures that their infrastructure sector may be facing in the decades to come. As the sector-specific research projects are primarily technology oriented, and most of them headed by technological experts, most of the infrastructure scenario analyses focused on the technology determinants of infrastructure development and on the technological research questions to be tackled in the research projects.

The generic research papers in this volume were not all based on infrastructure scenario analyses in a similar straightforward manner. In the papers by Ten Heuvelhof et al., Van Twist et al., and Weijnen and Bosgra, the authors chose to present the core dilemmas being faced by the national government, regulators, and design and control engineers, respectively, in the turbulent period of transition towards liberalized utility markets, and in view of the large uncertainties about the outcome of the transition process. For each type of decision on infrastructure planning, financing, engineering design and control, and public management these actors are or will be facing, the authors present the two extremes, and argue how to arrive at a balanced decision.

The paper presented by the infrastructure economics research group illustrates how ill-defined the concept of infrastructures really is. The classical picture of infrastructures as natural monopolies, based on highly capital intensive, physical networks needs drastic re-adjustment, in view of new technologies that break down the natural monopoly characteristic of existing networks on the one side, and, on the other side, in view of the rapid development from an industrial based economy towards a highly knowledge intensive, service oriented economy. Quoting one line of Bill Melody's presentation during the symposium: *'All infrastructure is knowledge'*. At this stage of the program, however, the research team is still too fully occupied with acquiring an in-depth understanding of the physical infrastructures and the dynamic behaviour of these complex systems, to even dare to include the hardly tangible concept of knowledge infrastructures in the subject area of their research.

The need for concrete and tangible results of our research is strongly felt by the policy makers confronted daily with difficult decisions on e.g., infrastructure planning and capacity management on the existing infrastructures. In this turbulent transition period, their time horizon is only too easily compressed to that of the day or tomorrow, yet they recognize that the choices being made now will largely determine the degrees of freedom for infrastructure innovation in the future. In the introductory paper to this volume by Westerduin, the challenges to achieve reliable, efficient, safe and environmentally friendly solutions to the current mobility problems are elucidated and, to effectively meet these challenges, they emphasize the role of technological innovation in the transport infrastructures.

This volume of papers, the first report of the Delft Interfaculty Research Centre for the Design and Management of Infrastructures on 'work in progress', provides a rich picture of the challenges for infrastructure design and management in the future as the researchers, from their variety of perspectives, perceive them. On behalf of the research team, I express my hope that the readers share our fascination for the extremely interesting, but highly complex research subject of infrastructures in transition, and that they will appreciate our first efforts to structure the subject matter. The larger the effort, the more rewarding the results will be. The research team is determined to continue its fundamental and systematic search for innovative approaches, methods and tools to support effectively the future design and management of infrastructures, and the policy making on infrastructure development and management.

Acknowledgement

Finally, I would like to express my gratitude to Wil Thissen and his staff for their efforts in preparing and guiding the infrastructure scenario analysis workshop in May 1998. Wil Thissen is also acknowledged for his perfect role as chairman of the First Annual Symposium. I would furthermore like to thank Mieke Boon for her determination to get things properly organized, and the inspiration she provided for the scenario workshop. For the production of this book, I would like to acknowledge Miranda Aldham-Breary for correcting our English, and Paulien Herder, Sandra Junier and Connie van Dop for their indispensable efforts in preparing the camera-ready papers.

Delft, March 1999
Margot Weijnen

A Scenario Approach for Identification of Research Topics

Wil A.H. Thissen
Delft University of Technology
Faculty of Technology, Policy and Management
Policy Analysis group

Abstract

Scenario approaches have been designed to assist in developing strategies under conditions of uncertainty. One of these approaches is described specifically. Its adaptation and application to identification of research topics on infrastructures are outlined.

1. Scenario approaches

The term 'scenario' is used widely to indicate pictures or images of possible real-world situations or developments. In the movie-industry, a scenario means a storyline. In military research, a battlefield scenario means a possible sequence of events emerging from the combination of actions of the enemy and reactions of one's own forces.

In policy making and planning - our prime field of attention - the term scenario is used to indicate possible future situations or developments. There is, however, not something like 'the' scenario method, but a wide variety of approaches has been developed each of which may be used to serve another purpose. All these approaches are being labelled 'scenario-approaches'.

The following types of scenarios are generally distinguished:

- images of *possible* future situations at a specified time in the future, e.g. a view of the Netherlands' housing situation in 2030
- images of possible future situations and of the path(s) leading from the present to that future situation
- images of a *desired* future situation, for example a sustainable society.

The first two types are called *explorative* scenarios, and are associated with forecasting, i.e. exploring possible futures starting from the present. The third type are called normative or *prospective* scenarios, in which case the search of a path leading from the present to the desired future situation is referred to as 'backcasting'.

Further distinctions can be made with respect to the aspects described in a scenario. Some analysts use the term scenario exclusively to describe possible exogenous situations not under the control of the planner or decision maker, for example the amounts and patterns of rainfall that may have to be dealt with by a sewer system of a city. Others use the term to describe possible action paths or policies of a decision maker, for example, in energy policy, a nuclear, fossil fuel or renewable energy scenario.

We will limit ourselves here to one of the various scenario approaches, developed specifically to support strategy formulation in the context of uncertain future developments.

2. Scenarios for strategy formulation amidst uncertainty

Planners in industry and government, in the past, have often attempted to forecast the most probable future situation their firm or agency would have to deal with, and then used this as a basis to develop or select a policy or strategy. During the past decades, however, an approach has been developed which, instead of attempting to select a single best *prediction* of the future, acknowledges the inherent unpredictability of future circumstances. Therefore, the focus is on identification of a spectrum of *possible* future situations (Van der Heiden, 1996, Schwartz, 1993, Wack, 1985).

Starting point is a situation in which an actor or a firm wants to develop a strategy, acknowledging that he or she has no control over exogenous developments that may, however, have a significant impact on the effectiveness of the strategies. Conceptually, the approach therefore makes a distinction among (see Figure 1 below):

- strategies or policies available to a decision maker, affecting the system over which the decision maker has some control (the management of a firm can control the firm's product strategy; the ministry of transport can influence the transport system through extensions of the infrastructure, changing traffic rules, etc.)
- factors affecting the effectiveness of the system over which the decision maker has no control, such as international economic developments.

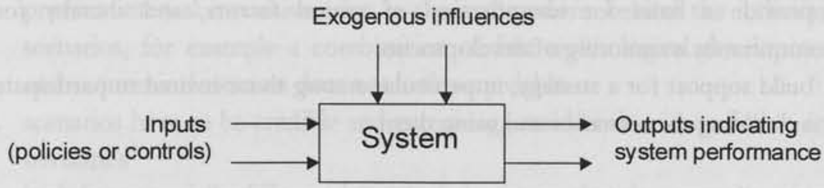


Figure 1. Inputs, outputs and exogenous influences to a system

The scenario approach discussed here generally includes the following steps:

1. Identify the decision context and objectives, i.e., what is the decision scope, what is the system and what is considered to be the surroundings, what is the time frame of relevance, what are the key indicators against which the performance of the alternatives is to be evaluated.
2. Identify those factors exogenous to the system that may significantly affect performance, and the future development of which is uncertain. Based on these, develop a small but varied set of possible future circumstances or scenarios in which the system will have to operate.
3. Identify strategies or policies aimed at reaching the objectives, taking possible future circumstances into account.
4. Evaluate the strategies' effectiveness for the different future contexts, and choose or compose a preferred strategy according to some principle, for example minimize the risk of large losses, maximise the chance of big gains, stay flexible, etc.
5. As time evolves, monitor developments as they occur, identify the development direction that is becoming real, and adjust the strategy as desired.

This approach has been used successfully in a variety of cases. A notable example is provided by Shell Oil Company in the seventies who, in a scenario-exercise, had identified the possibility of an oil crisis situation and adequate response strategies to it before the crisis actually occurred. Faced with the real occurrence of what was only a possibility before, Shell was ready to deal with it much better than the competitors.

Recently, the approach is being applied to support policy making in both the private and the public sector. The approach, in general, helps to:

- raise consciousness about the uncertainty and variety in possible future situations
- stimulate creativity in designing solutions to deal with the variety of possible circumstances
- make more deliberate choices when facing uncertainties

- provide a basis for identification of critical factors, and thereby for continuous monitoring of developments
- build support for a strategy, in particular among those invited to participate in building the scenarios and using them.

3. Developing scenarios

The generation of scenarios as mentioned in step 2 above is not trivial. It requires knowledge of the field and of the decision situation for which the scenarios are developed, imagination and creativity, and craftsmanship in the synthesis of a limited set of useful scenarios. In general, the following step-wise approach is suggested for scenario development:

(a) Specify the type of scenario one is to develop, and the relevant time frame. In addition to the general decision context, it must be decided whether the scenario will include a development path to the future or just a possible future situation at a given time, and it must be established for what future time frame scenarios will be designed (e.g., 2020 or 2050).

(b) Identify the exogenous factors that have an impact on system and policy performance, for example the demand for electricity for the energy supply system, or the climate situation for the water management system.

(c) Select those factors:

(1) the future development of which is highly uncertain, and

(2) changes in which will have a significant impact on the performance of the system or policy of concern.

(d) For the factors selected, identify the major driving forces behind change. For example, for electricity demand, these driving forces could be population size, general economic development, or changes in spending patterns by consumers.

(e) Reduce, if necessary, the number of relevant, independent driving forces to a total of 8 or less. This may be done by further selection based on variability and impact, or by aggregation.

(f) Design three or four future scenarios based on combinations of different developments in the driving forces.

(g) Label each of the scenarios by an easily recognizable name, and provide a brief but imaginative description of the essential characteristics of the future depicted by the scenario.

The design or synthesis step (f) in particular requires craftsmanship. Some important suggestions for this step include:

- prevent internal inconsistencies in the assumptions behind the individual scenarios, for example a combination of fast technological development and economic downturn does not look very viable
- scenarios have to be credible and surprising; avoid both science fiction and trivialities
- look for essentially different images of the future that do cover the variety in what is possible or thinkable
- avoid thinking in terms of probabilities, for example by developing one scenario that seems to be much more probable than others; similarly, avoid classifying scenarios as 'high', 'middle' and 'low' scenario, as the tendency of the users will inevitably be to adhere much more weight to the 'middle' scenario, rather than taking the possibility of all scenarios into account.

4. Application to research program design

The approach outlined above was developed to support decision makers in policy formulation amidst uncertainty. To a certain extent, this is analogous to developing a research program with the objective to provide contributions in the longer term. More importantly, one of the intentions of the program leaders for the DIOC on infrastructures is to identify research challenges at the overall system level in a systematic way, taking possible future developments and challenges into account. A structured approach that stimulates creativity in thinking about possible future situations is therefore needed. In addition, as the programs and subprograms are newly formed and require participation and co-operation by scientists from different faculties and research cultures, an approach that can help build a joint platform for deliberation is very welcome.

Therefore, the scenario approach outlined has been adopted to assist in identification of research topics. A few adaptations are necessary, however. First, the users of the approach are not policy makers confronted by the need to make decisions regarding the infrastructure, but researchers who need to make decisions about their research subjects. Therefore, the scenarios to be designed have to address not only the possible developments *exogenous* to the infrastructure systems of concern, but also *internal* factors essential to the functioning of the systems, such as the introduction of new technologies, or changes in the way the systems are being governed and managed. Second, as one of the key objectives of the first program phase is to broaden the field of attention and enhance creativity, the emphasis has been on identification of possible future scenarios and associated research or design questions, rather than on the selection of research topics. The selection will follow in a subsequent phase.

To stimulate creativity and prevent sticking to the present situation, scenarios have been developed for a future thirty years from the present: a point far enough in the future to allow for the possibility of major changes in the systems and the surroundings.

The exercise has been set up as a joint activity of the scientists in the program, sub-groups addressing the different infrastructures considered or other cross-cutting themes. A plenary workshop was held in May to kick off the process. As most participants in the program were not familiar with the scenario approach, the emphasis in the workshop was put on explaining the approach and on a first, quick effort at designing scenarios under the guidance of a person with experience in applying the method. Afterwards, using the workshop experience as starting basis, the subgroups elaborated the scenarios, leading to the results and views reported in the remainder of this volume.

References

Heiden, K. van der : Scenario's, the Art of Strategic Conversation. Wiley, Chichester, 1996

Schwartz, P.: The Art of the Long View. Century Business, 1993.

Wack, P.: Scenario's; Uncharted Waters Ahead. Harvard Business Review, sept/oct. 1985, pp. 73-89.

Wack, P.: Scenario's; Shooting the Rapids. Harvard Business Review, nov/dec. 1985, pp. 135-150.

Challenges for Infrastructure Policy-Making in a Changing Environment

Bram Westerduin

Director-General for Freight Transport

Ministry of Transport, Public Works and Water Management

In this highly developed society of ours, a high-grade infrastructure that functions well is one of our most basic needs. There is no discussion about that. It is only when we ask how we are to meet this need that discussion arises. That is not surprising, since infrastructure has an immense impact on both the environment and our economy. Here I mean not just how we construct new infrastructure, but also how we use what we have. Ninety-five percent of the infrastructure in 2010 exists of is planned already, so utilisation is at least as important as expansion. We are mainly looking for the solutions technology can offer us. The challenges for infrastructure policy makers have therefore to do with finding the best ways of utilising existing infrastructure, and to fit new infrastructure in our economy, in the space available, and in the environment.

This applies, I feel, whatever the infrastructure and whatever the sector. An interesting question is therefore whether we can discover overall patterns and overall design and management principles. As you will understand, I am very interested in the approach adopted here in the Delft university.

But I shall start from the angle I know best. My field is transport, and my primary interest is thus transport infrastructure. I shall therefore first give you my view of the future of transport. Against this background, I shall then suggest a number of topics for research, which, I expect, will challenge you to embark on more generic and comparative infrastructure studies.

We see a growing interdependence between infrastructure and its utilisation. That applies to all sectors. When it comes to transport, infrastructure policy is part and parcel of overall policy on mobility. The first questions we need to answer, therefore, are about mobility. How is it going to develop in our country, and in Europe as a whole? How do we assess these developments, and how do we approach management problems? Only then we can decide on ways of managing our infrastructure in the future.

Mobility is undoubtedly an immense problem area for society. It is also high on the political agenda. So how important are mobility and accessibility in relation to other social issues?

In the past few years, the ICES, the committee responsible for advising the government on economic policy has stressed the view that Dutch society has become largely dependent on transport, and that accessibility has become the dominant factor in economic and spatial troublespots.

I would like to give you a few facts to illustrate this.

The transport sector is good for 375,000 jobs, for generating 7% of our national income, and for contributing more than ten billion guilders to our balance of payments.

In the past, mobility had a vast influence on major land-use processes, such as suburbanisation, which in turn led to commuting. Even today, development proceeds alongside the main transport arteries. The decision in the 1960s to build a tangential highway network has led to new economic zones along the grid structures on the peripheries of our major cities. Economic corridors have been grown along the roads into the hinterland. But because infrastructure is both long-lasting and highly capital intensive, it can have a paralysing effect on land-use planning. I therefore believe that we can respond more effectively to the kind of natural processes I have just sketched.

Still on the subject of mobility, the next question that arises is the scale of the problem of accessibility, namely congestion and pollution.

Looking at how the transport system in the Netherlands and other densely populated parts of Europe now works, how our road network is slowly approaching gridlock, and how public transport can more or less go no further; and if we consider that mobility is increasing, that there are more and more cars on our roads, that more and more people are using their cars in their free time and on holiday, that road haulage will continue to grow rapidly and that we are not going to meet our environmental targets, we can only conclude that what we are dealing with is a social problem of the first order.

There are also a number of external factors that further complicate to find solutions.

- First, the problem is caused by complex social processes, and the question is whether we can identify angles from which we can best influence them. The Ministry can only play a minor role here.
- Second, there is no single solution that enjoys broad public support. Groups and individuals all have their own way of looking at the problem. For many companies, for example, transport costs play little or no role in their strategic decisions, and it is with amazement that I see people opting to get stuck in the traffic jam on the road from Almere to Amsterdam every

morning instead of taking the train on the adjacent track. Everyone seems to choose their own solution, and policy aimed only at the supply side does not work.

- Third, it is a complex problem because it affects a whole mesh of policy areas. In the current coalition agreement, for instance, the government states that over the next few years we are faced with radical decisions on how our growing population is to work, to live, to travel and to recreate in a clean, safe environment. A blueprint with wonderful, technological solutions is not the answer. We have to think instead of cohesive policy packages.
- Finally, the international dimension is beginning to play an increasingly dominant role. More regions are faced with the same problems, and we are enmeshed in international political, governmental and infrastructural networks.

So far a short outlook on the problems. How can we tackle them? My conclusion is that we have to develop a strategy with a multi-actor approach, with short term steps embedded in an overall long term strategy. The process of strategy development must be a process in clear steps, of with the basic first steps are an evaluation and assessment of the current situation, followed by a clear formulation of aims, vision and strategy.

I would now like to elaborate on each of these two steps. How, first of all, do we assess the current situation? I started my talk today by outlining the way things are at present. But which of these trends and developments do we need to take account of? What scenarios are possible?

The following trends have, I feel, the greatest impact on accessibility.

1. Growth in mobility:
   Despite the current global crisis, the trends are towards growth: the economy is growing, world trade continues to increase, the population is rising and ageing, and mobility continues to grow. At the other hand, the development of information and communication technology, together with the growing flexibility, makes people increasingly freer to choose when and how they travel.
2. A stronger position for the consumer, and flexible supply
   The production chain is becoming more and more demand-driven. This is leading to more flexible production methods and to new logistic processes.
3. The ongoing geographical spread
   Society is becoming more flexible, and we are tending to make more flexible use of space too. Living and working are gradually becoming footloose. People change jobs

and move house more easily, and distances are loosing importance. This is leading to a diffuse pattern of home-work relationships. To add to this, the economy is fanning out into corridors south and east of the Netherlands.

4. Stricter environmental standards
People demand a cleaner environment, and are both able and willing to pay for it. Translated into political terms, this will lead to higher environmental standards.
5. Innovative new technologies
It is clear that technology is going to play a greater role as we develop into a high-grade society and service economy. This means mainly the emergence of new systems and concepts, especially in logistics. Because innovation is very costly and entails considerable risk for the business community, and because it has far-reaching implications for society as a whole, it is higher and higher on the political agenda. The role of government is to create conditions and tighten up the requirements to be met by industry.
6. Internationalisation and regionalisation
Power is shifting to a higher level, namely to Brussels and head offices. On the other hand, competition between economic regions - which often straddle national borders - is beginning to intensify.
7. The market as playing field and increasing government intervention
The trend towards market forces is continuing; the government is becoming an increasingly powerful market player in its role as director and regulator. I believe that accessibility will increasingly be accepted as a scarce commodity, and the government will no longer be expected to guarantee free passage at all times and at any cost.
8. Strengthening the price mechanism
Finally, the trends I have listed will lead to the deployment of the price mechanism and it will come to be accepted that a price must be paid for commodities such as accessibility.

In short, quality is an increasingly important aspect of society in the Netherlands as in other parts of Western Europe. People want more choice, and they're willing to pay for it. This means an ever more important role for the price mechanism in many sectors, including transport.

What we now have to do is identify the problems connected with these trends and what we can expect on the transport front if policy is unchanged. The zero scenario, I believe, is that everybody seeks his own solution, rush hours become longer and longer, the system becomes less and less reliable, and poor accessibility prevents enterprises locating in the main economic centres. I predict an even more uncontrolled spread of activities away from existing urban areas. In the short term, congestion will, in any event, increase.

It is no longer inconceivable that one wet and windy autumn morning we see one long traffic jam, with the tailback leading seamlessly into its own front.

This zero-scenario is not acceptable and the question is what strategy should we choose, and what are our guiding principles?

Our main aim is to combine economic growth and sustainability. Our basic principles in this regard are to strengthen the idea of the Netherlands as a distribution centre, and to take account of the international policy framework.

That means our aim is to develop a transport system that:

a. supports the economy, since it:
- contributes to Dutch industry's international competitiveness
- generates employment
- supports the development of the regions;

b. serves the common good, since it:
- provides every member of the public with an acceptable level of accessibility
- does not place an excessive burden on the environment
- is safe;

c. is efficient and affordable.

How will this look like; do we have a shared focusing vision on that system of the future?

Sure is that developments in transport will go towards more efficiency, more intelligence and more choice. Transport in the 21st century will be marked by the provision of "multiple choice". So my vision of the future of transport is:

- The transport system of the future is inter- and multimodal. Goods and passengers will both be transported in chains of movements, in which nodes will play a major role. They will determine the quality of transport and have a major impact on spatial planning, to which urban and regional authorities will respond.
- Cars will be clean and energy efficient and will therefore still play an important role. Companies are working hard to achieve this. Multinationals like Shell and Daimler Benz work on concrete plans and others will not want to be left behind.
- The boundary between public and private passenger transport will become blurred, partly under the influence of technology.
- We see rapid mass public transport on national and regional transport axes for commuter and business traffic, linked to efficient networks centred on

the big cities and "customised transport" with a mainly social function in rural areas.

- Europe will have international networks of roads, waterways and railways for goods and passengers. I am worried about how good the intermodal links will be. Central and Eastern Europe will give priority to expanding the road network.
- The Netherlands' geographic position will still be a dominant factor. We will continue to concentrate on our two international mainports, which will intensify their efforts to add value.
- There will be a network of multimodal corridors between the mainports and the large European economic centres (road, water, rail, pipeline and fibre optics) with advanced distribution centres on which dense networks will be based, using the latest logistical concepts right down to underground distribution in cities.
- The price mechanism will be an ever more important determinant of transport choices (such as the multimodal mix). This will improve efficiency and lighten the environmental burden.
- The use of new technology will increase, especially with a view to improving efficiency and quality; and more and more will be invested in dedicated infrastructure for the transport of goods.
- Finally, the government will increasingly develop a role as director, while market players more and more taking care for realisation and operations.

This is how I see the future. With regard to spatial planning, it corresponds to the present government's policy towards compact cities and towards the controlled development of regionally differentiated corridors.

I realise that I am sketching an ambivalent picture: on the one hand, a transport system facing terminal congestion, and on the other hand, a society that strives for high quality. The question is: what strategy do we use to solve this problem, and what role is expected of the Ministry of Transport, Public Works and Water Management?

For the long term, I see the following five strategic lines:

- Utilisation and transport efficiency

  Building new infrastructure offers no relief on the short term. But it is possible to use available technology and intelligent systems to intensify the utilisation of the infrastructure significantly. We are already successfully doing so. The same with transport efficiency. Fifty percent of trucks are still driving empty. This concept of efficiency includes the goal of utilising our infrastructure twenty-four hours a day, within the environment parameters, of course. In the long term, too, we can achieve a much more intensive and safer use of the existing infrastructure, especially through the introduction of new technological concepts.

- Making the price mechanism effective

At the same time, it is necessary to use the price mechanism to gear supply and demand better to each other. I am thinking of not only road pricing, but also parking policy as important regulatory instruments. There have to be improvements to price determination, especially by including external costs. Transport costs are hardly taken into account in goods transport. The logistical process is seen as a commercial process. My impression is that the public, too, is hardly influenced at all by transport costs when choosing a home, and this element has completely disappeared from housing policy. This means that in pricing policy we have to speak about considerable fees and taxes if it is to affect distribution.

The political decision has now been taken to give priority to road pricing and pump the revenue back, which means it will cut both ways.

- Intermodal and multimodal

We need a substantial package of measures to achieve intermodal transport:: to coordinate different modes by completing networks, by building and improving nodes and the adjacent development of node technology, by developing new transport concepts for goods and passengers, incorporating the deployment of the price mechanism.

- Improving the infrastructure

The main goal is to improve the existing networks. My investment programme would be:

1. To complete the large European road, water and rail networks, especially the road network leading to Eastern Europe; to upgrade the connection with the Netherlands by constructing high-speed or dedicated railways and improve existing railways and waterways for goods transport. For the Netherlands, it is very important that Brussels press ahead with this. The construction of the Betuwe line and the southern and eastern stretches of the high-speed line are products of this strategy.
2. The development of economic and transport corridors. The goal is to boost a trend that is already under way and to channel it in a coherent and responsible approach to corridor development. That is, to ensure that transport flows freely, that economic opportunities are exploited, and that the corridors fit in with their surroundings. I see corridors as an essential part of the logistical network; as transport axes,

infrastructure hubs (road, waterway, railway, pipeline and fibre optics between important economic centres in the Netherlands and Europe). Corridors are economic development axes. Transport-oriented companies see the corridors as great places to set up business.

3. Eliminating bottlenecks in the main national road, water and rail infrastructure. First of all, we have to consider combining it with measures for utilisation, new logistical systems and innovative solutions for target groups etc. These measures should also be considered in a area wide context. Another important goal in this context is to improve transport safety.
4. the development of new public transport systems in and around metropolitan areas, in combination with urban development

- Controlling mobility

  There will be no support in the short term for curbing mobility, and there are also no real challenges. All the same, we need back-up policy aimed at controlling mobility, and it will have an impact, certainly in the longer term. Arguments for this are:

  - Pricing policy in transport will persuade people to make more conscious choices about mobility.
  - The potential of information and communications technology and the social flexibility that arises from it mean that people will be able to find other solutions and will be more inclined to accept controlling measures (teleworking as a lubricant).
  - Mobility will have to play a greater part in choices about land-use development

So far my vision on transport development. I would now like to examine a number of the consequences these trends in mobility have for infrastructure research. The future of transport, as I see it, gives us some important topics for research on infrastructure utilisation and planning. I shall put a few of them to you now.

First, how do we utilise our infrastructure to the full? We are doing all we can with what we have at our disposal, but there are a great number of uncertainties in relation to the use of innovative technologies. The main question is how to make our infrastructure more flexible, so that we can introduce concepts such as dynamic traffic management, automatic vehicle guidance systems, dedicated facilities, and road pricing systems. The trend in the transport sector, like many others, is towards the user pays principle, and the obvious step is to discriminate according to time of day, reason for travelling and target group. What we need to know is how we can introduce measures such as these in such a close-knit network, while avoiding undesirable side-effects. And we need to know whether - and to what extent - substitution between networks is feasible.

This brings me to the issue of intermodal networks. For these to develop, high-quality nodes are an essential condition. And we need the relevant technology. This is already being developed in the container transport sector, but we now need to look at other goods flows, as well as passenger transport. We will then have to decide on where our existing networks can best be linked. And it is vital to ensure they are anchored in land-use plans. For these nodes will present excellent opportunities for the development of new activities, and thus for added value.

The spatial impact of infrastructure is the next issue that calls for our attention. We are now witnessing the development of spatial and economic corridors close to infrastructure hubs. How should we respond? What direction should the relationship take between infrastructure, economics and space shortages? How do we prevent undesirable side-effects?

This brings me to the planning and construction of new infrastructure. The main problem in the transport sector is poor flexibility. We are confronted with long-term planning, and infrastructure is long-lasting, calls for considerable investment, and has a radical impact on the environment. However, we are now witnessing a move towards greater flexibility in its use, in the form of inter-modality and multi-modalities, new logistical concepts, and transport chains. What is more, as in the telecommunications sector, utilisation is now severing its links with management. It is important for us to make infrastructure more flexible if we are to facilitate new projects and respond to user demand and the availability of new services.

Finally, what roles should government and business play in constructing and utilising infrastructure? We are already seeing market forces at work in infrastructure utilisation. There have been moves towards public-private partnerships in both construction and exploitation. But how do we form successful partnerships?

To sum up, transport infrastructure policy-makers are faced with the challenge of ensuring greater flexibility by introducing innovative forms of utilisation, linking networks, and possibly moving towards multi-functionality. Up to now, we have taken too little note of developments and solutions in other infrastructure sectors. I am therefore very interested in comparative analyses across the sectors. May I wish you every success with your research programme.

Infrastructure Concepts and Classifications

A Framework for Scenario Analysis of Infrastructures in an Economic Perspective

Lucy Firth, Karin Boersma, Bill Melody

Delft University of Technology

Faculty of Technology, Policy and Management

Economics of Infrastructures group

Abstract

Technology, economics and policy are three interrelated domains of the phenomenon infrastructure. It therefore makes sense to view them in a integrated way. However there has been a tendency to deal with only one or two of these domains at a time, without an overall frame in which the different foci of the domains can be reconciled and their interdependence understood. The paper begins with a description of the differences, commonalities and complementarities of functional concepts of infrastructure adopted in engineering and economics. Policy debates that are relevant to this function-based defined concept are discussed. After this, a model that describes the interrelationship between technological, economic and policy characteristics of infrastructure is developed. Again relevant policy debates are used as an illustration. The final section of the paper considers the value of the model as a framework for an integrated approach to the technology, economics and policy issues surrounding the infrastructure for our future society.

1. Introduction

The concept of *infrastructure* is central to the domains of engineering, economics and policy. The professions have distinct foci. Engineers focus on the technology of infrastructure while economists focus on allocative issues associated with the establishment and operation of infrastructure. Policymakers mainly pay attention to infrastructures in response to problems that arise in their – national - economies. As technology, economics and policy are three interrelated domains of one phenomenon - infrastructure - it makes sense to view them in an integrated way. Something of the inter-relatedness of the technology, economics and policy of infrastructure can be understood from a brief overview of the policy trends in terms of technology and economics. Infrastructure has received a lot of attention in the policy domain over the past 10 or 15 years. This is largely due to the on-going liberalization of the formerly government owned and/or regulated monopoly infrastructure sectors, including telecommunications, transport and energy. Liberalization can be understood as a reversal of a previous trend to bring infrastructure under government control, if not ownership. Historically, governments intervened and sought to control infrastructure for various combinations of reasons including:

- military reasons - infrastructure is important for reaching military goals. In fact, the expression 'infrastructure' was coined in 1927 in reference to military constructions such as tunnels, bridges and culverts (Oxford English Dictionary, 1989). Government intervention in establishing 'civilian' infrastructure is recognized as a self-defense strategy e.g., telecommunications infrastructure in Sweden and transport infrastructure in Switzerland.
- social reasons - infrastructure enables various social goals to be achieved e.g., education infrastructure generally and telecommunications to remote areas in Australia. Governments intervene to ensure that infrastructure and its benefits are available (and affordable) to all.
- economic reasons - infrastructure facilitates economic activity. Governments intervene in infrastructure in order to enhance economic including generating employment through the construction of public works, overcoming market failure in the establishment and operation of infrastructure, and promoting growth and development. Without intervention, it has been argued, infrastructure either wouldn't be established or if it was it would be operated either inefficiently (e.g., duplication) or exploitatively (e.g., monopoly pricing).
- political reasons - infrastructure can imbue the government with kudos arising from military and social projects, from large engineering projects or from the economic benefits of improved economic performance. Moreover, the control of infrastructure can lead to political power.

Government intervention in infrastructure shaped technology in various ways. Assured government funding, for instance, led to massive network structures that may not have been developed under private investment. The associated big system technology reinforced the economic arguments for intervention. Intervention can lead the use of technology that is not technically optimal and to the direction of technological change that is not technologically preferred. An example of this, is the Hogesnelheidslijn from Amsterdam to Belgium that is a detour for the Amsterdam to Paris route. The reversal of the trend for government intervention in infrastructure can be argued to have come about because technological change has made competition viable by changing the cost structures of establishing and operating infrastructures. Policy changes, supported by technical and economic attributes of current telecommunications, are largely responsible for the on-going policy reassessment. Therefore, relationships between the technology, economics and policy of infrastructure are complex and require an integrated multidisciplinary approach.

However, there has been a tendency to treat the domains of economics and engineering separately and so to consider economics and policy together, or engineering and policy together without an overall framework to understand all three. The result has been a lack of understanding between the professions. It could be debated whether that lack of understanding is worse if each profession assumes that all professions mean exactly the same when they say 'infrastructure', or if they assume that the others are talking about something totally different and irrelevant. The scientification of the professions has lead to different rationales that focus on restricted aspects of reality (Snellen, 1987). Snellen's paradox is that while the professions become increasingly specialized and their sciences become increasingly abstract from the complex reality, policy makers increasingly rely on scientific knowledge to inform their policies. The solution would appear to lie in interdisciplinary approach that uses a common conceptual framework to research into the complex reality. This paper develops a framework in which the different foci of the domains can be reconciled and their interdependence understood. It does so from an economist's perspective that adopts a simple representation of the engineers' focus on technology. No definition of infrastructure is provided at this stage, as the paper investigates the domain and application of the concept of infrastructure.

Section 2 of this paper looks at the differences, commonalities and complementarities of functional concepts of infrastructure that are adopted in engineering and economics. Policy debates relevant to the functional definition of infrastructure are discussed. A model of the relationship between the technology, economics and policy of infrastructure is developed in Section 3. Again, relevant policy debates are discussed. Section 4 presents a conclusion that considers the value of the framework for an integrated understanding of the technology, economics and policy issues of infrastructure for the emerging economy of the 21st century.

2. Understanding the importance of infrastructure

There is a lack of clarity about and a lack of consensus over the meaning of the term 'infrastructure'. Each profession has what Button (1996) calls a 'gut feeling' for infrastructure. While a gut feeling approach may be adequate within homogenous groups of experts, interdisciplinary study of infrastructure requires a clear understanding of infrastructure that goes beyond 'what most people consider it to be' (Button, 1996: 148). Engineers focus on the technical aspects of the relationship between infrastructure and dependent functions. Infrastructure not only implies the relationship between physical structures such as roads and the dependent functions such as transport, but also that between intangible infrastructure such as procedures and the processes that depend on these procedures. Economists, on the other hand, use various expressions to refer to that which is fundamental for economic activity. These expressions include public works, public utilities, public investment, public capital, public goods (though not all public goods are infrastructure) and social overhead capital as well as infrastructure. The economists' 'gut feeling' has recently expanded to cover a widening array of phenomena including schools, recreational facilities and the legal structure. While the engineers' and the economists' uses of 'infrastructure' seem disparate, they have in common that they refer to a phenomenon on which other functions (engineering) or activities (economics) depend. This commonality implies an understanding of the importance of infrastructure based on its function. That is, infrastructure is important to engineering and to economics because of what it does.

Providing a precise definition of infrastructure in terms of that function is not an easy matter. Some functional definitions specify the functions of particular phenomena that are recognized as infrastructure. For example, the National Resource Council (1995) defines infrastructure as: 'Facilities and their operations and the operating and management institutions that *provide* water, *remove* waste, *facilitate* movement of people and goods, and otherwise *serve* and *support* other economic and social activity or *protect* environmental quality (NRC 1995: 121, emphasis added). A more universal functional definition with an economic focus is provided by Jochimsen and Gustafsson (1977) who follow Hirschman (1958) in defining infrastructure as essential to the functioning of the economy. They operationalize the definition to that which the World Bank funds, as it only funds that which is essential to the functioning of the economy. While such operationalization is attractively simple, it begs conditioning questions.

The identification of infrastructure as a facilitator of economic activity is attractive because it can be used to highlight the importance of linkages at various levels within the economy. Power lines, for example, facilitate the transport of electricity. Electricity, in turn, facilitates various activities throughout the economy. Moreover, some activities are directly dependent on electricity, while other activities can also use other sources of power. Some economic activity may be impossible with a particular infrastructure, while others may be merely more difficult. Layers of dependent relationships and degrees of dependence can thus be identified. A list that identifies phenomena as infrastructure and indicates their hierarchical level and degree of dependence could be drawn up with cognizance of the connections between activities. This would enable infrastructure projects to be prioritized for funding to overcome bottlenecks and to facilitate essential services. Hierarchical layers of dependence are also relevant to the engineering concept of infrastructure. Those layers may consist of physical structures or protocols. The Open Systems Interconnection (OSI) agreement by which heterogeneous computers are standardized to enable communication), for example, establishes a hierarchy of functional layers. These layers consist of protocols. The functioning of each layer depends on the ones below (PC webopaedia, 1998). Thus, in engineering and in economics, infrastructure is not an absolute concept. Rather, it is conditional on its relationship to a certain functionality of economic activity. That relationship is that the function or activity depends on the infrastructure. This commonality in the functional concept of infrastructure provides a basis for successful communication between the professions.

2.1 Relevant policy debates

The understanding of infrastructure as a phenomenon that facilitates economic activity is associated with at least two important policy debates in economics: those of economic growth and development, and of competitiveness. The essence of the relevance of infrastructure to these debates is discussed here.

Economic growth and economic development

The policy of establishing infrastructure to exploit the intuitive, causal link between the existence of infrastructure and economic performance, including productivity, growth and development was ratified by Rostow (1960). Rostow found that large infrastructures, such as the railroads in the USA, led to accelerated economic growth. Developing countries were advised to increase public capital expenditure particularly on transport (Goldin and Winters, 1995). Empirical work at the national level (e.g., Aschauer, 1989) and at the regional level (e.g., Florax, 1992) confirmed that infrastructure promotes growth. Moreover, infrastructure was argued to contribute not only to economic growth by increasing productivity, but also to social and economic development by providing amenities that enhance the quality of life. Kessides (1996) found that infrastructures' contribution to economic development was conditional upon its efficient operation and its relevance to the needs and wishes of the society. However, the causal link between infrastructure and economic growth has been questioned. Fogel (1964) used historic data to show that the railroads actually contributed little to the USA's growth. Gramlich (1994), in a review of the debate over the impact of infrastructure on growth, argued that econometric estimates do not provide convincing, consistent evidence of a relationship. Moreover, studies using differencing methods have found zero impact of public investment in infrastructure on productivity (e.g., Hulten and Schwab 1991; and Tatom 1993).

Despite the lack of consistent supporting econometric evidence, there remains a gut feeling that infrastructure, as functionally defined, is important to growth and development. As Batten states in the introduction to an edited volume on infrastructure and economic development: 'a disturbingly shallow degree of consensus can be gleaned from the contents of this volume. The uncontested part of that consensus says that a durable and efficient system of infrastructure *seems* to be a good thing for an economy' (1996: 11, emphasis in original).

Competitiveness

The definition of infrastructure as facilitating or enabling economic activity suggests there may be a relationship between the presence of infrastructure and competitiveness at both the national and company level. Nations compete to attract private investment by companies for such benefits as employment, technology transfer and export earnings (Mjoset, 1992). To attract investment, countries provide tax relief, regulatory support and infrastructure services. The infrastructure sought by companies and provided by nations has recently focused on communication, transport and storage infrastructures that orchestrate purchase, production and marketing functions. However, the effectiveness of infrastructure for attracting foreign investment is not clear. As Kessides states, 'although the provision of infrastructure is clearly not a sufficient condition for attracting private investment to a given location, differences in the quality of infrastructure can be an important factor at the margin in determining the choices among potential sites' (1996: 218).

Not all infrastructures that enhance competitiveness are tangible structures. Eliasson argues that competitiveness is enhanced by Dahmenian competence blocks of cutting-edge companies. The existence of these companies in a location promotes the generation and diffusion of the knowledge. As such they act as 'technical universities and research institutes, unintentionally providing free educational and research services, often in areas where such services are not supplied by existing educational institutions or where the nature of competence makes traditional educational institutions incapable of supplying them' (1996: 125). This raises the question of whether these blocks and the implied personal networks are infrastructure.

The complexity of the link between infrastructure and competitiveness is indicated by Porter (1990). Porter argues that innovation at the company level is the essence of the international competitiveness that brings dynamic prosperity. The propensity of a company to make innovative decisions, which will enhance its competitiveness, depends on the environment in which it operates. Infrastructure as a key element in that environment is important to competitiveness to the extent that it influences innovative decision-making. While infrastructure can enhance economic performance, relying on cost advantages due to infrastructure can retard innovation. A lack of infrastructure can, for example, create bottlenecks that lead to innovation and competitiveness. 'Innovation to offset selected weaknesses is more likely than innovation to exploit strengths' (1990: 83). The idea that infrastructure enhances international competitiveness by reducing operation costs is therefore misleadingly simple. As Reich says: 'it is not which nations own what, but which nations citizens learn to do what' (1992: 137).

Summary

The major points from this section are:

- there is considerable commonality between the engineering concept of infrastructure as performing the fundamental role of a platform for functionality and the economic concept of infrastructure as performing the role of facilitating economic activity.
- the relationship between infrastructure and engineering functions or economic activity is one of dependence. That dependence may be absolute or merely imply the facilitation of the function or activity. There are layers of functions or activities each depending on the ones below.
- the role of infrastructure suggests important policy implications for economic growth and development, and for competitiveness.
- scientific evidence does not consistently confirm the existence of a causal relationship between infrastructure and economic growth. Therefore, the policy of seeking to enhance economic performance through the establishment of infrastructure is not unanimously supported.
- the effectiveness of infrastructure to perform the function of facilitating economic activity and bringing the associate benefits depends on its suitability to the broader socio-political context in which it exists.

3. Integrated model of infrastructure

The functional approach to understanding infrastructure invites the classification of all important facilitators of economic activity as infrastructure. However, to economists not all facilitators of economic activity are infrastructure. If, for example, economic activity depends on resources such as land, labor and capital, the question arises: are these inputs infrastructure? If they are not, the simple description of 'facilitating economic activity' may lead to their erroneous classification as infrastructure. The classification of a facilitator as infrastructure in economics is based on that phenomenon having certain characteristics associated with the markets' failure to establish and operate infrastructure. Those characteristics result, in large part, from the technological characteristics of the infrastructure, and they provide the basis for the economic argument for policy intervention.

Various authors have sought to distinguish infrastructure from other phenomenon that perform the function of facilitating economic activity by identifying the characteristics of sectors or industries that are infrastructure (e.g., Stohler, 1977; Biehl, 1986 and Kessides, 1993). While not all of these writers explicitly set out to define infrastructure in terms of a set of characteristics, the identification of defining characteristics of sectors that are considered infrastructure, is central to their work. A characteristic-based approach to distinguishing infrastructure would ideally enable a list to be drawn up of all that is infrastructure. However, the selection of the sectors that are studied in the literature in order to identify these characteristics seems to originate from the 'gut feeling' of what sectors are important to the economy. Stohler (1977), for example, states that something can be said about the essence of infrastructure by adding up characteristics that are *generally associated with infrastructure*. Moreover, a degree of arbitrariness is imposed by each author. Stohler, for example, explicitly aims to link the classification of sectors as infrastructure to their having certain characteristics, but there is no apparent relationship between having those characteristics and being classified as infrastructure in his work. Stohler finds that transport, energy, education, research, health, waterworks and national defense are infrastructure sectors. That he does not consider the (tele)communication sector, may be indicative of the changing level of interest in specific sectors over time, rather than indication that he considers telecommunications not to be infrastructure. Biehl (1986) generally follows Stohler but does not mention national defense. Instead, he lists communication, social infrastructure and sport, tourist and cultural facilities as infrastructure. Kessides (1993) has a more limited scope to what is infrastructure that can be summed up in five terms: transport, electricity, water, waste and telecommunications. The findings of these authors on the characteristics of infrastructures are summarized in Table 1. Table 1 illustrates the diversity of characteristics that are looked at in economic literature and also interprets the characteristics mentioned by Stohler, Biehl and Kessides as relating to the practical domains of the establishment of infrastructure, the production of infrastructure services, the consumption of infrastructure services, and network coordination.

***Table 1**: Characteristics of infrastructure as identified by various authors.*

	Characteristics		**Practical Domain**	**Sectors identified as infrastructure**
	Technical	**Economic**		
Stohler	Long life time, big projects	Large sunk costs, large investment, high risk	Establishment of infrastructure	Transport, energy, education, research, health, waterworks, national defense
	Interdependence within sector	Economies of scale, large indirect costs, unprofitable exploitation, external effects, high risk, heterogeneous (incomparable) performance	Production of infrastructure good or service	
	Product is input for many sectors,	Lack of vision of individual consumers, non-excludible, no market price	Consumption of infrastructure good or service	
	Structural interdependence		Network coordination	
Biehl	Immobility		Establishment of infrastructure	Transport, energy, education, research, health, waterworks, communication, social infrastructure, sportive, tourist and cultural facilities
	Indivisible, restricted substitutability		Production of infrastructure good or service	
	Product is input for many sectors, non excludible	Non excludible	Consumption of infrastructure good or service	
	Band or point in network, system effects		Network coordination	
Kessides	Big projects	Contestable markets, large sunk costs	Establishment of infrastructure	Transport, electricity, water, waste, telecom
		Natural monopoly (economies of scale and scope), contiguity	Production of infrastructure good or service	
	Non-excludable, non-rival	Substitutes, low price elasticity, temporal patterns of demand, diversity of user needs, non excludible	Consumption of infrastructure good or service	
	Interlocking networks	Network externalities	Network coordination	

Source: Stohler, 1977; Biehl, 1986; Kessides, 1993.

Table 1 indicates that although the literature does not present an unambiguous and generally-agreed upon list of characteristics, several characteristics can be interpreted as central to the economic concept of infrastructure. Moreover, those characteristics apply to the practical domains, as discussed here.

- *Characteristics associated with the practical domain of investment in the establishment or construction of infrastructure.* The technical feature of big systems generally involves large capital investment with a delay in returns during construction and installation. Given the historic lack of venture capital this implies market failure in that the infrastructure is not established. The durability of infrastructure implies that, once online, returns continue for the long term. This implies that a perspective longer than the short-term, profit-maximizing perspective of business is needed to fund the investment. There is a high degree of sunkness (or specificity) of investment. This means that infrastructure, if abandoned, cannot be sold for other uses, as could, say, a hammer. This specificity increases the risk of investment as it reduces the options for a revenue flow. Therefore, where the technology of infrastructure calls for large up front investment in assets that are peculiar to the infrastructure, there is a tendency for the market to under-provide infrastructure in the absence of intervention.
- *Characteristics associated with the production of infrastructure services.* The *production* of infrastructure services is subject to increasing returns to scale. This implies a natural tendency for monopolization because large-scale firms can exploit cost advantages not available to smaller firms. An obvious example is that once a railway is built to serve one passenger, it can be used to carry other passengers and goods for trivial extra cost. The monopolization is reinforced because the existing operator has the advantage over potential entrants to the industry.
- *Characteristics associated with the consumption of infrastructure services.* The *consumption* of infrastructure services is subject to some degree of non-rivalness and non-excludability. *Non-rivalness* is due to the technical nature of the service, and means that consumption by one party does not impair consumption by another party. Within bounds, the service is not depleted by use and therefore potential users are not rivals. An extreme example of this is the security against flooding provided by the Netherlands' systems of dykes and canals. The security afforded one citizen by these measures is not reduced by that offered to another. This is also an example of a *non-excludable* infrastructure service. A citizen cannot be excluded from security, for example, if they do not to contribute to the cost of that security. Where infrastructure has the technical attributes of non-rivalness and is at least partially non-excludable, there is a tendency for individuals to avoid contributing and so there is under investment in that infrastructure. Where this 'free-riding' is a problem the government has the power to coerce contribution through taxation, and then fund infrastructure services under

private or public governance. Therefore, the technology of infrastructure can lead to an inability for the market to operate competitively and viably.

- *Characteristics associated with the coordination of network infrastructures.* Infrastructures consist of nodes (e.g., stations, electricity production plants, and telephone switches) and connecting links (e.g., railway, high voltage electricity cables, and telephone lines) that together make up a network structure. The functioning of (infrastructure) networks is characterized by structural interdependence within the system. That is, the components in the system are complementary to each other and have to be combined to create a service. This interdependence results in a need for a degree of central co-ordination. The technical network attribute of infrastructure is therefore a further reason for Government intervention to perform (or regulate) the role of coordinator.

The essential function of infrastructure and its technical and economic characteristics together form the reason that infrastructure has been singled out and subjected to intervention that has not been imposed on other facilitators of economic activity. The past, current and ongoing developments in infrastructure policy can be understood within the tripartite relationship between the technology, economics and policy of infrastructure, as depicted in Figure 1.

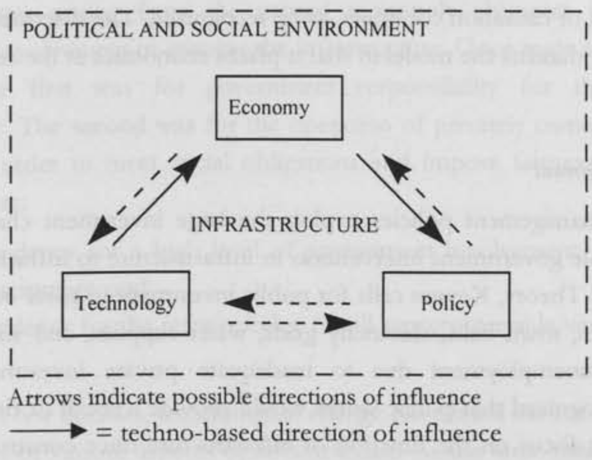


Figure 1. *Model of infrastructure's tripartite relationship: technology, economics and policy*

Figure 1 depicts the three interdependent domains of infrastructure within a broader setting of the political and social environment. The arrows indicate that each member of the tripartite relationship is reciprocally influenced by each other member. The solid lines depict the *techno-based* direction of influence. By 'techno-based' it is meant that, without influences from the political and social environment, the technology of infrastructure determines the economic characteristics that indicate market conditions and policy. Historically, the technology of railways, for example, required major construction work that implied enormous investments, which were beyond the private market resources in some countries. The policy to intervene to ensure construction of railways can be seen in terms of the economic characteristic that was technologically-based. In the real world there are other factors, including feed back, which make the other direction of influence possible. These are indicated by the dotted lines. The policy in response to the technologically-based economic characteristics may influence (or merely reinforce) the technology in an ongoing dynamic process. The policy of funding railway establishment protected railway technology from incentives to developer less massive technology. The technology of infrastructure may also be driven by policy rather than merely respond to it.

3.1 Relevant policy debates

The relevance of the characteristics of infrastructure to policy debates is indicated by the following discussion of selected policies. The discussion emphasizes the importance of technology by focusing on the techno-based direction of causation. This should not be taken to imply technological determinism. The importance of the factors associated with the reversed direction of causation continues to be recognized. The discussion is from an economic perspective that fits the model in that it places economics at the center between technology and policy.

Keynesian demand management

Keynesian demand management policies exploit the large investment characteristic of infrastructure to enable government intervention in infrastructure to influence economic cycles. In the General Theory, Keynes calls for public investment in *public works* including building roads, houses, town halls, electricity grids, water supplies, and so forth (1936: 106) to overcome unemployment due to inadequate private investment demand. Although Keynes recognized that public works would provide a social or financial return over time, he did not focus on the function of infrastructure once constructed. Rather, the focus was on boosting economic activity to produce employment beyond the 'primary employment provided by the public works' (1936: 117) by public capital investment operating through the multiplier (Brown-Collier and Collier, 1995).

Technological change leads to a change in the type of infrastructure suited to Keynesian demand management policy. The introduction of radio-based mobile telephony, for example, removes the telecommunications sector from the specter of massive construction works. On the other hand, the introduction of cable TV and associated roll out of cable placed the television industry in the group that requires massive investment. Generally, the reduction in attention to policies of anti-cyclical infrastructure construction in developed countries can be seen as much as a movement away from such massive projects to connect locations and to control the environment as it is an ideological movement away from economic planning.

Liberalization and privatization

The recent and on-going debates on the liberalization, privatization and deregulation of infrastructure are centered on various characteristics that impact on the market's capacity to co-ordinate the establishment and operation of infrastructure (Kessides, 1993). The debate is not about governments abdicating responsibility for a need that they previously shouldered, but about identifying the limits of that need, and intervening appropriately. An historic lack of private venture capital for investment in big system infrastructure led to popular agreement that there was a role for government in the *establishment* of infrastructure in the public interest (Kessides, 1993). There were two general options. One option was for government responsibility for planning and establishing infrastructure, and for government ownership of that infrastructure. The other option was for the government to assist private capital investment, for example, by granting monopoly licenses as an incentive, or by arranging development funds. Concern over the anti-trust issues arising from the natural monopoly character indicated a need for government involvement in *operating* the infrastructure. Once again there were two general options. The first was for government responsibility for the operation of the infrastructure. The second was for the operation of privately owned infrastructure to be regulated in order to meet social obligations and impose fairness and efficiency. Two outcomes were:

- a tendency for a high level of government involvement in all aspects of infrastructure, and
- a tendency for the responsibility for all aspects to reside with one party.

Over the past few decades, technological change has altered the nature and importance of characteristics of various aspects of infrastructure. At the same time social and economic development and a change in the understanding of the functioning of the economy have led to a reconsideration of the appropriate level of intervention. Recent economic arguments have indicated that government involvement compounds the problems of market failure with those of government failure. The solution has been seen to be the unbundling of various aspects of infrastructure ('sectoral planning and policy-making, ownership, regulation, financing, execution of investment, and/or operation and maintenance' (Kessides, 1993:*x*)) with intervention kept to a minimum and restricted to those aspects where it is essential. The optimal mix is argued to be incentive-based efficiency through privatization and liberalization, supported and conditioned by government involvement in planning (Melody, 1997; Kessides, 1993). The reduction in the welfare of those negatively affected by liberalization should be addressed, if necessary, by other means open to the government (Kessides, 1993).

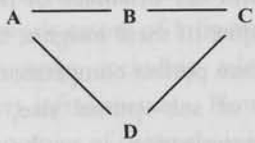
The reassessment of the Government role coincides with technological change that reduces the emphasis on big systems in some infrastructures (e.g., telecommunications and energy). This has reduced the need for investment funds, shortened the period before revenue is earned on that investment and reduced the natural monopolization tendency. It may well be asked why technological change led to liberalization during the 1980s and 1990s when it had led to regulation in earlier times (Winseck, 1998). That question cannot be addressed fully within the tripartite relationship model developed in this paper. Rather, it makes sense within the broader social and political environment in which the model is embedded.

What can be understood within the model is that technological change has altered the extent to which goods and services are rival and excludable and so altered the level of expected market failure. The growth of the private capital market has reduced the importance of capital resources as a limiting factor for investment in infrastructure. There is thus an increasing capacity for private enterprise to undertake many of the aspects of infrastructure establishment and service provision, and a continuing tendency for that enterprise to be regulated in the public interest. The outcome of the debate on privatization, liberalization and regulation of infrastructure is a recognition 'that there are fewer activities requiring public intervention than once was believed; and that public intervention, when justified can be exerted through less distorting policy instruments than those traditionally used' (Kessides, 1993: ix).

Economics of networks

The relevance of the technical network nature of infrastructure for economics and for policy is less well understood than are those of the establishment and operation of infrastructure. Although networks are present in every sphere of human activity and are important to technical, social and economic development (Batten, Casti and Thord, 1995), conventional economic theory does not deal with their particularities and thus is inadequate for the their analysis (Economides and Encaoua, 1996). Conventional economics deals with relationships between components as either substitutes or complements. In networks, nodes and links can be both complements and substitutes at the same time. For example, two railways between the towns of A and B, and between B and C can be complements in producing transport services between A and C, see Figure

2. At the same time they can be part of two alternative substitutable options in the transport service from D to B.



$AC = AB + BC$

$DB = (DA + AB)$ *or* $(DC + CB)$

Figure 2. *Components as both complements and substitutes*

Further, networks are subject to an external effect whereby the value of the network is enhanced by its size (i.e., the number of nodes and links). In telecommunications, for example, an additional subscriber increases the potential number of other subscribers that any subscriber can reach. Providing a complementary link to the existing network increases the value of that network to each subscriber. In other words, the value of a unit of a network good increases with the expected number of units sold (Economides, 1996), which is contrary to conventional economic reasoning. Finally, the traditional treatment of networks in economics focuses on efficiency issues relevant to a single owner of the entire network. As a result of the liberalization of network markets, the focus is increasingly on issues of interconnectability and compatibility between firms. 'In a network where complementary as well as substitute links are owned by different firms, the questions of interconnection, compatibility, interoperability, and co-ordination of quality of services become of paramount importance' (Economides, 1996: 678).

Much of the literature on the economics of networks has focused on game-theoretical solutions to problems attributed to peculiarities in the investment incentives (e.g., Doyle and Maher, 1993) and in market conditions (e.g., Hendricks, Piccione and Tan, 1995). Networks exhibit a positive critical mass, that is, they require a minimum size in order to be profitable. This results in a heavy minimum threshold investment (Trebing, 1994), which illustrates the importance for the network nature of infrastructure for the establishment and operation of infrastructure. Therefore, the implications of the network nature of infrastructure are important to the understanding of its establishment and operations as well as its co-ordination. Optimum market conditions are also affected by the existence of network externalities. Economides (1996), by modeling strategic decision making in variously-specified games, finds that 'perfect competition will provide a smaller network than is socially optimal' (1996: 682). Monopolists might also fail to behave as predicted by conventional economics. 'In the presence of strong network externalities, a monopolist exclusive holder of a technology may have an incentive to invite competitors and even subsidize them' (Economides, 1996: 691). Competitors in the telecommunications equipment industry may seed each other's knowledge in order to have a united voice in the standards forums. The conditions for these have not been adequately researched, especially in connection with the dynamics of path dependency (Boisot, 1995; Langlois and Robertson, 1996). In spite of these insights, the conventional finding that monopoly will produce worse results than perfect competition may still apply. While perfect competition generates a network of sub-optimal size, monopoly may exacerbate this by overcharging. The lack of determination as to the optimal market structure is compounded by decision making under oligopoly in which the precise specification of the game influences the outcome (Katz and Shapiro, 1985).

The issues raised by the economics of networks indicate that because the network characteristic of infrastructures impacts on the other practical domains it has relevance to policies for the establishment and operation of infrastructure as well as for policies for the coordination of infrastructure. Two examples are that the prescription of unbundling becomes less generally applicable, and proscription of merges becomes less universal (Economides, 1996). The economics of networks, while highlighting the inadequacies of standard economic thinking for dealing with network infrastructures, is indeterminate in suggesting policy options.

Summary

The major points from this section are:

- the function of infrastructure is not adequate to explain the level of intervention that it has attracted. That intervention can be understood within the context of the technical characteristics and associated economic characteristics of infrastructure.
- no defining economic characteristics can be found in literature. Therefore, it is not possible to separate infrastructure from other facilities, goods or services.
- four categories of characteristics of infrastructure are important to the practical domains of the establishment, production and consumption of infrastructure services and the coordination of infrastructure networks: those due to the big system nature of infrastructure, those due to pricing problems in the production and consumption of infrastructure goods and services, those due to the network nature of infrastructure.
- there is a tripartite relationship between technology, economics and policy of infrastructures. The relationship is one of reciprocal determination. The direction of causation between technology and policy *via* economic characteristics can be understood within this model.

- the socio-political context in which infrastructure exists and operates influences the direction of causation between technology and policy. This is illustrated by a short discussion of Keynesian demand management and recent liberalization and privatization policies.
- the network nature of infrastructure is central to future policy for network infrastructures including telecommunications, information, energy and transport infrastructures. However, the current level of understanding of the economics of networks is inadequate to inform policy.

4. Conclusion: Towards the future

Although infrastructure is central to the successful operation of our society and economy, it is poorly understood and subject to many features that render it difficult to plan for and to manage. These problems are exacerbated by the fact that technological, economic and policy aspects of infrastructure are dealt with separately, by different professions. Better policy outcomes may be achieved by interdisciplinary cooperation based on a better understanding of the meaning and application of the concept of infrastructure in each profession. Section 2 of this paper argued that there is sufficient commonality to work towards a common understanding. A further exploration of the concept of infrastructure may lead to a framework for interdisciplinary research on infrastructure. The review of economic insights that was given in this paper can be used as a basis for this. Integrating concepts from policy, legal and engineering disciplines will strengthen the framework. Furthermore, such a theory-based concept can be supplemented by using empirical data. The goal of this effort would be an operational concept of infrastructure, that is both multidisciplinary and generally applicable. This operational concept should be applicable to all sectors of the economy and should enable both scientists and policy makers to determine whether or not something is infrastructure in a clear and consistent manner.

The call for a better understanding of infrastructure and the associated economic and policy implications applies to all kinds of networks. Economists must bring to the interdisciplinary projects on infrastructure a more highly developed economics of networks than the discipline presently has. The economics of networks must be suited to inform policy not only on the coordination of network infrastructures, but also on the establishment and operation of network infrastructures. Those networks are heterogeneous, and not all built or tangible. Along with transport, energy and telecommunications networks there are also, for example, exchange networks (c.f., Hakansson, 1990; Granovetter, 1973) and learning networks (Rogers, 1982; Kobayashi, 1995). These less tangible networks are essential to economic activity. Networks between people and firms, for instance, have long been recognized as central to exchange and learning relationships in industry. Such networks are increasingly important in the emerging knowledge economy in which knowledge is the most important input and output and learning is the most important activity. The nature of learning networks may be quite different to the physical or electronic networks that dominate our thinking about infrastructure today, and so present renewed challenges for economics and policy makers.

There is a clear need for research programs that focus on the interplay of the technical, economic and policy issues of infrastructure. Work on these areas has begun and must continue. Ideally, such programs would provide a comparative study of technological, economic and administrative development in infrastructure and their public management, in order to identify the generic requirements for the design, operation and management of infrastructure facilities, and the decision making processes linked to them (c.f., Weijnen, 1996).

References

Aschauer, D. (1989), Is Public Expenditure Productive? *Journal of Monetary Economics* 23/2, pp. 177-200.

Batten, D. (1996), Infrastructure and the Complexity of Economic Development: An Exploratory Introduction, in: D. Batten & C. Karlsson (eds.), *Infrastructure and the Complexity of Economic Development*, Berlin: Springer, pp. 1-13.

Batten, D., J. Casti & R. Thord (1995) Introduction, in: D. Batten, J. Casti & R. Thord (eds.), *Networks in Action: Communication Economics and Human Knowledge*, Berlin: Springer-Verlag, pp. 195-228.

Batten, D. & C. Karlsson (1996), Preface, in: D. Batten & C. Karlsson (eds.), *Infrastructure and the Complexity of Economic Development*, Berlin: Springer, p. *v*.

Biehl, D. (1986), *The Contribution of Infrastructure to Regional Development*, Luxembourg: Office for Official Publications of the European Communities.

Brown-Collier, E. & B. Collier (1995), What Keynes Really Said About Deficit Spending. *Journal of Post Keynesian Economics* 17/3, pp. 341-355.

Button, K. (1996), Ownership, Investment and Pricing of Transport and Communications Infrastructure, in: D. Batten and C. Karlsson (eds.), *Infrastructure and the Complexity of Economic Development*, Berlin: Springer, pp. 147-165.

Boisot, M. (1995), Is Your Firm a Creative Destroyer? Competitive Learning and Knowledge Flows in the Technological Strategies of Firms. *Research Policy* 24, pp. 489-506.

Doyle, C. & M. Maher (1993), *Investment Decision: The Role of Market Structures in Vertically Separated Network Industries*, Department of Economics Working Paper 9403. Cambridge: University of Cambridge.

Economides, N. (1996), The Economics of Networks. *International Journal of Industrial Organization* 14, pp. 673-699.

Economides, N. & D. Encaoua (1996), Special Issue on Network Economics: Business Conduct and Market Structure. *International Journal of Industrial Organization* 14, pp. 669-671.

Eliasson, G. (1996), Spillovers, Integrated Production and the Theory of the Firm. *Journal of Evolutionary Economics* 6, pp. 125-140.

Florax, R. (1992), *The University: A Regional Booster? Economic Impacts of Academic Knowledge Infrastructure.* Avebury: Aldershot.

Fogel, R. (1964), *Railroads and American Economic Growth: Essays in Econometric History.* Baltimore: John Hopkins University Press.

Goldin, I. & L. Winters (1995), *The Economics of Sustainable Development.* Cambridge: Cambridge University Press.

Gramlich, E. (1994), Infrastructure Investment: A Review Essay. *Journal of Economic Literature* XXXII, pp. 1176-1196.

Granovetter, M. (1973), The Strength of Weak Ties. *American Journal of Sociology* 78/6, pp. 1360-1380.

Hakansson, H. (1993), Networks as a Mechanism to Develop Resources, in: P. Beije, J. Groenewegen & O. Nuys (eds.), *Networking in Dutch Industries*, Garant: Leven and Apeldorn.

Hendricks, K., M. Piccione, & G. Tan (1995), *Entry and Exit in Hub-Spoke Networks*, Department of Economics Discussion Paper 95/20, Vancouver: University of British Columbia.

Hirschman, O. (1958), *The Strategy of Economic Development*, New Haven: Yale University Press.

Hulten, C., & R. Schwab (1991), *Is There Too Little Public Capital? Infrastructure and Economic Growth*, Discussion Paper. Washington: American Enterprise Institute.

Jochimsen, R. & K. Gustafsson (1977), Infrastruktur. Grundlage der marktwirtschaftlichen Entwicklung, in: U. Simonis, (ed.), *Infrastruktur. Theorie and Politik.* Koln: Kiepenhauer & Witsch, pp. 38-53.

Katz, M. & C. Shapiro (1985), Network Externalities, Competition and Compatibility, *American Economic Review* 75/3, pp. 424-440.

Kessides, C. (1996), A Review of Infrastructure's Impact on Economic Development, in: D. Batten & C. Karlsson (eds.), *Infrastructure and the Complexity of Economic Development*, Berlin: Springer, pp. 213-230.

Kessides, C. (1993), *Institutional Options for the Provision of Infrastructure*, World Bank Discussion Paper. Washington: The World Bank.

Keynes, J. (1936), *The General Theory of Employment Interest and Money*, London: MacMillan and Company.

Kobayashi, K. (1995), Knowledge Networks and Market Structure: An Analytical Perspective, in: D. Batten, J. Casti & R. Thord (eds.*), Networks in Action: Communication Economics and Human Knowledge*, Berlin: Springer-Verlag, pp. 195-228.

Langlois, R. & P. Robertson (1996), Stop Crying over Spilt Knowledge: A Critical Look at the Theory of Spillovers, paper presented at the *MERIT Conference*, August 25-27, Maastricht.

Melody, W. (1997), *Telecom Reform: Principles, Policies and Regulatory Practices.* Lyngby: Den Private Ingeniorfond, Technical University of Denmark.

Mjoset, L. (1992), *The Irish Economy in a Comparative Institutional Perspective.* Dublin: National Economics and Social Council.

National Research Council (1995), *Measuring and Improving Infrastructure Performance*, Washington D.C: National Academy Press.

PC webopaedia (1998), http://www.pcwebopaedia.com/OSI.htm, April.

Porter, M. (1990), *The Competitive Advantage of Nations.* London: MacMillan.

Reich, R. (1992), *The Work of Nations.* New York: Vintage Books.

Rogers, E. (1982), *Diffusion of Innovations* (3rd ed.). New York: The Free Press.

Rostow, W. (1960), *The Stages of Economic Growth.* Cambridge: Cambridge University Press.

Trebing, H. (1994), The Networks as Infrastructure – The Reestablishment of Market Power. *Journal of Economic Issues* XXVII/2, pp. 379-389.

Serageldin, I. (1995), *Nurturing Development: Aid and Cooperation in Today's Changing World.* Washington: The World Bank.

Snellen, I. (1987), *Boeiend en Geboeid. Ambivalenties en Ambities in de Bestuurskunde.* Alphen aan den Rijn: Samson H.D. Tjeenk Willink.

Stohler, J. (1977), Zur rationalen Planung der Infrastruktur, in: U. Simonis, (ed.), *Infrastruktur. Theorie and Politik.* Koln: Kiepenhauer & Witsch, pp. 16-37.

Tatom, J. (1993), Paved with Good Intentions: the Mythical National Infrastructure Crisis. *Policy Analysis*, Aug 12.

Weijnen, M. (1996), *Design and Management of Infrastructures.* Delft: Interfaculty Research Programme of the Delft University of Technology.

Winseck, D. (1998), *ReConvergence: A Political Economy of Telecommunications in Canada.* Cresskill: Hampton Press.

Competition Engineering in Network Based Industries

Paradigms, Practices and Pendulums

Mark J.W. van Twist, Ernst F. ten Heuvelhof

Delft University of Technology

Faculty of Technology, Policy and Management

Public Management group

Abstract

In a short span of time, the idea that deregulation together with privatization and liberalization will lead to a more efficient management in network based industries has been made a principle of policy almost all over the world, although the speed of action and radicality between countries may differ a lot. Competition engineering is thought to be the key to improvement in performance. The expectation is that consumers will eventually profit from this by lower prices, increased customer-orientation and more product innovations.

In this paper we describe the shift in paradigms that forms the basis of the developments in the different network based industries and the new practices incited by this paradigm-shift. Furthermore, we put the developments in the network based industries in a historical perspective. Is this then the predicted end of history? Are all the changes we can still expect in the future just the aftermath of the privatization, liberalization and deregulation that are taking place at present? Or should we view these changes as just another movement of the pendulum, that will inevitably incite a counter-movement at any point in time?

1. Introduction

1.1 Network Based Industries

A number of industries in our society can exist only because there are networks of transport and distribution that ensure the possibility of production and service (Armstrong and Doyle, 1995, p.2). Examples are the telecommunication and energy branches, as well as public transport and drinking water supply.

Because of the crucial importance of transport and distribution networks, these industries are commonly called infrastructure industries or network based industries. In the telecommunication industry the infrastructure exists of a mix of transmission media and switchgear. Electricity supply makes use of a network built of a hierarchy of increasingly higher voltage transmission links. For water treatment, use is made of delivery and storage facilities for water and sewerage networks. And the railway network for example, can be unfolded into tracks, signalling-systems, earthworks (tunnels, bridges et cetera) and of course the stations.

1.2 Importance of Network Based Industries

Network based industries supply us with electricity, fuel and water, facilitate data exchange for telephony and broadcasting, transport people and distribute goods. Herewith, network based industries are essential to the economy. The products and services these industries supply directly influence a society's affluence and an economy's productivity and its international competitive position. We are dependent on these network based industries to such an extent that the slow down or (temporary) halt of just one of them has disastrous consequences to modern society. The economic damage of one day without electricity or telephone can hardly be calculated and without public transport maybe not all comes to a hold, but still the consequences are grave.

1.3 Developments in Network Based Industries

Probably, the large importance of network based industries to economy and society are a major explanation for the fact that until recently most of these industries were government owned (Europe: public monopoly) or at least were under close government supervision (United States: regulated private monopoly)

Nevertheless, over the last ten years this has been changing rapidly. All over the world, countries are studying the possibilities of and experimenting with deregulation in these industries; in combination with privatization of public companies and a liberalization of the markets for infrastructural products and services (Kessides, 1993; Klein and Gray, 1997). In a short span of time, the idea that deregulation together with privatization and liberalization will lead to a more efficient management in network based industries has been made a principle of policy almost all over the world, although the speed of action and radicality between countries may differ a lot.

According to this idea, competition engineering is almost always thought to be the key to improvement in performance. Therewith the expectation is that consumers will eventually profit from this by lower prices, increased customer-orientation and more product innovations. The (expected) positive effects of privatization, liberalization and deregulation are stressed constantly in political debates. However it really is too early yet to be able to form a balanced opinion on the effects that the developments in network based industries induce. As yet there is no serious empirical proof of the advantages as promised by the new developments.

1.4 Paradigms, practices and pendulums

It is not the ambition of this paper to offer a detailed survey of the effects that are now slowly becoming apparent in the different network based industries. There is no room for such an elaboration here and moreover we can refer you to other publications (Van Twist and Veeneman, 1999). Instead we would like to elaborate on a number of reflections on the effects of competition engineering in network based industries that are more general and beyond specific industries.

Thereto, we will describe the shift in paradigms that forms the basis for the developments in the different network based industries and the new practices incited by this paradigm-shift. Furthermore, we would like to put the developments in the network based - industries in a historical perspective. Is this then the predicted end of history? Are all the changes we can still expect in future just the consequence of the developments that take place now: privatization, liberalization and deregulation? Or should we view these changes as just another movement of the pendulum, that will inevitably incite a counter-movement at any point in time?

2. Shifting paradigms

In our view, a paradigm shift in the ideas on network based industries can be discerned. We speak of a paradigm shift, because this concerns not only an ideological but also an intellectual shift. What is interesting about the paradigms in the ideas on network based industries is not only that they contain intellectual as well as ideological dimensions, but also that they describe a situation, make clear what is wrong about that situation and, at the same time, open new perspectives on how things can and should be done differently and better. In this respect Aucion (1990, p.116) tells us: "Ideas which gain ascendancy in political circles are best described as "paradigms" for the simple reason that they combine both intellectual and ideological dimensions. They are models, which have appeal, because they appear to describe reality, they offer an explanation for the same, and they prescribe ways to change in desired directions. In these respects, they have adherents beyond the explicitly partisan circles of politics. At the same time, they are represented in terms that are simple and easy to digest. As a result, they are ideas which can be communicated to politicians and bureaucrats in forms which do not presuppose that either politicians or bureaucrats will have to read the original sources of the ideas".

2.1 The "classic" paradigm

The main assumption that forms the basis of the 'classic' paradigm is that network based industries share specific, mainly economic, features that make exclusion of competition desirable and even necessary.

Among others, these features are as follows (Bauer, 1998, pp. 3-6):

- the high capital intensity and high capital outlay that are needed to construct and manage the infrastructure without which the industries could not be;
- the economic position of power the companies in network based industries have acquired in relation to consumers, because the market is like a natural monopoly;
- the vital importance of the products and services that these industries supply for other industries and private households; indispensable for everyone and mostly without substitutes being available;
- the strategic importance of the network based industries for economic development and national security.

It is recognition of these features that ensured that products and services of the infrastructure-based industries were supplied by highly regulated private monopolies in the United States and by public monopolies in Europe until recently. The system of rules that was formed over the last decades is based on the paradigm that network based industries should not be exposed to competition on the market and that government intervention is essential to protect consumers against possible abuse of the power of monopoly and against possible cut-throat competition. (Bauer, 1998, p.36).

The specific economic characteristics of the network based industries entail high overhead costs, that can be passed on to the consumers in very differing manners. When government regulation/ government property is renounced, this could all too easily lead to abuse of power on prices, packages of services and investments. Also destructive forms of competition can arise without government regulation.

From a historical point of view, it can be concluded that the 'classic' paradigm is justified by negative experiences with the originally private and unregulated service provision of the network based industries. The economic concept of 'natural monopoly' is used as an explanation for disappointing output of private companies and as explanation for imperfections of the market as a regulating mechanism in the network based industries.

On the strength of the 'classic' paradigm, regulation and government ownership can be seen as the answer to disappointing performances of companies in the network based industries in the free market. (Bauer, 1998, p.27). Availability of services in remote areas and for low income groups, fairness in price-fixing for different categories of consumers et cetera et cetera then are the variables that are particularly used to measure those performances.

The paradigm cannot only clarify what is wrong with admitting the market forces in network based industries, but can also offer a perspective on how things should and could be organized differently and for the better in network based industries. Public property and government regulation can be seen as the result of attempts to control the political and economic power of companies in network based industries; power that can lead to dissipation of scarce means, unfairness in price-fixing and inequality in the standard and the quality of service rendering.

2.2 The "new" paradigm

By now a next, 'new' paradigm has presented itself in the discussion on network based industries. In this paradigm, the promotion of market forces by allowing competition in the production of former public utilities is essential.

In the 'classic' paradigm introduction of competition in the network based industries is impossible by definition, because after all they are natural monopolies in this view. We speak of a natural monopoly when the cheapest way of production is made possible by letting one and only one provider serve the entire market. The introduction of competition by allowing several providers to operate, only stimulates inefficiency and evokes destructive forms of competition by the superfluous construction of more infrastructural facilities in this case (George c.s., 1991, pp. 362-363). For example, when dealing with the introduction of competition in the railway business, it seems rather dubious to ask of the prospective competitor to construct a second railway line next to the existing one, except maybe on the busiest stretches.

The duplication of infrastructure through the construction of a second sewerage system, a second gas supply net or a second electricity grid is also often considered inefficient and superfluous in practice. From the viewpoint of the 'classic' paradigm, competition in such a situation cannot be envisioned very well and is at least undesirable practically speaking. The 'new' paradigm on the other hand is based on the assumption that, since through strong ties to the infrastructure, it is not the entire industry that is characterized as a natural monopoly, but just parts of that industry (George c.s., 1991, pp. 362-363). For example, by distinguishing the infrastructure itself from the services rendered via that infrastructure, certain aspects of the industry can be opened up to the market and subjected to competition. To be specific: Possibilities to have more than one company use the same infrastructure for transportation of goods and people can be investigated in order to facilitate competition in the railway business.

The 'new' paradigm epitomizes a criticism on the model of the public monopoly (or regulated private monopoly) as the obvious way of organization in network based industries. The main idea behind the 'new' (or rather renewed) paradigm, is that competition should and can be given more room in network based industries.

An important explanation for the irrepressible rise of the 'new' paradigm can undoubtedly be found in the response in society today to politically tinted debate in which the superiority of the private company in a competitive market is placed opposite the public utility company that has to work in the monopolistic surroundings of the public bureaucracy.

Therewith, the paradigm shift that can now be discerned can partly be explained from a revival of the economic and political "laissez-faire"-ideas, based on a belief in competition engineering as goal in itself and the market as the only force that can really do justice to individual freedom and personal choices. However, this is only part of the explanation. Another part is that, next to this ideological shift also there is a change in the discussions on network based industries in intellectual respect, that compels reflection on the correctness of the 'traditional' paradigm. We will clarify this in the next paragraph.

3. Competition engineering: designing the market

The idea that energy supply (gas and electricity), public transport (bus and train) and telecommunication (telephone and broadcast) should be seen as industries in which competition is not possible, can be traced back to the dependence on infrastructure in these industries. In each of these industries the production of goods and services is inextricably bound to a transportation and distribution network of which the average costs decrease considerably through intensity in use. Therefore, large economies of scale can be reached. Even more important is that each of these networks deals with sunk costs. These are investments in the infrastructure needed for production and service, but which cannot be recovered on the customer.

The basis for the idea in the classic paradigm that network based industries are natural monopolies that allow no room for competition as a public utility is formed by exactly this recognition of these economies of scale and sunk costs (George c.s., 1991, pp. 365). Then, how should we understand the 'new' paradigm's desire to allow more room for competition? Where in the network based industries do we find the possibilities to introduce and enhance competition? In the following subparagraphs, we will describe four alternatives that incidentally do not exclude but complement one another.

3.1 Competition on the infrastructure/competition in the market

By making a clear distinction between possession and management of the infrastructure on the one hand and the exploitation of the infrastructure for the delivery of goods and services on the other hand, a first opportunity for competition engineering arises.

Competition on the infrastructure is made possible through isolation of specific irrevocable costs linked to the production of a specific good because of its relation with the infrastructure. These costs are isolated in separate management structures that are unrelated to the exploitation. At the same time, possible objections of dissipation of capital are no longer valid just like that (Baumol, 1982).

Several variations are possible here. One variation is leaving the infrastructure itself in public domain and subsequently allowing competition amongst private parties by letting them use the infrastructure for their services to their customers. For example in the railway-business this would mean that the railway network remains the property of the government, whereas the exploitation of the trains using the railway is left to market parties that are in competition with each other. In this example, competition is made possible by isolating the sunk costs and by dispelling the necessity for parties potentially joining the market of train-transportation to construct their own rail-infrastructure.

Another variation in facilitating competition on the infrastructure is to indeed leave the infrastructure in the hands of private parties (usually the former monopolists: the incumbents), yet by way of regulation ensure that other parties on the market have access to the transportation and distribution network, that is in the hands of the incumbents, under reasonable conditions (Gable and Weiman, 1996). For example, this is applicable in the telecommunication industry in which the former monopolist still firmly owns the network needed for wired telephone services and in which the market parties must have access to each other's networks all the time in order to be able to process the telephone calls going on over these networks.

In this case, the incumbent that disposes of the most widespread network will always be stimulated strongly to avoid or discourage the entry of competitors. For example, this can be achieved by charging too high a price for interconnection (that is means linking of networks). Therefore, the conditions for access are essential if the infrastructure is left in the hands of the market parties themselves (Klein and Gray, 1997, pp. 1-2).

3.2 Competition for the infrastructure/contestable markets

One of the opportunities for competition engineering in network based industries is the stimulation of competition on the infrastructure, in other words competition in the market. Another possibility is to let competition for the infrastructure arise. In the literature this construction is known as the forming of contestable markets. In this instance, the monopoly on the infrastructure is given to the bidder that offers the best service to the customers for the lowest prices via a franchise-construction (Klein and Gray, 1997, p.1). In case of competition for the market, companies contend for franchises that give them the right to be the sole provider of goods or services in a particular segment of the market for a specified period of time. This construction is meant to allow mutual competition and is, among others things, used in the allocation of broadcasting and telephone frequencies and the allotment of particular bus or railway line exploitation.

To illustrate how mishaps can occur when competition enters into network based industries, an example in bus transport is often used. The story goes that busses in Great-Britain were racing one another and completely irresponsible overtaking manoeuvres were used to pass each other in order to get to the bus stops first to take all the passengers that were waiting there. A problem like this can be avoided by structuring the market in such a way, temporary monopolies are allocated (in this example exploitation of specific bus lines). Therefore, the proper alternative here is competition for (parts of) the market and not competition in the market. In case of competition for the market, companies compete for the right to be sole provider, thus monopolist, for a specific period of time. By connecting the monopoly position to time limits and holding out the prospect of renewed allocations regularly, monopolists are stimulated by the authorities to keep offering good quality and to keep in touch with the customer.

3.3 Competition between infrastructures/monopolistic competition

Industries that seem like natural monopolies by way of their infrastructure based character can opened up to competition in quite a different manner. An alternative to competition on the infrastructure or competition for the infrastructure as described above, is the stimulation of competition between infrastructures (Van de Velde c.s., 1996). This form of competition is also known as intermodal or monopolistic competition in the literature. For example, in passenger transport trains face the competition of cars and busses. In telecommunication wire-bound networks become more and more exposed to competition of mobile telephone services. In future there will possibly even be competing wire-bound networks, since the costs of construction of such infrastructures are decreasing considerably. It is possible to induce such forms of competition consciously. For example by allowing busses to exploit the same long distance stretches as the train network does and to allow train companies to buy up taxi services to be able to reach similar or even better track coverage than presently possible for bus companies. Competition between infrastructures can also be enhanced by stimulating technological innovations aimed at the creation of multi-functional infrastructures: telephone lines that can transmit television signals, television lines that can handle telephone connections, pipeline networks that can transport not only ethyl or gas or water, but all of these products (and more).

3.4 Competition with the infrastructure/monopolistic competition

A final opportunity to encourage competition in network based industries is the development of alternative methods of production with which goods and services of network based industries can be produced, without having to make use of infrastructural provisions.

Consequently, in the latter alternative the infrastructure based character of production and distribution disappears. A market force that we could describe as competition with the infrastructure now comes into existence. For example, in the water industry bottles of mineral water become the competitor of tap water. An example from the electricity industry is the development of a home generator (micro total energy principle with Stirlingmotor), that can in future supply separate households with electricity and heat, like a small power station at home. Large-scale introduction of such new technologies not only offers competition to the present way of production and distribution of electricity, but can possibly even signify the end of the infrastructure in the long run.

4. Ambiguous practices

4.1 More market, less government?

According to the line of reasoning of the 'new' paradigm, a number of developments have been started up within the network based industries that have to create more scope for the market than there is at present and have to lead to production of utility provisions with less government intervention than is the case at present.

Privatization

Governments dispose of their shares in utility companies. Companies in the utility industry will become more separate from governments than they are now. At present, shares of utility companies mostly are partly or all in governmental possession. In time, many governments will dispose of their interests in utility companies. A large number of councils have e.g. sold their interests in cable companies. In time, public authorities will dispose of their interests in energy production and energy distribution companies. The future handling of the shares in waste processing companies and bus companies is under discussion periodically. Disposal of shares does not consequently entail an automatic termination of the special tie between government and these companies. For example, some councils have claimed continuance of direct or indirect influence on the program supply and subscription tariffs of 'their' cable companies even after sale of the shares.

Liberalization

Parts of the market for utility services will come into the possession of private companies. Besides the companies that already operate in the utility industry, new companies will emerge on the market. The essence of liberalization is that the markets become accessible to new companies. Markets that used to be supplied by only one government owned company as monopolist will become open markets in as far as possible, that is to say markets in which several companies compete. Some of these companies will offer the same services and goods that the traditional public utilities offer. As an example of this case, in time a foreign energy producer will offer its electricity to Dutch private consumers. There will be no difference to the consumer, since to him electricity is electricity by and large. But there will also be new companies that offer new goods and services. In the telecommunications industry this is already the case. Many new parties try to sweep the Dutch market with services that are new in part.

Deregulation

In future, utility services will be supplied with less government intervention: a larger part of the goods and services that are produced by the network based industries will be coming from companies that no longer have special ties with the government or even never had any ties with the government. When regarding the decreasing government intervention from the point of view of the present utility companies, we speak of privatization. When we regard that same development from the perspective of the entire market, we speak of liberalization. From the point of view of the government itself, we also speak of deregulation. Here a decrease in the sanction-applied limitations the government can impose on organizations and individuals is meant. In this way, within the network based industries more possibilities arise to form and establish provision of goods and services in accordance with one's own views.

4.2 More market, more government

More often than not it is thought (by supporters as well as by opponents) to be a matter of course that competition engineering in network based industries entails more market and therefore less government. However, in practice it turns out that competition engineering evokes a much more complex change that is responsible for radical shifts in public and private responsibilities, but that cannot be adequately described in terms of "more for the one, and therefore less for the other". In that respect, it is rather confusing to describe the changes that are going on presently in the supply of utility provisions in terms of deregulation, liberalization and privatization.

Deregulation?

A concept like 'deregulation' is rather misleading, when defining the developments that are now taking place in the network based industries.

Instead of current abolishment or diminishing of regulations in network based industries, we in actual fact rather witness an on-going process of dictation of new regulations (and often not less but more). An example is the telecommunication industry. Here, we currently see that new regulations that hardly or not at all existed before, are being drawn up in many countries (Hudson, 1997, p.76).

After all, there is no need for a separate system of regulation where telecommunication is part of a public monopoly. In such a situation, decisions on frequency-division, standards and prices are simply made by the same government company that is responsible for the utility itself.

When there is room for competition this changes. When countries allow competition, for example in electricity supply or in public transport, they have to ask themselves new questions, such as the following. How do we make sure that these utility provisions remain accessible to the entire population? How do we ensure that the production of utilities remains affordable for everybody? How do we realize our environmental goals and which are the means available to give social considerations due weight to in a competing market?

Moreover, by allowing competition, the government is suddenly forced to bring about a 'level playing field', with equal opportunities for former monopolists and new accessors. At the same time, that same government has to promote the interests of consumers that are difficult to reach, or are on the outside and commercially less interesting. New questions of policy come into play here: organization of access to bottleneck facilities, breaking of monopoly positions and resistance to cross-subsidization.

Even in a competing free market regulation is needed to ensure fair competition between market parties and to ensure that the company's activities eventually do not only promote their individual interests, but also the common, social interest (Hudson, 1997, p.76).

What is misleading about all this is that under the pretext of deregulation there is, in actual fact, at the same time also a development to the contrary, namely the introduction of regulation that did not exist formerly or at least to a much lesser extent.

Liberalization?

Liberalization is another concept used to describe the developments that are currently going on in the network based industries. On further consideration, this term also is rather misleading. Liberalization stands for a "liberation" of the market by allowing the entrance of potential competitors. The goal of course is that competition will occur. But to create competition in network based industries it takes more than allowing newcomers to enter the market (Hudson, 1997, p.76).

To create competition several hindrances have to be overcome, such as:

- incumbent advantages: the established monopolist will be in a position of considerable power. Technical, financial and legal (brand name) advantages, arising from the specific position of the incumbent easily make entrance of the market unattractive to new providers.
- cross-subsidization: income from non-competitive parts of the market could be used by the established parties to gain advantage in the competitive parts of the market. This also deters accession.
- bottleneck facilities: in order to be able to operate in a competitive manner, new providers have to dispose of infrastructural provisions that often still are in the hands of the former monopolists. This simple fact can easily be abused to deter potential accessors.

Thus, we cannot speak of liberalization in the sense of "liberation". On the contrary, if introducing competition is the goal to be reached, it is essential to take a large number of measures that are aimed at letting the market function unhindered and in a responsible manner.

The extent to which the market can actually be seen as liberalized is dependent on interconnection conditions and access rules. In this sense, not only interconnection of networks is needed, but also (seemingly) technical matters such as price fixing for the use of the infrastructure and the dissection of the network into separate components that are of importance to create real competition, need to be laid down. Equal opportunity for all parties is of the utmost importance for the creation of competition, but will not be realized by a "liberation" of the market (Hudson, 1997, pp.73-74).

Parties in the market that are owners of the infrastructure will try to protect their monopoly position in service rendering via the network by refusing interconnection or by fixing the entrance price at a higher level than the costs they make themselves or higher than the accessors can afford. To facilitate competition adequate conditions have to be created. Rules have to specify the technical criteria as well as the price of interconnection. Otherwise accessors will be forced to lower the standard of their services or will not be able to afford the interconnection with the existing network.

Interconnection rules are also needed to bring providers into line technically. Consumers should not have to be concerned about large differences in the quality of the service rendered, depending on which technology is used by which provider.

Liberalization as such, in the sense of leaving free, without at the same time setting additional rules for former or existing monopolists is not sufficient to actually create competition in network based industries.

Possibilities for competition can only arise when the government assumes its responsibility and designs and regulates the market in such a way that there is a 'level playing field' with equal opportunities for established monopolists (incumbents) and new accessors on the market. Therefore, the government responsibility goes way beyond liberation of the market by allowing access to potential competitors, subsequently to simply withdraw from the market (Kahn, 1998, pp.17-18).

In conclusion, it can be established that the term 'liberalization' is rather misleading as characterization of the changes currently taking place in the network based industries, since these changes entail more or less the opposite of a liberation of the market.

Privatization?

Not only the terms liberalization and deregulation are misleading when describing the developments presently taking place in network based industries. Upon closer study, the notion of privatization is not particularly clarifying in this respect either.

In itself privatization does not entail much more than the conveyance of property from a public sector organization or institution to a private sector one. Therefore, problems that might exist with monopoly positions in a particular sector simply remain after privatization. Most certainly a market or one of the above appointed forms of competition do not arise automatically after conveyance of the ownership of an organization or institution to the private sector. If it is at all possible to argue favourably for privatization in itself, the argument should be based on a direct connection between the property (public or private) and the efficiency of the organization. The argumentation in this respect is not particularly balanced and persuasive. Although a number of studies into the effects of possession on operating results have been published, the impression they convey is very variegated.

Some studies show the superiority of public companies (see: Millward, 1982), while others on the other hand conclude that private companies do better, since by definition, public institution have to make higher costs (Borchering et.al., 1982).

If we are to believe an organization like the OECD (1992, p.29) the most important explanation for this variety in findings is that "the varying interaction between ownership and competition and the effects of regulation where competition is absent". This explanation deflects us from straightforward assumptions on the superiority of one form of ownership compared to the other and leads us back again to the significance of competition and makability of the market.

In that respect, privatization in itself really lacks exactly that which is essential to the network based industries, namely creating possibilities for the market and advancing competition. For that reason, in the electricity industry liberalization of the market was put before privatization of the utility companies as regional monopolies.

Although competition engineering is presented as a way to replace the government's pathologies by the superior forces of the market in highly ideologically biased argumentation, in practice there is no such substitution (Kettl, 1993, p.14).

In fact, government and market are more and more thrown onto each other and have become increasingly interwoven by the renewed and intensified attention to the makability of the market, the use of new and innovative instruments (for example auctioning and franchising) and variations on regulation in conformance with the market.

All in all, the term privatization is rather misleading as designation for a development in which privatization of public companies (if they exist) is far less essential as is the governmentalizing of the private sector as a result of new ambitions to (re)design the market in the network based industries.

5. Changing pendulums

5.1 The end of history?

If we are to believe Francis Fukuyama we are witnessing the end of history. In his view, the universalization of the liberal democracy as the ultimate form of government marks the end of history (in the Hegelian sense of a continuous confrontation of ideologies).

At the moment, it seems rather fashionable to announce the end of all kinds of things: our history, our ideology, our democracy, our national states et cetera. It must be the season for it: the end of the century and of a millennium even is near.

Still, that does not do away with the following question. Are we about to enter a post-historic period with the prospect of centuries of endless boredom where network based industries are concerned? Will all changes that are to be expected in future only be the aftermath of the privatization, liberalization and deregulation that are taking place at present? Or should we see these developments as the umpteenth movement of the pendulum, that sooner or later will inevitably be followed by a countermovement?

In our opinion the history of the network based industries indeed shows the characteristics of a pendulum-like movement at first sight. However, on further consideration, matters are more complicated, because the notions with which we describe the current developments are misleading: they suggest a movement in one direction, whereas at the same time, there is also a movement in the opposite direction. However, it would be a mistake to think that therewith the predicted end of history is near: the search for dilemma-proof arrangements continues.

5.2 At first sight: a pendulum-like movement

Infrastructural provisions such as roads, canals, railways, gas and electricity lines, sewerage and water systems were completely in private possession until into the nineteenth century. Interference by the government was minimal, if at all (Klein en Roger, 1995, p.1).

This changes in the twentieth century. Not only do infrastructural provisions themselves become more and more governmentally owned, but also companies producing goods and services via those infrastructural provisions become more and more regulated or even nationalized. Government interventions based on the 'traditional' paradigm as described above are the cause of this.

In the course of the seventies and eighties of this century a next change took place in many countries whereby the emphasis shifted from the government back to the market, because of dissatisfaction with efficiency, customer-orientation and innovativeness of companies in the network based industries.

A flood of privatization, liberalization and deregulation sweeps the world based on the 'new' paradigm already explained above. It seems this movement fits a historical development characterized by a pendulum-like movement (Klein en Roger, 1995, p.1).

The next 'swing of the pendulum' is surely to be expected, when we realize that, although privatization, liberalization and deregulation offer promising perspectives to the network based industries, they also entail substantial risks.

It is one thing to state that there should be competing markets in network based industries, it is a different matter entirely to actually realize this in practice. Here we find that no wish for the future, however promising, can be realized without overcoming some difficulties and hindrances first.

We can hardly speak of a pendulum-like movement where under the pretext of deregulation, in actual fact (re-)regulation is going on, where liberalization really is anything but liberation of the market and where privatization involves a movement that accomplishes exactly its opposite.

That conclusion is yet again enhanced by realization of the fact that the paradigm shift taking place now in the idea forming on network based industries is more that a ideologically biased re-discovery of the classic market ideas. As has been said before, it certainly also entails an intellectual progress in idea forming on possibilities for market functioning in network based industries.

The re-discovery of the market does not entail an unconditional return to last century's society, although the contrary is sometimes contended very firmly (a/o. Hoogerwerf, 1995, p.78).

5.3 On further consideration: complementation between government and market

Competition has to be carefully engineered and regulated. For that, from the point of view of the 'new' paradigm and in light of the existing practices, the starting point will always have to be a complementary relation between government and market and not the monopolization of one order principle at the expense of the other.

Withdrawal of the government as owner of the infrastructures and as owner of the service rendering companies on the infrastructures bears a consequence; the government will subsequently have to play a new and different role. This role entails the engineering and regulation of a market that investors find trustworthy and consumers find fair and legitimate and, moreover, which can guarantee efficiency for the economy as a whole.

After all, it is unmistakable that the introduction of market functioning and competition potentially entails merits as well as risks (Kettl, 1993, p.164).

For example, a potential danger is that private companies try to penetrate the market of public utilities by undercutting in the expectation that they will be able to increase prices once the monopoly position has been acquired.

Thus, the danger in this case is that after privatization, liberalization and deregulation in the network based industries companies will be developed that gradually succeed in obtaining a position of monopoly or that can abuse their own share in the market to prevent the accession of new competitors by means of forming a conglomerate with other companies (oligopoly).

Sometimes companies reach informal agreements amongst each other not to compete in price when governments turn to them to realize public means via private production, for example by inviting tenders. This is a recurring problem in road-construction.

Finally there is the danger of fraud. Corruption can occur when the market is called in. A historic reason for involvement of bureaucracy in the beginning of the twentieth century was the protection of public interests against corruption by private parties.

The much-praised disciplining effect of the market exists only if and in as far as there is mutual competition between market parties that awards success and punishes failure (Kettl, 1993, p.180). When the market's disciplining effect somehow is not realized, it is inevitable that problems arise, varying from conflicting interests and abuse of economic positions of power to fraud.

Markets should be carefully engineered and regulated in order to function well (Osborne en Gaebler, 1992, pp.104-106). Whether it is about public transport or telephony, companies operating in network based industries can be inclined to skim the most profitable segments of the market: the most cost-effective bus routes or those consumers that are willing and able to pay the most for telephone services.

Whenever markets are not carefully designed or regulated, competition can be undermined. For example, when private companies are given concessions to operate a power plant for the period of 20 to 30 years, it can safely be assumed that the disciplining effect of the market through introduction of competition incentives will hardly be at play here.

Even when private companies do not assume a monopoly position, they can still gain enough political and economic power to undermine competition (Osborne en Gaebler, 1992, p. 106). When competition is not engineered and regulated sufficiently careful, private bus companies for instance may be tempted to spend large amounts of money on lobbying the legislator and/or on bribes to acquire and maintain their contracts. Contractors involved in waste-disposal services could be tempted to apply their political influence against measures such as recycling and source policies aimed at reduction of the amount of household refuse.

On the other hand, competition that is carefully engineered and regulated can possible ensure more socially fair results that a government monopoly can (Osborne en Gaebler, 1992, p.105).

Private parties can be forced to render commensurable wages and prices or to stimulate specific desired behaviour. It is important to realize that essential values and standards and public goals the government has, do not have to be lost with the introduction of competition. Private contractors can well be compelled to render service to all segments of the market in order to avoid skimming.

Carefully designed and regulated market functioning can serve public interest and therewith is the opposite of selling out public responsibilities to the 'free market'. The free market does not exist, that is to say if a market free of government regulation is meant. All legal markets are structured by government rules. The only market that is not in one way or another regulated by the government is an illegal market.

5.4 Government regulation: from substitution to condition for market functioning

To achieve a complementary relation between government and market and to maintain it, new and innovative forms of regulating are needed and not an unconditional belief in the self-regulating functioning of the market or a sale of government provisions to the market.

In that respect it is not unimportant to conclude that regulation was mostly used as a substitute for competition whenever competition was not thought to work or when it failed in practice. This continued until halfway through the eighties. Nowadays a different view is taken.

Now regulation is not so much so regarded as a substitute for competition, but as a supplementation of competition intended to remove possibly negative effects of competition such as unequal accession for service providers or large regionally-bound differences in price (Bauer, 1998, p.4).

With the introduction of an approach to public utilities that is more favourable to competition, the basic principle of government regulation has shifted in the direction of facilitation of competition (for example through free access and interconnection policy) and guidance of the transition to more competitive market structures.

More and more, traditional forms of regulation are replaced by new, innovative instruments of regulation that are in conformance with the market. An example of this is the substitution of rate base/rate of return-regulation by performance based-regulation. Another example is the use of 'green-power' certificates in the power supply industry. And further the use of concession conditions as a way to achieve public goals can also be regarded as a form of regulation that is more in conformance with the market.

More and more emphatically, regulation is seen as a precondition to competition instead of as a substitute thereof, as it was usually seen in the 'classic' paradigm of network based industries as public utility provision (Bauer, 1998, p.8).

6. Conclusion

6.1 Unresolved tensions, remaining dilemmas

Although the changes in the network based industries that are currently taking place are drastic and far-reaching, we can ascertain simultaneously that many of the underlying problems of policy have not changed in essence. The central tensions and dilemmas remain essentially the same, in spite of shifting paradigms in the idea forming on network based industries.

This should not surprise us. We are used to calling the products and services of network based industries public utilities, because they stand for certain values. Among others, these values are affordability, accessibility and reliability. Most (if not all) of these values can be interpreted in different ways. Furthermore they clash with each other and with other values that we also find to be important in the production of utilities: efficiency, customer-orientation and innovation. Such conflicts in values are hardly ever resolved completely with the restructuring of an industry, no matter how drastic the changes might be. Take, for instance, the distributive goals that are connected to the production of utility provisions: affordability for all income groups and accessibility to all. Because of their monopoly positions in network based industries, until recently, utility companies were able to render services under cost price to consumers with a low income or to customers that would be confronted with very high costs because of their remote living locations. In this regard, examples from the telecommunication industry are cross-subsidization of urban to rural telephone users and of long distance callers to local callers.

When monopolies are laid on the table and competition is allowed, consumers that pay prices that are much higher than the costs made in their cases will start looking for other providers. And new entrants on the market will want to provide their services to exactly those market segments where the prices surpass the costs (such as bulk consumers, big money makers) and in that way gradually eliminate the possibilities for cross-subsidization by the utility companies.

Distributive goals cannot and will not (have to) be relinquished just like that in a situation in which market forces and competition come into play. However, other instruments that are in conformance with the market will have to be found to achieve these goals; these are instruments of which practice value has to be proven as yet (an example of such an instrument would be formation of a fund for the financially weak via a general surcharge on the consumption rate of the infrastructure).

The current developments in network based industries change the context in which conflicts on the realization of, for instance, distributive goals are pronounced, negotiated and fought out. Furthermore they influence the relative weight that parties involved will attribute to specific values.

Finding a new equilibrium in these matters can relieve persistent tensions, but will also cause other tensions or revive forgotten problems. No matter how successful a new equilibrium might be at a certain point in time, it will not be everlasting.

6.2 The narrow margins between 'Yes, if 'and 'No, unless'

The developments currently taking place in network based industries are not (only) founded on ambitious and compelling lines of argument, but (also) in practical policy forming. Bringing about changes in network based industries turns out to be a matter of narrow margins. In competition engineering, it is not a matter of a principled and inspired choice between market and government, but rather a practical and subtle consideration of 'yes, if' and 'no, unless'.

Competition engineering in network based industries cannot be seen as similar to a substitution of market for government. Instead, it is a matter of careful experimentation with new equilibria and with the organization of an instructive environment by allowing a certain differentiability in finding those equilibria. Furthermore it is a matter of preventing damage to society by avoiding and/or postponing arrival at points of no return as much and as long as possible.

References

Aucoin, P. (1990), Administrative Reform in Public Management: Paradigms, Principles, Paradoxes and Pendulums, in: *Governance: An International Journal of Policy and Administration*, vol.3, nr.2, April, pp.115-137.

Bauer, J.M. (1998), *The role and evolution of public utility regulation. Tasks of regulation in an era of partial competition*, Paper presented at 40th NARUC Annual Regulatory Studies Program at Michigan State University, East Lansing, Michigan, August 3.

Batten D.F. and C. Karlsson (eds.) (1996), Infrastructure and the complexity of economic development, Berlin:Springer.

Baumol, W.J., Panzar, J.C. and R.D Willig (1982), *Contestable markets and the theory of industry structure*, San Diego.

Bennet, J.T. and Johnson, M.H., *Better Government at Half the Price: Private Production of Public Services*, New York: Caroline House Publishers.

Biehl, D. (1986), The contribution of infrastructure to regional development, Final report, Office for Official Publications of the European Communities, Luxembourg.

Colling, T. (1994), Commerce versus Politics, Compulsory Competitive Tendering and the Determination of Employment Policy in a British Local Authority, in: Clarke Thomas, *International Privatisation; Strategies and Practices*, Berlin/New York: Walter de Gruyter, pp.119-138.

Cox, K.R. (1972), Man, location and behavior: An introduction to human geography, New York.

Demsetz (1968), Why regulate utilities? in: Journal of Law and Economics, nr.11, pp.55-65.

Diamond D. and N. Spence (1989), Infrastructure and industrial costs in British industry, London.

Donahue, J.D. (1989), *The Privatization Decision: Public Ends, Private means,* New York: Basic Books.

Gabel D. and D. Weiman (1996), Sustaining Competition in Network Industries trhough Regulating and Pricing Access, Dordrecht: Kluwer Academic Publishers.

George, K.D. and C. Joll, E.L. Lynk (1991), *Industrial Organisation: Competition, Growth and Structural Change*, in New York: Routledgefourth edition.

Hatry, H.P., *A Review of Private Approaches for Delivery of Public Services*, Washington D.C.: Urban Institute Press, 1983.

Hoogerwerf, A. (1995), *Politiek als evenwichtskunst, Dilemma's rond overheid en markt*, Alphen aan den Rijn

Israel, A. (1992), *Issues for Infrastructure Management in the 1990s*, World Bank Discussion Papers, Washington D.C.: The World Bank

Kahn, A.E. (1998), *Letting Go: Deregulating the Process of Deregulation*, East Lansing, Michigan

Kessides, C. (1993), *Institutional options for the provision of infrastructure*, World Bank Discussion Papers, Washington: The World Bank, Washington

Klein, M. and N. Roger (1995), *Back to the Future: The potential in infrastructure privatization*, in: Public Policy for the Private Sector Series, World Bank

Klein, M. and P. Gray (1997), *Competition in Network Industries - Where and How to Introduce It*, in: Public Policy for the Private Sector Series, World Bank

Naschold, F. (1996), *Modernization of the State: Structural reforms and innovation strategies of the public sector*, in: F. Naschold and C. von Otter, Public Sector Transformation, Rethinking markets and hierarchies in government, Amsterdam/Philadelphia: John Benjamins Publishing Company

National Research Council (1995), *Measuring and improving infrastructure performance*, Washington: National Academy Press

OECD (1992), *Regulatory Reform, Privatisation and Competition Policy*, OECD/OCDE, Paris

Robinson, C. c.s. (1994), *Regulating the Utilities: The Way Foreward*, Institute of Economic Affairs

Rosenthal e.a. (1997), *Openbaar bestuur, beleid, organisatie en politiek*, Alphen aan den Rijn

Savas, E.S. (1987), *Privatization: The Key to Better Government*, Chatnam

Simonis, U.E. (ed.) (1977), *Infrastruktur, Theorie und Politik*, Köln: Kiepenhauer & Witsch

Stiglitz, J.E. (1988), *Economics for the public sector*, New York, pp. 1-23, 61-90

Twist, M.J.W van, Veeneman, W.W. (1999), *Marktwerking op weg, Over concurrentiebevordering in infrastructuurgebonden sectoren*, Utrecht, Lemma

Velde, D.J. van de, P. van Reeven and L. Sleuwaegen (1996), *Marktwerking in Nederland: Het openbaar vervoer*, OCFEB, Erasmus Universiteit

Vickers J. and G. Yarrow (1991), *Privatization: An Economic Analysis*, Cambridge

Wolf, C. (1993), *Markets or governments: choosing between imperfect alternatives*, Cambridge, pp.1-35, pp.59-103

Conceivable Administrative and Judicial Variations

Ernst F. ten Heuvelhof, Martijn Kuit, Helen D. Stout

Delft University of Technology
Faculty of Technology, Policy and Management
Public Management group

Abstract

A description of possible changes to administrative and judicial arrangements which might influence the shaping and management of infrastructures is given in this article. The changes described are substantial; we are not saying that they will happen or even that they are likely, merely that they are conceivable

In assessing the likelihood that the changes described here will take place, the recent implementation of major changes in this area, such as liberalization and privatization developments, should certainly be taken into consideration. These changes have had a variety of causes and start times; developments within the EU can trigger changes in the Netherlands, but sometimes the origins lie within the country: for instance, the financial relationship between government and utility companies. Technological developments may also be associated with change.

The changes we describe are principally linked to the roles distinguished in the infrastructure regime. These roles are played by the government, the utility companies and the end users.

1. Introduction

A description of possible changes to administrative and judicial arrangements which might influence the shaping and management of infrastructures is given in this article. The changes described are substantial; we are not saying that they will happen or even that that they are likely, merely that they are conceivable. The chance that these changes will take place, together with their profound significance, mean that it is well worth thinking through their consequences for the shaping and management of infrastructures in this way. In assessing the likelihood that the changes described here will take place, the recent implementation of major changes in this area, such as liberalization and privatization developments, should certainly be taken into consideration. These changes have had a variety of causes and start times; developments within the EU can trigger changes in the Netherlands, but sometimes the origins lie within the country: for instance, the financial relationship between government and utility companies. Technological developments may also be associated with change, although here it can be difficult to distinguish between cause and effect. However the changes are initiated, their effects are profound and are felt across a broad front. We have no reason to believe that these changes are going to slow down in the years to come; on the contrary, there are signs that recent discussions and changes have released forces that have yet to take full effect.

The changes we describe have been ordered in accordance with Figure 1, which distinguishes between the three groups of actors invariable found to be associated with the provision of a given utility: government, utility companies, and end users. These three groups operate within the context of a certain regime, an 'infrastructural regime'. This is shown as the area within the dotted line in figure 1, while the solid lines represent the relations between the groups. The regime comprises the roles and regulations which give concrete shape to the utility companies and which form and facilitate future developments. The changes we describe are principally linked to the roles distinguished in the infrastructure regime; besides these, we describe a smaller number of changes within government and end users which, though they lie outside the regime, will nevertheless affect the design and management of infrastructures.

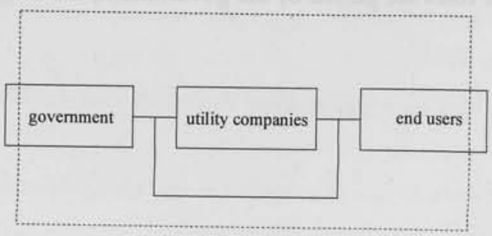


Figure 1. *Infrastructure regime*

2. Infrastructure regime roles

The following roles may be distinguished within the infrastructure regime:

- **Owner**. An infrastructure or network always has one or more owners. Who is the economic owner? Who is the legal owner? How has this ownership been framed? What benefits and disbenefits are associated with this ownership? What are the owner's powers and responsibilities?
- **Manager**. The manager is responsible for the running of the network; the manager is often required to allocate network space and make it pay its way.
- **Service provider**. The service provider uses the infrastructure, and needs it to make his services available, usually to end users. By the same token, they pay fees to the network manager or owner in payment for this use of the infrastructure space.
- **End user**. The end user is the final consumer of the services that the service provider makes available on the network. The end user pays for the enjoyment of this service.
- **Legislator**. The traffic of services and payments between the owner, the service provider and the end user is invariably subject to regulation. In the absence of specific regulation, then ordinary regulation - including public and private judicial law – applies; however, it is almost invariably the case that the infrastructure regime also includes specific legislation.
- **Regulator**. Where rules exist, they must be policed. As for the rules, there will always be a baseline level of compliance supervision, but given the special character of infrastructure provisions and the special legislation that usually applies to them, this supervision often takes special forms.
- **Dispute arbitrator**. The interests involved in infrastructures and the services they provide are considerable; the services are held to be essential, large amounts of money are involved. The many changes still on the agenda mean that many of the areas in which these interests are at work remain unclear; the relationships between actors are changing, roles are changing, some interests are as yet undefined. This is fertile ground for conflicts, which then have to be settled, here, too, conventional arbitration arrangements will be available in the event that no special arrangements have been made.

The following points may be noted about these roles:

- All roles exist in all infrastructures, though not always in an explicit form;

- The roles are divided between the various actors involved. Governments, semi-government organizations and utility companies are prominently involved in the division of roles. It is possible for an actor to fulfil more than one role.

We shall now outline each of these roles in turn, and describe the conceivable changes for each role. After having exhausted this list, we shall mention a number of possible changes likely to affect more than one role.

2.1 Owner

Parts of an infrastructure network may be owned by a single legal entity ('concentration') or by a group of such entities ('diffusion'). Various types of concentration and diffusion can be found in different infrastructure sectors. A single actor, for instance, may own different links in the production column: the same legal entity might own the drinking water network, the sewage network and the sewage treatment plants. Certain links in the chain may be concentrated to an owner horizontally: for instance, a single party might own several comparable links in different infrastructures. Yet another form of concentration concerns the scale of the networks involved: a single party might own an entire regional, national, or even international network. An extreme form of this type is when *all* infrastructures are in the hands of a single entity.

The degree of diffusion of ownership rights can vary between infrastructure sectors. Separate links may have different owners; different parts of a network may be distributed between a number of owners. Diffusion, too, has its extremes; consider the situation, for instance, in which the end users are also the owners of the infrastructure. In addition, different combinations of concentration and diffusion can exist; for instance, there can be horizontal concentration in specific parts of the network as a whole.

The degree of concentration or diffusion of ownership rights can have an historical background. In the Netherlands the various links in the water infrastructure sector have traditionally been in the hands of water companies (drinking water supply), municipalities (the sewage network) and water boards (sewage treatment plants). In other countries, examples can be given in which ownership rights in the water sector are vertically concentrated with a single entity, which may also possess the entire regional network or multiple networks. Economies of scale and economies of scope are often cited as the justification for this kind of concentration; remarkably, the opposite trend can be seen in several other infrastructure sectors, in which ownership rights are being spread across more parties. An important reason for this is that the 'monopolistic' character of various network links can differ; one link may be subject to a strongly regulated regime, while another is not. Each regime requires the owner to adopt a specific approach.

2.2 Manager

As we have said, the manager generally allocates network space and ensures that it pays its way. In performing this allocation, managers may employ one of two main principles: one based on technical and administrative issues, and one based on technical and commercial issues. Note that both principles comprise a technical component; it cannot be denied that technology plays a dominant role in the infrastructure sectors being considered.

The technical and administrative issue-based variant is based strongly on traditional views on infrastructure use: infrastructure is a collective good, and all factors tending to obstruct the free use of the infrastructure should be eliminated. This variant is characterized by a high degree of inertia; once the supply of services has been guaranteed, then all that remains to be done is to consolidate the situation. Large organizational changes in the sector are taboo, and there is little variation in the number of organizations with which the manager has contractual relations.

The technical and commercial issue-based variant is characterized by dynamism. Business and commercial motives determine the manager's decisions; the infrastructure is seen not as a collective good, but as a product with which money can be earned. The even-handed treatment of different clients and across-the-board prices do not apply. There can be large numbers of contractual relations and a wide variety of parties with which these contracts are framed. The number of these relations can change markedly over time. In short, the manager does whatever is best for the organization, and this may include opportunistic behavior.

2.3 Service provider

Changes in scale and attention area

Service providers employ the infrastructure to make their services available to end users, they pay for using the infrastructure and they are paid by the end users for the services they provide. There are a great many different kinds of service providers, but two variables determine their nature: the scale level of the service provider, and the scale level of the market in which the service provider is operating. The scale level of the service provider can be regional; for instance, the organization might operate only in a certain area; or it can be national or international. The same scale level distinctions apply to the market in which the service provider operates. The variety of possible combinations open to a service provider is shown schematically in matrix form in Figure 2.

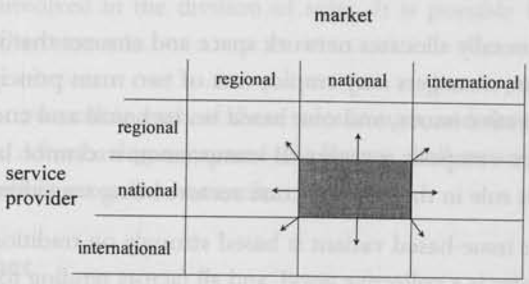


Figure 2. *The nature of the service provider*

A service provider may operate in the regional-regional cell: in this case its organization is organized at the regional level and its market is also a regional one. An example of this type of provider might be a water supply company. The water market is generally a regional one; the high cost of transporting water means that there is usually a financial disincentive to build networks that cross regional borders. In the US, for instance, water service providers are generally organized along regional lines. Historically, paternalistic motives ensured that every region or town possessed its own water supply company. A Dutch example of a service provider which used to be found in the national – national cell is the passenger travel section of the *Nationale Spoorwegen*, the national train company. The *NS* is a nationally-operating company whose market was originally limited to the Netherlands.

As has been indicated in Figure 2, a service provider can move activities into other cells. Vertical movement is within its own control. A service provider operating at the national level can decide to divide its activities into a number of regional compartments, to move independently into foreign markets, or to undertake a strategic alliance with another organization in order to move onto the international market. An example of vertical displacement is given by the merger of two regional electricity suppliers to form a national electricity company. Naturally, this vertical movement can also be in the other direction.

Horizontal displacement, however, is something that the service provider cannot usually initiate or resist. Horizontal movement, particularly towards the right, can result from a market opening up to new players; movement to the left or the right can be the result of technological innovations opening up new possibilities for providers. This last scenario is one in which providers exercise a degree of control.

Integration versus fragmentation

Looking at the production column within infrastructure sectors, different service providers can be distinguished in several ways. Depending on their place in the production column, they either provide services to other service providers or directly to the end users; each represents a given link in the production column. For instance, an electricity producer supplies services to an electricity distributor, who in turn supplies services to electricity end users.

The number of links in the entire production column will vary between infrastructure sectors, as will the occupation of each link by service providers. A service provider, for instance, may operate several links; equally, a given link may be represented by several providers. There may be certain developments involved in this setup. For instance, we speak of 'vertical integration' when a given provider controls an increasing number of links in the production column, for example, when the same company supplies drinking water and deals with waste water.

We speak of horizontal integration, for instance, if a service provider controls the same links across different production columns. These might be identical production columns, for example, when two or more service providers merge; or production columns from different infrastructures, for example when a single service provider controls the same link in the electricity network and in the water network, for instance, electricity distributors sometimes also collect water rates on behalf of water boards.

Far-reaching horizontal and vertical integration can ultimately lead to the creation of 'utility supercompanies', service providers active within more than one infrastructure and which control large sections of the production column.

The degree of integration varies per infrastructure and per country. In fact, infrastructure sectors may be subject to various degrees of fragmentation, with service providers electing to concentrate on a specific task. This can lead to the splitting up of companies. In the US, three infrastructure sectors now show strong vertical fragmentation: gas, electricity and telecommunications. In several states the production column of the electricity sector has been split into four parts: generation, transmission, distribution and aggregation. However, the American water sector is undergoing a strong integration.

2.4 End user

Relation between service provider and end user

The judicial relationship between service provider and end user can be shaped in one of two ways: either the parties involved can determine their mutual rights and duties by means of a private law contract, or the legislator can prescribe the rights and duties that each party undertakes to uphold when entering a legal relationship. Both private and public law provisions are invoked.

The principle of autonomy, of the right to self-development, is the fundamental principle underlying the concept of contractual freedom; the restraints under which each party are placed are freely entered into. When two parties choose to define their relationship by means of a contract, then the law holds that their express agreement is of sole importance to the establishment of their legal relationship. The general law provisions of the Dutch Civil Code contain practically no limitations of the freedom of individuals to enter into contractual agreements. The principle of autonomy therefore also implies that a private law contractor is entitled, in principle, to contract whatever it wants with whoever it wants. In this connection a distributor, for instance, is not obliged to enter a supply agreement with a certain end user (perhaps because the end user refuses to pay a certain price); and by the same token, an end user is free to opt *not* to be an 'end user'. The expectation is that the very special character of infrastructures will diminish in importance and that the relation between service providers and end users will be normalized to a higher degree.

The concept of the 'abuse of authority', a tenet well known to jurisprudence, forms an important limit to the principle of autonomy. The law considers an abuse of authority to exist when one party, in all reasonableness, cannot legally exercise this authority after having weighed up the interest of this exercise of power against the interests damaged by so doing. In certain circumstances, failure to meet this proportionality requirement mean that any damage has to be compensated for by the party causing the damage. In such cases, 'ordinary' legal stipulations pertaining to evidence and the like, as found in the Dutch Code of Civil Procedure, apply; rules that were framed from the standpoint that parties to such legal conflicts would be a match for each other. The legislator can prescribe to contractual parties which rights and duties are attached to their contractual relationship. An important reason for the legislator to intervene in the autonomy of contractors and independently to influence social relations in this way is the need to protect weaker groups, particularly in the context of structural inequalities between parties; in such cases the legislator limits the contractual freedoms between parties by establishing special contract provisions, such as those affecting employment contracts and tenancy contracts.

These points may mean that the agreement between a provider and an end user is qualified as a special contract in the sense that the Civil Code may specify a number of constraints to which the parties are bound; as a consequence of which, for instance, the distribution company has an obligation to connect and supply the service concerned and the end user, besides having the right to enjoy these services, also has an obligation to accept such connection and to provide a demand for this service. Another reason for government not to leave the question of rights entirely up to the contractors alone is the damage to wider interests that might be incurred by a failure to meet contractual agreements.

It should be noted that in an important number of cases, these wider interests are not (or not directly) linked to the interests of a private party, and that in such cases it is not enough merely to effect measures which protect the interests of the weaker, in the sense described above party to a contract, examples of this would include environmental values and general economic welfare. In these cases, the legislator will invoke public law to apply judicial steering to contracting parties, by imposing certain rights and duties. Within this framework, the legislator may elect, for instance, to apply special price structures to infrastructural products and services.

Supply-driven versus demand-driven consumption

With regard to the consumption of services by end users, two extremes may be distinguished: consumption is determined by the service provider ('supply-driven') or by the end user ('demand-driven'). 'Supply-driven' implies that the end user has little or no power to influence the quantity or quality of the goods or services being supplied. The service provider offers a standard package at a fixed price, such as a minimum number of minutes of telephone time per month for a certain price; and it is the service provider who determines the quantity and quality of the package. Technological and financial motives will inform the definition of optimality underlying the choices made with regard to this quantity and quality.

'Demand-driven', on the other hand, represents a situation in which the end user enjoys unlimited flexibility in deciding the quantity and quality of the services the end user opts to consume. For this reason, the service provision process is entirely dependent on the demands made by end users on quality and quantity. In this model, the end user has a dominant influence on the service provision process; in consequence, the market is characterized by a great variety of different service packages.

In principle, the end user is then free to set the level of service quantity and quality; in practice, this freedom is often limited by the pricing structure employed by the provider. 'Inclining block rates', for instance, mean that the per unit price increases with the number of units consumed; seasonal rates also influence the freedom of the end user.

In terms of service quality, the end user usually has little freedom of choice, however, the Dutch water sector is an exception to this rule: in several parts of the Netherlands, users particularly large-scale users can choose between several levels of water quality. The transport sector also comprises several quality distinctions e.g. first / second class train travel. Nevertheless, choice usually remains limited for the end user.

End user flexibility can also take on other forms; for instance, an end user might be able to negotiate the price of the service to be delivered. The consequence of this is that the service provider spends a great deal of time negotiating with all its clients, and the enormous transaction costs that this bilateral negotiation generates will certainly have a negative effect on the sales price. As a reaction to this, 'aggregators' may enter the market: organizations or individuals that bring together certain service providers and groups of end users. In the American electricity sector, aggregators already form a fixed link in the infrastructural chain in several states.

2.5 Legislator

The legal provisions which apply to service infrastructures can be distinguished by the way in which they arose. When framing legislation, the legislator can call on two fundamentally different strategies: the vertical and the horizontal. In principle, the vertical approach means that the regulator unilaterally enacts prevailing law; here, the term 'unilateral' should be understood in the context of the relation between those setting standards and those required to abide by them. The distinguishing feature of the vertical approach is that the framing of legislation accords no special position to social groups or organizations, either formally or materially, whose interests are affected by the legislation in question *as such* - by which it is implied that institutionalized interests have no special or exclusive influence in advance on given decisions, including the framing and passing of legislation. This said, of course, it is quite possible that in a given instance a written or unwritten law will assert that the interests being advocated by certain groups or organizations must find some sort of expression in the results of the legislators' deliberations; for instance, by offering financial compensation for damage occasioned by the establishment of certain laws.

This vertical approach is generally known as the 'command and control' approach, one in which the legislator makes use of a monopoly of coercion without which the legislators rules cannot be pushed through. The legislator figuratively and literally *prescribes* the law by forbidding or compelling certain behaviors. Such a method of regulation can be particularly well defended from the point of view of the classic concept of the principle of legality, the principle that government intervention should be based on formal laws which accord powers to the government body concerned. For instance, the general rule-and-decision model - the administrative code of practice which expresses the principle of the legality of government - can be seen at work in the design of the Electricity Act, which comprises a general provision forbidding utility companies to supply electricity to (particularly small-scale) end users, unless a license to do so has been specifically granted. According to the commentary, the license system was introduced primarily to protect 'captive consumers', that is, consumers who (as yet) have no freedom of choice between suppliers. The Act describes the general conditions that a utility company must meet to be taken into consideration for a license. The body issuing the license is also authorized to attach further conditions to its conferral; the principal issue here is that the management exercises unilateral authority on the basis of general, abstract rules which apply equally to identical situations.

In the horizontal approach the legislator is a partner, together with wider society and its representatives with which it is on an equal footing, to a process of dialogue and negotiation which ultimately produces legislation. This approach is also known as 'negotiated rulemaking'. The criterion which distinguishes the horizontal from the vertical approach is that regulations are in principle created in participation with, rather than in the absence of, the organizations and groups affected by the proposed regulations.

Legislators establishing law in a procedure that can be characterized as involving dialogue and negotiation with societal representatives can manifest in two ways: formally or informally. In the latter case, the legislator goes through the appropriate, unilateralism-based procedure, but at the same time makes special, informal arrangements which are invisible, in principle, to the formal instruments normally employed by the legislator. The contract is one of the best examples of judicial forms in which dialogue and negotiation are embedded, and when a horizontally-working legislator also formally operates as such contract law is generally employed.

The information which was needed to make these laws was often controlled by the traditional utility companies. That's why they were extremely important in making these laws. Other actors, incl. users, didn't have an important function. In the future more actors will influence the process of rulemaking.

2.6 Regulator

The preventative or repressive supervision exercised in infrastructures can be organized in a number of ways, though in principle, two modalities may be distinguished: the sector-specific regulator and the general regulator. The sector-specific regulator is a regulatory authority appointed for the exclusive supervision of a certain infrastructure. It has no dependency relationship with other regulators, nor is it linked with them in any other way. The general regulator, like the independent regulator, is charged with policing compliance with sector-specific regulations; it should be noted that no distinction can be drawn between the two modalities in this respect.

The difference between general and specific supervision lies predominantly in the kind of *organization* in which this sector-specific supervision is embedded; in the case of general supervision, no special regulatory authority is appointed. A Dutch example of sector-specific supervision in the telecommunications industry is given by *OPTA*, the

Onafhankelijke Post en Telecommunicatie Autoriteit (the Netherlands Regulatory Authority for the Telecommunications and Postal sector); another Dutch example in the electricity sector is the *Dienst Uitvoering en toezicht Elekriciteitswet* (the Implementation and Electricity Act Supervision Service). Besides supervision, sector-specific regulators are often also responsible for providing services in the sector concerned, and in this capacity they possess regulatory and other decision-making authorities within the meaning of article 1:3 of the *AWB* (the Dutch Act on General Administrative Law, the third tranche of which contains general provisions with regard to regulatory powers).

Combining different types of authority in this way has a number of important advantages; one, for instance, is that the experience gained in carrying out the implementing role can be put to excellent use in the regulatory role. An example of a Dutch general regulator is *NMa*, the *Nederlandse Mededingingsautoriteit* (the Dutch Competition Authority), which is charged with policing compliance with the Competition Act as well as carrying out the necessary tasks arising from the Competition Act. In principle, the Competition Act applies to all businesses and organizations active on the Dutch market, and to all economic sectors therein. The criterion for its application in any area is that economic activities are involved.

One argument for this modality is that infrastructures exist in which market forces are gradually taking hold, and in which it will eventually be true to say that a full market situation, in the sense of the Competition Act, obtains. An argument for sector-specific supervision is that stricter supervision is temporarily needed during a transition process towards market relations, supervision which calls on specific powers (e.g. price control measures). Broadly speaking, in selecting between these two modalities it will be necessary to weigh up the importance of a consistent competition policy (coherence and uniformity of normative and conceptual frameworks) and considerations having to do with the more sector-specific organization of sector-specific supervision.

2.7 Dispute arbitrator

The settlement of the disputes that arise in infrastructure management and use can be organized in any of four different ways. At the institutional level, the following modalities can be distinguished: the general arbitrator and the sector-specific arbitrator, each of which can be said to be working either on an incidental or a structural basis.

The general dispute arbitrator has a broad area of competence, in the sense that it is authorized to settle the disputes that arise in a number of different societal areas and is therefore not confined to a single area. As far as judicial arbitration is concerned, for instance, a civil court can be described as such a general arbitrator. According to Article

112, section 1 of the Constitution, civil (or 'ordinary') courts are authorized to hear disputes arising from a litigant's claim that their subjective right, as derived from objective civil law, has been infringed: in constitutional terms, 'disputes concerning civil rights and claims'. In principle, the competence of the court to hear such a dispute is unaffected by the area (e.g. infrastructural provisions, housing, work, school) in which the dispute arose. From the point of view of legal unity, this modality has important advantages.

The sector-specific arbitrator is a conflict settlement authority set up with a view to settling the disputes that arise in a given infrastructure area. Examples would include an 'electricity court', 'telecommunications court' or the like, arbitrators whose competence covers a single, well-defined societal terrain. An example of such a special court in current Dutch law is the *Tariefcommissie*, the Tariff Committee, an independent legal college set up in 1935 which deals exclusively with tax disputes arising from customs, import and excise duties. In this working area the Tariff Committee has evolved a unique identity, and has been able to retain its character principally because its judgements are not open to appeal (Tariff Committee Act, Dutch Penal Code 1994, 7).

Several arguments for a special court can be put forward: the need for courts having special skills, the 'sensitivity' of society towards outside intervention (the legitimacy issue), and the wish to create a purpose-built procedural law model rather than simply adopt a standard model.

The broad distinction being made between general and special dispute arbitration can be refined by applying the criteria 'incidental' and 'structural'. Incidental arbitrators are available on demand, so to speak, and as such have no permanent status. Such arbitrators enjoy a number of important advantages; their *ad hoc* arrangements create flexibility and vitality which make tailor-made settlements possible. Compared to incidental arbitrators, structural arbitrators are 'fixed'.

The organization of the dispute arbitrator is relatively unchanging; the arbitrator's conduct is fairly predictable and, in this sense, the dispute arbitrator is also trustworthy. Structural arbitrators have the significant advantage that the 'closedness' of their organization basically rules out suspicions of favoritism, something which incidental arbitrators face more often; the likelihood of consistent policy is raised; and incentives exist to amass knowledge and expertise.

Other aspects that can be employed to make further distinctions between arbitration modalities include the distinction between judicial and political dispute settlement, the distinction between the material and formal law that applies, and the composition of the arbitrating team, for instance, whether judgement is passed by a board or by a single judge, by experts or by laymen.

3. Changes likely to affect more than one role

The changes likely to affect more than one role can affect either part or all of the infrastructure regime.

Role distribution

The first role-transcending change to be considered concerns the distribution of roles between agents. How many agents will be involved in any given role? In many infrastructures, it used to be the case that while roles were distributed between various actors, these actors were closely linked. Most roles were in public hands, and those private actors who had a role to play maintained close contacts with public authorities; however, a trend is emerging in which roles are being spread over more than one actor. Today many separate agents are jointly responsible for the management and use of infrastructures, where in the past these responsibilities lay with one body. The old monopolists and incumbents are being split up into several agents. The role of supervision is being given closer attention and is being accorded new representatives. New, private agents have entered the scene, who are much less intimately linked with government than used to be the case. This trend may well continue; more private agents may be accorded specific roles. It could even occur that a role is given to several private agents in competition with each other. New governmental and semi-governmental agents are also appearing, and international organizations are unmistakably gaining ground. It is not hard to imagine that these trends will continue.

Another important development is that the roles of owner, manager and service provider are becoming spread across various agents. Where these roles used to be combined in a single body, today they are becoming increasingly separated. The most radical separation would entail that the three roles came into the hands of three entirely different bodies, entirely separate from judicial, economic and staffing points of view.

Whatever form role distribution takes in the future, it will be as well to remember that this form will not be stable. Agents will want to tinker with role distribution, since they have a strategic interest in fulfilling more than one role; after all, the more roles they can control or influence the better; however, while agents will want to fulfil more than one role, others will want to prevent them from doing so, since this weakens their own position. Whatever the outcome of this process, it will remain a dynamic one.

Institutional design

A second role-transcending change is concerned with the institutional design of three players: the legislator, the regulator and the dispute arbitrator can all be composed in different ways. Their institutional design can be arranged along one of two lines, namely private law and public law. The distinction between private and public law is primarily one of principle: it may be said that the public-law organization of a given function, the responsible authorities, are making it clear that this function is of such societal importance that it has to be carried out under public accountability; however, in practice the criterion of public-law and private-law design has little discriminatory power.

In principle, whether the institutional design is along public-law or private-law lines bears little relation to the degree of influence that government can exert, though it should be noted that there are many private-law institutions in which government exerts a predominant influence. There are various strategies that government can employ to exert this influence, ranging from strict subsidy conditions to the appointment of government representatives to executive or regulatory boards. In the event that arrangements for all or part of an infrastructure give preference to regulations derived by the traditional, appropriate channels of democratically legitimized decision-making at the highest level, then the legislator is also the legislator in the formal sense.

The alternative is that the regulations are devised by an organization which forms no part of the organization of public government and which exercises its authority independently. In this respect mention can be made of self-regulation, a method of setting general-application regulations in alternative social fora, usually interest groups. The scope of these regulations is 'general', by which it is meant that they apply equally to all those within a certain, functionally delineated community, for instance, all energy producers. It should be noted that this generality concept differs from the usual interpretation of the generality principle in law. In a state under rule of law, self-regulation comes down to ensuring that decision-making bodies take all relevant interests into appropriate account, for example by creating mechanisms to ensure the 'equality of arms' (since there will always be stronger and weaker parties).

If regulations are framed in formal law then the application of a vertical strategy in the regulatory process is more appropriate than when regulations come about in alternative social fora; given its decision-making structures, in the latter case regulations should be seen rather as the outcome of a negotiation process. In fact, the wish to create regulations in a cooperative procedure of dialogue and negotiation can be a reason for the formal legislator to attribute or delegate powers to alternative social fora.

Like the legislator, the regulator and the dispute arbitrator can also be organized along either public-law or private-law lines. This applies to both regulatory modalities, the sector-specific and the general. If choosing the private-law option, it should be remembered that a certain distance should be retained between the field of supervision and the supervising body to prevent a conflict - or apparent conflict - of interests, for instance by delegating supervision to involved but opposing interests, e.g. energy consumers as the watchdog of distributor tariff observance. As for the dispute arbitrator, there can be a demand for independence, only on condition that a dispute arbitrator endeavors to stay at a formal distance from conflicting parties can the arbitrator expect these parties to accept its findings. A state-appointed judge who comes to a decision on the basis of fixed and public procedural rules represents, in this sense, the most traditional conflict arbitration modality. A special court can be designated for a certain infrastructure, while the ordinary court deals with infrastructure disputes. Another possibility is that contending parties opt to entrust their dispute to a third party, such as an arbitration council. In principle, this third party can be either a general or a sector-specific arbitrator and can operate either on an 'on-demand' or on a permanent, fixed-procedure basis. One reason for appointing an arbitrator in a given area can be to take advantage of the arbitrators special knowledge. Given the technological dimension of infrastructures, it is not inconceivable that the possession of technological expertise, as well as judicial knowledge, will become an important part of the satisfactory settlement of disputes.

European developments

A third important variable influencing the development of infrastructural networks is the development of the European Union: is Europe becoming larger and more imposing, or is it showing increasing regionalization? If the former, then decision-making processes at the European level will become more important; decisions will determine the national policy of member states, which will display increasing unity and similarity. Member states will also endeavor to set up Europe-wide infrastructural networks by linking national networks. Initiatives have already been taken to set up such European networks.

A positive effect of a very large Europe would be that its size creates enough space for individual member states to profile. An increase in the number of member states will promote equality between the participants to European decision-making processes.

An extreme variant of this scenario is that the separate member states are entirely subsumed by a large single Europe, with national governments giving way to an all-embracing European government. Strategic decisions would then be taken only at the European level, and national governments would become no more than the executors of EU policy.

In the alternative developmental option, regionalization, member states attach great importance to their independence and make use of the EU to promote their own national aims. The European Union then exists, but there is much dissent during decision-making processes and EU policy is disjointed and incoherent. Member states each develop their own policies and laws. Differences between member states widen and integration between them becomes more problematic. No Europe-wide infrastructural networks arise; networks remain separate and independent. There are no incentives to develop European networks; the member states become islands in a European sea.

4. End user

A number of general societal developments may be distinguished which principally affect the end user.

The decline of the nation-state

Modern society can be described as the *information society*, one in which technological knowledge and information form the most important sources of power, as opposed to violence or money, as has been the case in the past. In the information society, the development of information technology plays a crucial role; the origins of this new world lay in revolutionary changes in the production process arising from this information technology, i.e. technologies which have shrunk space and time to negligible proportions. This technology effectively underlies a new economic and political order, one which knows no hierarchies but which is based on the capacity for innovation, competition and flexibility (the 'techno-economic paradigm'). Information technology has been able to develop in relative autonomy - in other words, practically undisturbed by any curtailing influences from without - and at great speed. There has been ample commercial incentive to do so, since economic globalization and the associated intensification of competition in numerous markets have become important business factors. The power vacuum left by the loss of classic political ideologies which the postmodern tradition calls 'the end of the big stories', has also been responsible for the lack of guidance or leadership in the development of information technology.

The coming of the information age has brought about fundamental changes in society, in the sense that the 'classic' basic institutions have been changed beyond recognition - particularly in the areas of social relations, power systems and ideology. The 'classic' power structure that used to hold the world together has fallen apart; the nation-state has lost practically all its sovereignty. Today's society is organized along global lines, and to an important degree it is split up along economic (market segment) lines. The scale of society is limited, in the sense that the welfare state is being dismantled. The historical compromise between capital, labor and state has broken down. All over the world, institutions are engaged in a continuous battle for control over knowledge and information. Strongly innovative multinational companies have the best chance of winning these battles (the so-called *powershift*: the growth of a new power elite). This is because global organizations which work as a network of flexible subsidiaries are in a position to circulate this information at lightning speed and to (re-)configure their capital and labor accordingly, anywhere in the world, and at any moment, in the most profitable way.

In the information society not much is left of the 'citizen', in the classic sense of 'citizenship' which refers back to old presuppositions about the nation-state. In the modern political and economic world, the individual plays no more than a modest role as a consumer, an actor who can choose between a certain number of pre-configured 'lifestyles'. Parallel with the decline of the nation-state, powerful new sources of identity formation have arisen: the World Wildlife Fund, Greenpeace, Amnesty International, Medecins Sans Frontières, and so on. The logic of the information society has, as it were, invoked its own resistance.

The new social movements have also adopted the networked form, and they also employ information technology to communicate between nodes in an unlimited whole that, in principle, can reproduce itself without limit. Individuals unable to link up with these networks can become socially marginalized; they can either passively accept this, the 'hermit option', or they can identify with groups that take explicit exception to this kind of social alienation e.g. fundamentalist religious groups, or neo-nazis.

In urban development terms, the decline of the nation-state has important consequences, since centralized zonal planning on behalf of the public interest therefore also falls away; supply and demand then determine the purpose to which land is put, its price, the cost of the buildings erected and therefore the category of owner. In the absence of centralized direction, the information society becomes, in planning terms, an 'urban jungle' in which every green space will eventually be built on by whoever has the financial means to do so, while older buildings in inner cities are left to decay, forcing those able to afford better to seek it elsewhere. In principle the concept of 'environment', or indeed of the *city* per se, plays no role in this weighing up of interests. Since the traditional structures of descent, family, fatherland etc. have fallen apart, modern man now directs his desire for identification principally towards the company of which he is a part ('corporate identity'). Fixed residential addresses are a thing of the past; the new company man follows his multinational employer 'wherever it goes'.

End user attitudes

In connection with, and partly in reaction to, the decline in social structures and cohesion, individuals can adopt one of two attitudes with regard to the goods and services made available, whether directly or indirectly, by infrastructures. The consumer of the future can be selfish and egocentric, directing his activities towards the gratification of his own needs and the search for personal happiness. This 'hedonistic' consumer recognizes few moral boundaries, in the sense that he considers his own ease and comfort to be of the greatest importance. The only limit he recognizes is the budget available; in principle, lack of money forms the only reason for restraining his urge to consume. The social costs of hedonistic behavior are greatest where individual pleasure is at the expense of collective pleasure - for instance, the private consumption of electricity versus the environmental pollution generated by the production of electricity.

The consumer of the future may however, in full appreciation of the threats which community and public spirit are facing, choose to display a social conscience, in the sense that her choices bear witness to a strong public conscience. Particularly in view of its potential as an alternative source of identity formation, 'green consumers' can develop and evolve within the framework of the global 'green movement'. The chief concern of this movement is that humanity deals as wisely as possible with scarce natural energy resources, particularly by reducing energy use to a minimum. This is advocated by promoting the virtues of a sober and frugal lifestyle, by encouraging households to produce their own energy or use 'green energy', and by establishing special hallmarks and certification systems, i.e. 'green electricians', 'green manufacturers', 'green contractors', 'green packaging'. The green movement could potentially grow into an alternative global community with its own shops, newspapers, schools, and other facilities.

5. In conclusion

Table 1 below shows the possible variations per infrastructure regime role, together with the other changes that have been discussed. An arrow indicates the direction of the possible shift. These shifts can vary between infrastructure sectors. A single arrow indicates the most likely direction of shift.

The large number of variations means that the construction of administrative and judicial scenarios is a complex business; after all, the number of potential scenarios is huge, with large numbers of possibilities lying between the extremes indicated; however, it is possible to indicate a smaller number of logical, extremely probable combinations. For instance, a technological and commercial culture will not be able to develop at great speed without judicial facilitation, perhaps by leaving the design of contracts between various market players open to their own determination rather than unilaterally presenting a required design.

Moreover, a number of other combinations are impossible, since certain extremes are mutually exclusive. For instance, the information society is by its very nature 'organized' on a global scale, and will therefore have little relation to the course of European development. The same is true of its relationship to judicial arrangements, whose institutional source remains the national legislator. The large number of possible and impossible combinations will not be discussed further here; this overview is intended principally as a toolkit which scenario-builders can employ as they see fit.

Table 1. *Overview of the possible variation per infrastructure regime role*

Infrastructure regime roles

Owner	concentrated with one agent	⇔	spread across more than one agent		
Manager	technical / administrative culture	⇒	technical / commercial culture		
Service provider	regional scope	⇔	national scope	⇔	international scope
	regional market	⇒	national market	⇒	international market
End user	prescribed relationship between end user and service provider	⇒	private-law contract between end user and service provider		
	supply-driven	⇒	demand-driven		
Legislator	unilateral legislation	⇒ ⇒	negotiated rulemaking (whether or not formalized)		
Regulator	sector-specific national	⇔ ⇒	Generic international		
Dispute arbitrator	sector-specific	⇔	general		
	ad hoc	⇒	structural		

An Engineering Perspective on the Design and Control of Infrastructures

Explorations into a Generic Approach to Infrastructure Scenario Analysis

Margot P.C. Weijnen*, Okko H. Bosgra**

** Delft University of Technology*
Faculty of Technology, Policy and Management
Industrial Systems and Environmental Resources group

*** Delft University of Technology*
Faculty of Mechanical Engineering
Mechanical Engineering Systems and Control group

Abstract

In this contribution infrastructures are explored as a subject for scientific research from a design and control engineering perspective. As physical infrastructures are technical artifacts, they are technology intensive by nature. New technologies have dramatically changed infrastructure networks and their subsystems in the past, and will continue to do so in the future. In order to unravel the complex interactions between technology, economy and society that govern the selection of technologies for infrastructure development and the design of the system structure, a deeper understanding of infrastructures and their evolutionary behavior is needed. A systems engineering approach is considered to be an effective approach to arrive at a formal description of infrastructures in generic engineering terms, enabling a comparative analysis between different infrastructures. On the basis of an exploration of technological, social and economic trends, the core dilemmas for infrastructure design and control engineering in the next decades are identified. Finally, a set of working hypotheses is formulated and their research implications explored, on the basis of which a research plan is proposed that will enable a deeper understanding of those generic factors and phenomena that are critical to the design and operation of infrastructure systems.

1. Introduction

Infrastructures form the backbone of the economy and society, they have a significant influence on the design of our physical environment and our way of living. The infrastructures providing us with the public utilities of energy supply, water supply, (tele)communications, mobility of persons and goods, collection of wastes and waste water, are particularly vital to the well-functioning of the economy and our households, and contribute greatly to our individual health and well-being. The availability of reliable infrastructural services is generally recognized to be a *conditio sine qua non* for the economic viability and social stability of a modern society.

The crucial role of infrastructures in our society originates from their enabling role towards other sectors of the economy. Infrastructures enable the provision of basic services that are needed at the basis of almost any value adding chain in the economy. Without these basic services, social and economic development are severely hampered. This is part of the reason why infrastructure planning, construction and operation, in most countries, used to belong to the public domain. This situation, however, is rapidly changing. In most industrialized countries, public utility functions are being privatized, and utility markets liberated. These changes seem to favor, as well as to be enabled by, certain types of technological innovations.

The assumed interaction between, on the one side, the regulatory and economic framework of infrastructure sectors, and, on the other side, the technical alternatives selected, is the central motivation for the research project introduced in this paper and the interfaculty research program it is part of. Whereas a generic research perspective to infrastructures is known to be applied in the fields of public management and infrastructure economics, technological research into infrastructures has so far been highly sector-specific, dealing with specific physical infrastructures and their components. In this paper it is explored how infrastructure design and control engineering strategies may be dealt with from a generic perspective.

2. Motivation from a social and economic perspective

In the current processes of change towards market liberalization, privatization of public utility companies and sector re-regulation, new technologies come into the picture. Given the high capital intensity of physical infrastructures and their long physical durability, it is of the utmost importance that the investment decisions being made now and in the near future can be thoroughly evaluated on their long term effects. Infrastructures have a strong direct influence on our way of living. More indirectly, they may change our way of living through their influence on the natural environment, as a consequence of the way they make use of natural resources, the way spent utilities and other emissions are discharged, or simply through the space occupied by infrastructural installations. Many infrastructures, in particular the surface transport infrastructures, profoundly change the landscape and largely determine the degrees of freedom for physical planning in the future. Thus the question arises: How flexible are the physical infrastructures with respect to their ability to be adapted to future changes, foreseen and foreseeable, in the social and economic environment? How well can technological innovations be accommodated? In other words, how can we make sure that the infrastructure facilities keep pace with rapid economic and social developments, rather than blocking opportunities? Given the internationalization of the economy, the span of control of national governments in steering infrastructure development and controlling infrastructure regulation seems to be steadily shrinking, and the driving forces for economic and social changes are increasingly generated in an international playing field.

3. Motivation from an engineering perspective

The public utility services mentioned in the introduction of this paper: mobility of persons and goods, telecommunications, energy supply, water supply (for drinking and other household and industrial functions), waste and waste water collection, all rely on the existence and well-functioning of physical infrastructures. These physical infrastructures are technical artifacts, hardware systems designed and operated by engineering professionals. The mere fact that physical infrastructures are engineering products, and products of an engineering culture, however, is only part of the motivation to study infrastructures from a generic engineering perspective. The infrastructure sectors are increasingly technology intensive, most notably so the telecommunications sector. Technological innovations may gradually or radically change an infrastructure sector:

New technologies may yield new infrastructures that consequently generate new services: e.g., glass fiber networks for ISDN services versus the traditional copper cable network for traditional telephone and fax services. New technologies may thus create competition between networks, and may lead to existing infrastructures becoming obsolete.

New technologies may undermine the natural monopoly position of existing infrastructures: e.g., wireless telecommunication technology has created an attractive and affordable alternative for the copper cable network based telecommunication services.

New technologies may equip existing infrastructures with new functionalities, so that an infrastructure once designed for one specific function becomes multi-functional: e.g., electricity cable networks may be put to additional functions such as telephone services, internet services, multi-media services. This phenomenon of different utility functions converging on the same, multi-functional, physical infrastructure, creates new opportunities for traditional utility companies to compete in the market.

New technologies may cause a restructuring of the physical network and the corresponding utility market: e.g., the enormous changes brought about in the economy of scale of power and heat supply (e.g., industrial co-generation), induce a shift from a largely centralized power supply system to a much more distributed supply system. New players can thus enter the market. Large scale users already have a choice between dependence on the grid and autonomous power supply, and even small scale consumers are envisaged to have that choice in the decades to come (depending on the price-performance ratio of e.g., micro-co-generation systems and PV-systems being developed). The emergence of highly cost-effective, distributed supply options in the power sector has created a need for third party access (TPA) to be organized on the electricity transport and distribution network.

New technologies may be introduced when existing infrastructures need to be expanded (e.g., to serve new residential areas and new industrial areas being developed), and may locally replace parts of an existing infrastructure in the course of de-bottlenecking and maintenance programs. This phenomenon is referred to as system integration: due to different functional and operational specifications, the integration of new subsystems in an existing infrastructure often turns out to be a complicated matter. The technical complexity of system integration may only be aggravated by barriers in legislation, regulation, administration, organization and management, etc.

New technologies originating from the control engineering field may create new possibilities and opportunities e.g., to improve the reliability of the utility supply, to improve the safety of the system for operators and users, to enhance infrastructure capacity, or to control the allocation of scarce infrastructure capacity to different providers and users.

The obvious impact of technological innovations on the operation and development of existing infrastructures and on infrastructure innovation, justifies deeper research into infrastructures from an engineering perspective. The importance of a *generic* research perspective to infrastructures, however, in spite of the pronounced differences at the level of their technical components and subsystems, has not been justified yet. The assumption that a generic engineering perspective to infrastructures is meaningful, is based on the observation that infrastructures, notwithstanding the pronounced differences at the level of their technical components and subsystems, share a large number of characteristics at the aggregated system level, i.e. at the network level. The working hypothesis of this research is that a set of meaningful analogies and commonalities can be identified, that will generate insight into the behavior of infrastructures as aggregated systems, and will justify the development of generic design and control strategies for infrastructures.

4. Infrastructures and infrasystems - working definitions

In spite of the abundance of infrastructure research in some disciplines, particularly in the field of economics, a clear definition of infrastructures appears to be lacking (see paper by Firth, Boersma and Melody). Infrastructure definitions in the technical literature generally focus on the immovables, i.e. the installations and interconnections that make up the network connecting suppliers and users. This network does not necessarily have a fixed character in the dimensions of time and space, as in the case of the infrastructure for wireless telecommunications. Certain infrastructures are useless without carriers, e.g., road and railway systems. All kinds of auxiliary equipment as well as human intervention are needed to make an infrastructure system perform its function(s). Considering the heterogeneity of infrastructures with respect to the type of products and services they provide as well as the type of technology involved, an attempt is made to introduce a general definition, or at least a general working definition of infrastructures, applicable to all infrastructures which will be studied in the TU Delft interfaculty research programme for the design and management of infrastructures: the infrastructures for transport of persons and goods, telecommunications, water supply, energy supply, waste (and waste water) collection and disposal.

Physical infrastructures

The basic facilities, equipment, and installations needed to provide the utility products and services crucial for the growth and functioning of an economy, community or organization. The collection of basic facilities, equipment, and installations refers both to the hardware of the infrastructure (needed to fulfill the basic transport, distribution, storage and processing functions of the infrastructure) and to the safety and control engineering systems (not only hardware devices, but including operational procedures, organization and management) needed to make the system function according to its functional specifications. The first and foremost functional specification of a physical infrastructure is that the system is able to match supply and demand of the utility product and/or services, on a range of time scales:

- *the evolutionary time scale, i.e. a time scale varying between decades and years, depending on the type of infrastructure and the rate of system innovation*
- *a range of operational time scales, with respect to e.g., seasonal and daily demand fluctuations.*

The hardware of an infrastructure not only includes the stationary parts, such as installations, cables and pipelines, but also the moving parts, such as the carriers needed to make use of the infrastructures for transport of persons and freight. It should also be emphasized that the definition of a physical infrastructure includes both hardware and software, the latter consisting both of technical and human 'software', as both are needed to make the infrastructure function.

The previous sections of this paper evidently lead to the conclusion that physical infrastructures cannot be developed as isolated systems, considering their strong interaction with the economic, social and regulatory environment. This interaction is not a one way interaction. On the one side, technological innovations may give rise to the development of new infrastructures or change the behavior of existing infrastructures, thus inducing a need for market restructuring, new legislation, et cetera. On the other side, new economic and regulatory frameworks will favor certain technological options and certain infrastructure configurations over other alternatives. As the subject of the interfaculty research program on the design and management of infrastructures includes this interaction between physical infrastructures and their environment, we re-introduce the integrative concept of infrasystems:

Infrasystem

A physical infrastructure in interaction with its physical, economic and social environment. An infrasystem is the integrated system of a physical infrastructure network, with all its physical and organizational attributes (e.g., carriers, control and safety systems, operational procedures, organization and management structures), the network of actors involved, and the rules (procedures, administrative arrangements) they apply intra- and inter-organizationally to make the system function in its institutional context.

The concept of infrasystems as complex socio-technical systems was introduced by Thomas Hughes [20] in his analysis of the development of the electricity infrastructure. It is the ambition of the Delft Interfaculty Research Center for the Design and Management of Infrastructures to unravel the complex interactions between technological, economic, administrative and organizational factors in infrasystems, and to do so not only from a curiosity driven analytical perspective, but also from the synthesis perspective, aimed at the development of strategies, methods and tools which can assist policy makers and utility companies in the planning, design and management of infrastructures.

The authors of this paper are focused on unraveling the technical complexity of infrastructures and infrasystems, and to make these insights operational for design and control engineering purposes. It is the aim of the authors of this paper to explore the technical characteristics of infrastructures and infrasystems and thus derive a systematic research framework needed to analyze the evolutionary behavior of infrasystems and to develop design strategies and tools to deal with the technical complexity of system integration questions as well as with the design of new infrasystems.

5. Analogies and commonalities between physical infrastructures

5.1 Analytical framework

So far, quite a few analogies and commonalities between the physical infrastructures in different public utility sectors, as well as between the infrasystems in different sectors, have been identified. In this section a preliminary list of commonalities and analogies will be drawn up. The purpose of this list is, rather than trying to be exhaustive, to make a preliminary attempt at defining a systematic framework, enabling a comparative analysis of the infrastructure sectors mentioned as the research objects of this paper.

In the previous sections, physical infrastructures have been found to share the following general features:

- high capital intensity
- long physical lifetime (ranging from centuries to decades)
- hardware dependent
- network character
- increasingly technology intensive
- public utility nature of products and services.

A deeper comparative analysis of the properties of physical infrastructures will be approached from two angles:

Phenomenologically:

How can the infrastructures, as they are, be described and characterized in such a way that a comparative analysis can be made? In this part of the work the focus is on a functional characterization of infrastructure systems and subsystems and a characterization of network morphology. At which aggregation levels in the system do we find meaningful analogies between different infrastructures? To what conclusions do these findings lead us regarding the potential for generic design strategies and tools?

Evolutionary:

How were the current infrastructures brought into being? How can we characterize their historic development and do we find meaningful analogies in their evolution patterns? Can we formally define the starting point of the evolution of a specific infrastructure? What will the infrastructures underpinning the economy and society in 2030 look like if we simply extrapolate the evolutionary patterns of the infrastructures we know today?

The first angle will help us to test the working definitions of infrastructures and infrasystems, as without a formal definition of the research subject, its evolutionary behavior can not be analyzed. In the sequel of this section, the phenomenological analysis will not, as yet, focus on network morphology, but be restricted to a first attempt at a generic functional characterization of infrastructures. The subsequent analysis of the long term dynamics of infrastructures, the second research angle, might give us a clue as to the bandwidth for control of infrastructure development as a function of its evolutionary phase. The operational control of infrastructures to match supply and demand on an hour to hour and day to day basis is not examined in this paper.

5.2 Characteristic functions of infrastructures

A number of functions can be distinguished in all five infrastructure systems subject to this study which, combined, enable the satisfaction of a basic need that is vital to the functioning of other sectors of the economy and society:

1. extraction/production of resources
2. conversion/processing of resources
3. transportation, storage and distribution of untreated and/or processed resources
4. end-use conversion and disposal by the user

Energy infrastructures

Energy infrastructures basically satisfy the needs for light, heating or cooling, mechanical and electrical power needed to supply other functions:

1. petroleum and natural gas production, coal mining, uranium ore mining (or extraction of uranium from other sources)
2. processing of resource (concentration, mixing, other feed preparations) and conversion to power and heat
3. transportation and storage of oil, gas, coal, etc.; transportation and distribution of ready-to-use electricity, gas and heat
4. end-use conversion of electricity to light, heat, mechanical work, etc.; end-use conversion of gas to space heating , hot water, etc.

Water infrastructures

Water infrastructures satisfy the basic human needs of drinking and washing water supply, in addition to household, industrial and agricultural uses as energy carrier, nutrient carrier, industrial solvent, etc.

1. ground water extraction and surface water intake
2. purification of raw water to drinking water quality or other desired quality specifications
3. transportation and storage of raw water; transportation, storage and distribution of ready-to-use water (drinking water quality or other quality level specified by e.g., industrial consumer)
4. end-use of water (drinking, cooking, washing, flushing, etc.) by consumer and disposal as waste water (possibly after industrial waste water treatment) to sewage system, local waste water storage system or surface water

Waste infrastructures

Waste infrastructures satisfy the needs for a clean and safe living environment and protection of the natural environment. In addition waste infrastructures may serve the purpose of making secondary material and energy resources available for reuse:

1. extraction of waste, i.e. collection of waste from waste producer, waste water collection through sewage system
2. separation into fractions for separate treatment or direct treatment of waste (water) to produce useful products or environmentally harmless waste products
3. transportation and storage of raw wastes; transportation and storage of treated wastes; distribution of useful products (e.g., secondary raw materials)
4. end-use (reuse) of separated components; end-use (reuse) of useful waste products; disposal of harmless waste fractions into the environment; final disposal of hazardous wastes and waste fractions in safe disposal sites

Transport infrastructures

Transport infrastructures (road, railway, waterway and airway systems, including the carriers needed to effectuate transport) satisfy the needs for mobility of people and goods

1. the transport infrastructure does not contain a specific extraction/production function other than the production of transport fuels needed to move carriers over transport lines
2. collection of people/goods for embarkment/'loading' in discrete units (containers, carriers)
3. transportation of people/goods in carriers over roads, railroads, waterways or airways (carriers being cars/busses, trains, ships and aircraft) and distribution to final destination, with possible interconnections between different transport modalities (and possibly intermediate unloading, storage and reloading)
4. disembarking/unloading at final destination

Telecommunication infrastructures

Telecommunication infrastructures satisfy the basic human need of communication in the case where the parties wanting to communicate are physically separated as well as the basic economic need for information and data exchange between individuals, companies and sectors in a largely internationalized economy

1. emission of communication signal

2. processing of communication signal into transportable signal
3. signal transportation and distribution to intended receiver
4. signal decoding at site of receiver

From this preliminary attempt into a functional decomposition of the physical infrastructures in the five sectors, it is evident that the infrastructures do not fit perfectly into a generic functional framework. In cases 1-3, the central functions of the infrastructures can be described as bringing about the conversion of physical matter into a form ready-for-use and transporting the latter to the consumer. In cases 4 and 5, however, only the transportation and distribution functions matter, e.g., the production of goods (or people!) is not considered part of the transport infrastructure. Conversion in these cases, if it occurs (e.g., the conversion of transport fuels), is only relevant at a lower system level, supporting the transportation and distribution function.

Unlike cases 1-4, infrastructure sector 5 is not concerned with the transport and distribution of physical matter (or a physical matter derived product), but with non-tangible communication and information.

In case 1, neither the production function nor the end-use conversion are generally considered to be part of the infrastructure. The power infrastructure includes the dedicated processing of fission and fossil fuels to produce power and heat as well as the transportation and distribution of power to the end-consumers. The natural gas infrastructure includes the storage of natural gas, its treatment to establish certain quality specifications and the storage, transportation and distribution of on-spec natural gas. In case 2, the production function is considered part of the infrastructure, as ownership and/or management of the raw water supply are generally in the same hands as the subsequent treatment and the transportation and distribution of the product to the consumer. In case 3, a clear picture of an infrastructure is lacking, except in the case of a sewage system for collection of waste water (municipal waste water, mildly contaminated or pretreated industrial waste water) and rain water run-off. For other wastes various collection systems have been set up, each followed by different transportation, treatment, reuse and disposal paths. All wastes other than waste water use the traditional transport infrastructures to be transferred from the waste production site to treatment and disposal sites. The transport infrastructures, case 4, are for the purpose of comparison with cases 1-3 considered as consisting of the combined system of transport ways, carriers and interconnections between transport modalities. The telecommunication infrastructures, in case 5, stretch from the signal emission and transformation devices through the transportation and distribution facilities to the signal reception and decoding devices. In all five infrastructure categories the control and management systems involved are considered part of the infrastructure.

The transport and distribution functions are characteristic functions of all physical infrastructures. *All physical infrastructures thus have the fundamental character of a transport and distribution network.* This network character does not necessarily imply the existence of direct physical connections between supplier and user through cables or pipelines. A wireless telecommunication infrastructure can also be described as a network. *In all physical infrastructures, one or several storage functions can be identified.* This statement holds for the energy infrastructure when we consider it as an aggregate system composed of specific infrastructures for electricity, gas, transport and heating fuels, and heat (e.g., hot water or steam). In the specific case of the electricity infrastructure, however, storage is a bottleneck due to the lack of technological options for large scale storage of electricity. Large scale storage of excess supply can only be established through (re-)conversion to potential energy or chemical bonding energy.

In many infrastructures, the nodes in the network contain processing facilities, where primary natural resources or secondary resources (waste, waste water) are chemically and/or physically converted. In the infrastructures for transport of persons and goods, the nodes in the network have the function of conveying persons or freight from one carrier to another. If the transport infrastructure is seen as an integrated infrastructure system composed of the subsystems road, railway, waterway and airway transport infrastructures, the nodes also represent the function of inter-modal connections: changing from one transport modality to another. *The processing function in the nodes of infrastructural networks, should thus be interpreted in generic terms as: changing the nature of either the matter being transported, or its carrier, or its transport modality.*

In addition to the functional characterization of infrastructures, many other characteristics must be examined, to arrive at a classification of infrastructures on the basis of meaningful analogies in structure and behavior. The most prominent of these characteristics is network morphology, which will be the priority research question. A preliminary list of infrastructure characteristics to support a comparative analysis between infrastructures is given below:

- functions of infrastructure, subsystems and elements
- mono- or multi-functionality
- type of input (specific type and quality versus multiple types and quality variation)
- type of output (specific type and quality versus product and quality variation)
- network morphology
- intensity of use of different links in the network, as a function of time and place, type of supplier or user
- user group topology

- supplier group topology
- transport mode (continuous or discrete)
- transport velocity
- transport direction (uni- or bi-directional)
- geographical scale of the network.

In addition to this list of characteristics to be examined in a systematic comparative analysis of infrastructures, future research will aim at identifying the set of design and operational performance criteria applied to infrastructure networks, their subsystems and components. In an analysis of infrastructure evolution, special attention will be paid to possible changes in the set of design criteria (shifting dominance of certain criteria, addition of new criteria), either as a result of strong influences from the social and economic environment in certain historic periods, or as a function of the evolutionary phase of the infrastructure itself.

5.3 Evolution of infrastructures

At first glance, all infrastructures seem to have followed a similar development path:

1. an infrastructureless stage of local supply and use
2. a stage when the traditional transport infrastructures (roads, waterways, railways, airways) were used to supply fuel (wood, peat) and enable communication (using dedicated carriers such as carrier pigeons and mail coaches)
3. a stage when dedicated infrastructures evolved, mostly through private initiative, starting at the level of local networks (local telephone networks) or local connections (railroad)
4. a stage of network expansion to a larger geographical scale, when the local networks were interconnected, followed by expansion of the network to rural areas, as the public utility character of the service was recognized and the infrastructure was adopted by the public domain. The stage of network expansion to regional, nation-wide or international transport and distribution networks was often stimulated by the emergence of new technologies that enabled economies of scale.
5. a stage of network intensification, with an increasing density of interconnections and end-user connections, and infrastructure capacity enhancement (e.g., through faster carriers or the introduction of more intelligent control systems).
6. at this moment a number of technology innovation trends seems to trigger new stages of infrastructure evolution. Some of these trends seem to work out analogously in different infrastructure sectors, whereas others seem to be of a more sector-specific nature. These new trends will be discussed in section 6.

The waste infrastructure still seems to be in phase 2, with the exception of a dedicated infrastructure brought into being for municipal and (some) industrial waste water, most of which is collected through sewage systems and cleaned before being discharged to the environment. The collection of most types of wastes, however, still depends on the surface transport infrastructure, with dedicated waste carriers being used for the transport of wastes to dedicated processing and disposal sites.

Waste water infrastructure development did not start until the value of its collection, cleaning and safe discharge were recognized as crucial for public health and the quality of the environment. The same argument holds for the installation of dedicated waste processing facilities and disposal sites. As the value of waste as a source of secondary materials is bound to increase with decreasing abundance of primary resources, and as the deterioration of the physical environment by inadequate waste treatment and disposal and even the space occupied by waste disposal is becoming more and more unacceptable in densely populated areas, the future development of more advanced, dedicated waste collection and processing infrastructures is not unlikely.

The value aspect is of evident importance in the ongoing discussions on the design of infrastructures for new residential and industrial areas. Residual heat resulting from high temperature industrial processes and thermal power plants, traditionally considered a waste product to be discharged to surface water and the atmosphere, is now being considered as a potential resource for district heating and other purposes. Provided the costs to install and operate a distribution infrastructure for residual heat and other (water and energy) utility wastes are competitive, spent utilities may replace the supply of 'fresh' utilities (generally derived from primary resources) through the traditional public infrastructure. Considering the generally low value of wastes and spent utilities, their re-use is only economically justifiable if the distance to the user(s) is short enough to make the benefits of using cheap 'wastes' outweigh the costs of installing a dedicated transportation and distribution network.

The current situation of the waste infrastructure seems comparable to the historic situation when the transport infrastructure underpinned the communication infrastructure (in times when messages were carried by dedicated messengers or when mail was transported by mail coaches) and the energy infrastructure (transport of wood, peat and coal to the end-user). The classical transport infrastructures, such as the natural waterways, man-made waterways and old road systems (such as the Roman road system) seem to have had a major influence on urbanization and industrialization patterns, in combination with the availability of natural resources. One of the reasons why the degree of urbanization of the Netherlands in the 17th century was far more advanced than in many surrounding countries, is speculated to be that peat, amply available as a high energy density fuel, could easily be shipped into the cities through the intricate maze of Dutch waterways and canals.

Most infrastructures, however, have evolved from a situation of dedicated carriers using the road, railroad and waterways infrastructures to the modern situation of a dedicated infrastructure for transport and distribution of one specific utility service. In the Netherlands the quality and availability of energy and water infrastructures have for decades been so high that many economic activities rely totally on these basic facilities, as the need for local backup facilities in case of infrastructure failure has been negligible. As a result, the transport infrastructure system is still the major backup in case of failure of infrastructures for drinking water and energy. When the drinking water infrastructure does not function consumers depend on bottled mineral water (to be purchased and transported to their homes by themselves) or on tank loads of drinking water (to be supplied and locally distributed by the water utility company). Similarly, when the energy infrastructures do not function, consumers can revert to buying bottled gas or to installing their own generators and fuel storage facilities, unless the energy utility company continues to service its customers by carrying a generator and fuel supply to the users site. All these backup options require the use of the classic transport infrastructures, the road infrastructure in particular.

Within the surface transport infrastructures the trend is towards specialization, and the development of dedicated new infrastructures is still proceeding. Within the road infrastructure system, special strips of road are designated for dedicated use by freight traffic, public bus traffic, and car-poolers. New railway lines are being designed for dedicated use by high speed passenger trains, others for dedicated use by freight trains.

Now the question can be raised as to what justified the construction and operation of dedicated infrastructures for specific utility services instead of the use of dedicated carriers on the traditional transport infrastructures. The answer is probably found in a combination of added value through improved reliability and quality of service, cost reduction through upscaling of hardware facilities, and the enforcement of a service obligation on the public utility companies. In economically developed and civilized nations, public utility services are reliable and guaranteed access is available to every citizen at an affordable price.

6. Technological extremes - Dilemmas for infrastructure development from a design and control engineering perspective

Throughout the evolution of the existing physical infrastructures, technological innovation has played a role in e.g,

- the intensification of processing facilities and transport functions (increasing throughput, e.g., through larger capacity and speed of carriers)
- improving the quality and the reliability of service
- protecting public health and safety, and the natural environment
- improved matching of supply and demand, particularly through increasing use of advanced control engineering systems and ICT.

The current changes towards liberalized markets for infrastructure products and services, privatization of public utility functions and the new regulatory regimes, will affect the selection of technological alternatives for the planning of new infrastructure systems and for incremental changes to existing infrastructure systems. The technological extremes of the infrastructure playing field in the decades to come will be explored in the following section of the paper.

Technological innovation is partly an autonomous process, curiosity-driven and thus catalyzed by its own progress. When new technologies become available on the market, enabling cheaper utility supply, product and service differentiation or new services, new choices are created for consumers, administrators and policy makers. As mentioned previously, however, the interaction between technology and the socio-economic environment is not a one way interaction. Technology development itself is also driven by social and economic trends. The current changes towards liberalized markets for infrastructure products and services, privatization of public utility functions and the new regulatory regimes, will affect the selection of technological alternatives for the planning of new infrastructure systems and for incremental changes to existing infrastructure systems.

The method followed does not conform to a regular scenario analysis as described by Thissen in this volume. Rather, a number of trends is identified. On the one side, technological trends of a more or less generic nature, i.e. apparent in a number of different infrastructure sectors, are examined for their consequences for future infrastructure design and control. On the other side, a number of dominant societal and economic trends is examined to see which type of technologies or technological innovations are favored by these trends. Thus an attempt is made to analyze the technological extremes to which the infrastructures may be stretched in different economic and social development scenarios for the year 2030, to provide insight into how current investment decisions might affect the future adaptability of infrastructures to changes in their environment.

6.1 Technology innovation trends

Distributed utility supply

In many infrastructure sectors new technologies have emerged or are emerging that enable cost-effective utility supply at a relatively small scale. In the electricity sector, this has caused a proliferation of industrial co-generation units at the sites of large industrial consumers, mainly process industries, which have large demands for heat rather than electricity. It is envisaged that micro-units for co-generation may become a cost-attractive option for household consumers in the next decade. Wind turbine technology is already cost-effective, and photovoltaic technology may become so in the decades to come. In addition, other decentralized systems, e.g., small scale gasification units for biowastes, may find their way to the energy market. This strong trend towards decentralization of utility supply down to the level of the individual user is also observed in other infrastructure sectors. In passenger transport, the preference for the use of a private car rather than public transportation can be compared with a preference for small scale, autonomous utility supply. In the drinking water infrastructure, new technologies, membrane separation technology in particular, may also lead to decentralization of utility supply in the near future. In the waste water infrastructure the trend towards decentralization is obvious, since most large scale industrial 'consumers' (producers of industrial waste water) have long installed their own waste water treatment facilities. In telecommunication, the increasing preference for mobile telephony rather than depending on the fixed network, again fits in the same trend towards decentralization. The downside of this trend is obvious: the large scale centralized facilities which are part of the 'old' network, are facing expensive overcapacity.

Multi-functionality

A still very young trend is concerned with the convergence of different functions on the same infrastructure, thus converting a previously mono-functional infrastructure in a truly multi-functional network. The most convincing example of multi-functionality is the use of the electricity cable for voice and data communication, and for cable television. As most infrastructures were purposely designed for mono-functionality, the existing system may not allow for any new functions. An option derived from the idea of multi-functionality is the idea to combine functions which are essentially supplied through different infrastructures, e.g., by drawing a glass fiber cable through a gas pipeline. Another idea related to multi-functionality is the use of information and communication technology to get information from the end-user system to the supplier, thus enabling the supplier to assist the consumer e.g., in managing his electricity costs by tele-starting certain pieces of end-user equipment at times of excess supply, when the price of electricity is lower. Such tele-control might be considered an option for the future for water using household appliances as well.

Competing networks

Particularly in the information and communication sector new technologies give rise to entirely new networks being installed, competing with the plain old telephone system, POTS. The new ISDN, ADSL and wireless telecommunication networks offer new services and a higher quality of the traditional utility service. The telecommunication sector so far seems to be the only one where new networks are cost-effective, and where thus effective competition between networks can be established. Another example is the installment of a new gas transport and distribution network for the supply of a different gas quality (higher calorific value) at a competitive price to a group of mainly large scale industrial users. In other infrastructure sectors, e.g., in the electricity and water sector, competing networks may be established at the local level, relying on relatively small scale decentralized utility supply systems.

Penetration of information and communication technology

The revolutionary development of information technologies in the past decades has resulted in their penetration into every sector of the economy, including the infrastructure sectors, and in many households. Combined with communication technologies an entirely new world of information and communication services has evolved. A high added value is particularly established in new software and specialty services (tele-banking, tele-shopping) rather than in the physical hardware (cables, chips, satellites, etc.) underpinning these services. The use of information and communication technology opens entirely new possibilities for the operation of infrastructural networks, including intensive, on-line supplier-client relationships: e.g., with the use of sensors on user equipment, maintenance requirements may be tele-monitored and equipment may be tele-started at times when e.g., cheap power is available. It is evident that information and communication technology already plays an important role in infrastructure capacity management. Moreover, advanced information and communication technology is an enabling technology to make a market work efficiently (spot markets, power exchange, Third Party Access on the network).

6.2 Societal and economic trends

Individualization of society

As marketeers have already learned, consumers cannot be easily categorized anymore in groups with a more or less consistent consumption pattern. The individualization of society is apparent from consumers who identify with different social groups for different aspects of their life and consumption behavior. In the consumer product markets, this situation has given rise to an enormous differentiation of products and services, enabling consumers to express their individuality through their consumption behavior. The public utility sectors, however, traditionally provide only a limited number of products and services. For the captive consumer hardly any choice used to be available other than being connected or disconnected to the particular utility grid. Tariff differentiation also used to be very limited for captive consumers. In a free market situation, however, with a choice between utility providers, an emancipated consumer is expected to select provider and product critically on the basis of product quality, quality of service and price/quality ratio.

In the liberalized telecommunications market this has already resulted in an enormous variety of tariffs and subscription charges. With a view to the imminent liberalization of other utility markets e.g., the electricity market, utility companies are now embarking on product, product quality, service level and tariff differentiation. Making the utmost use of information technology is a must to design a sensible and profitable differentiated tariff system, as the company needs detailed insight into the usage patterns of its clients. One of the new products being offered by some power distribution companies is 'green' electricity. New services offered include e.g., the resumption of power supply within an agreed time interval, should a power cut occur. Water distribution companies have, in some industrial as well as residential areas, installed facilities to distribute B-grade water for applications requiring less than drinking water quality.

Sustainable development

It is evident that part of the recent innovations in the supply of power and water meet the widely accepted need for a sustainable development. In spite of the unclear definition of sustainable development, the use of renewable materials and energy, and the reuse of wastes and spent utilities are generally agreed to contribute to sustainable development. In the waste sector, this has resulted in separate collection systems and dedicated treatment facilities for different types of wastes, in fact, in the emergence of waste infrastructures. The costs of discarded product removal, reuse of parts and materials and safe residue disposal is included in the price of more and more products, such as cars and refrigerators. Environmental protection and energy efficiency policies, aided by financial and other policy instruments, have greatly stimulated the development of cost-effective technologies contributing to sustainable development. The planned 'greening' of the tax system is expected to give another incentive encouraging users to more sustainable consumption behavior. Environmental policies, however, have not prevented continuous increase in energy and water use, and car mobility. Both car ownership and annual mileage are still increasing. In spite of the heavy taxation on car ownership, road and fuel taxation, newly introduced taxation on energy use and the increased value added tax on drinking water use, consumption behavior seems hardly to have been influenced.

Liberalization of utility markets

This is a pronounced trend in most, if not all, Western economies. The market is being trusted to provide a better ratio of price and quality of service, as competition will force suppliers to work as efficiently as possible. In the same line of thinking public utility companies are being privatized, as private companies are expected to be more efficient, flexible and innovative than public utility companies. The outcome of the current process of transition will among others depend on the number of competitors and the distribution of market power among them, and thus on the structure of the market and the quality of market regulation. It is still an open question how innovative privatized utility companies will be in a liberalized market. Will competition with new parties entering the market force them to be innovative, or will the creation of shareholder value prevail over risky investments in research and development?

Internationalization of economies

As national economies become more and more interdependent and international trade barriers are removed, the liberalization of utility markets is also being effectuated on an international scale. Thus utility companies that once operated in a closed market under government protection, now have to face competitors from abroad, with access to other resources. In the global telecommunications sector, fewer and fewer companies can survive, especially if they are too small and not a partner in one of the strategic alliances spanning the globe. In the electricity sector, a similar trend of merging and formation of strategic alliances is already observed. For the drinking water sector, as the networks at present hardly exceed the regional scale, this development towards a reduced number of larger players is proceeding at a much slower pace.

Summarizing our explorations into the consequences of technological innovations envisaged, both curiosity-driven and society-driven, the following generic dilemmas for future infrastructure development can be identified from a design and control engineering point of view:

- **Decentralized utility supply versus centralized utility supply?**
  One extreme is fully autonomous, network independent supply for all users, the other extreme is a totally centralized supply, relying on one large scale transport and distribution network with a limited number of large scale processing facilities. Intermediate options are e.g., semi-autonomous supply with grid-connection, local networks competing with a national grid, or a single (inter)national grid supplied by a large number of decentralized supply units. How can the designer determine the optimum degree of (de)centralization of supply and scale of the network?
- **Local optimization versus overall system optimization?**
  Depending on the market structure, ownership and management responsibility of the infrastructure (sub)system(s), it may not be possible to optimize the infrastructure system as a whole, but to optimize only on the scale of subsystems.
- **Systems integration (local innovation) versus overall systems innovation?**
  Can new technologies for supply, end-use, storage or transportation be accommodated in the existing infrastructure or should the overall infrastructure system be restructured and innovated?
- **Mono-functionality versus multi-functionality of infrastructures?**
  Should an infrastructure by definition be designed as a system dedicated to one specific function or can (future options for) multi-functionality be accommodated for in infrastructure design?
- **Supply side management versus demand side management?**
  The answer to this question has profound technical consequences for the design and operation of an infrastructure, even though it is not formulated as a technical dilemma.
- **Stakeholder value (e.g., sustainability) versus shareholder value (e.g., profits)?**
  The complexity of system integration and the selection of (new) technologies for new infrastructures is only aggravated by the uncertainties about the outcomes of the current transition process towards a liberalized market. In this situation, the long term social and economic impacts of current decision making are at risk of being neglected in favor of short term business interests. Depending on which one of the two extremes prevails and which infrastructure is to be designed, the designer may make radically different choices for network structure and technologies. Privatized utility companies in a liberalized market are generally

expected to aim for the creation of shareholder value as the dominant performance criterion for infrastructure design and control engineering.

- **Robustness by redundancy/durability versus robustness by flexibility/responsiveness?**
  With regard to the large uncertainties about the outcomes of the current transition process, this is probably the most relevant question from a design and control engineering perspective. This question more or less encompasses the previous design dilemma's, which all point at a need to keep sufficient degrees of freedom in the design of an infrastructure to enable it to respond adequately to changes in its physical, social and economic environment. It is quite likely that future infrastructure designers will be challenged to design for a short lifetime of the infrastructure and easy dismantling, rather than to design for extreme durability. Rather than designing an inert infrastructure in which robustness draws on redundancy, the key challenge for future infrastructure design is envisaged to be to design for flexibility in accommodating new technologies and responding to market dynamics.

7. Research objectives

The motivation for research into the generic properties of physical infrastructures as well as infrasystems was indicated in the previous sections. In the present section the delimitation of the problem area will be investigated, and how scientific research can be pursued from a number of working hypotheses will be outlined. So far, the discussion of the research area evolved from public utility services, towards physical infrastructures, and onwards to the concept of integrated infrasystems. The systems point of view comes naturally from the many analogies and commonalities identified in a comparative analysis between infrastructures at the aggregated system level. Most infrastructure sectors are currently in the transition process towards market liberalization, privatization of public utility companies, and re-regulation. Other prominent driving forces towards change in most infrastructure sectors are internationalization and technological innovation.

A feature of the public discussion and the ongoing research on infrastructures is that both generally remain limited to a specific infrastructure sector. This is particularly true for technological research, which often focuses on sector-specific infrastructure subsystems and components, leaving the basic structure of the infrastructures unquestioned. As a result, the interactions between infrastructures and substitution options receive little or no attention in the technological literature. Rather than focussing on sector-specific features, the authors of this paper will therefore focus on the technological analogies and commonalities between the infrastructures for energy, water, waste, telecommunication and transportation.

The first objective of this paper is to explore if a generic perspective on the technological features of infrastructures is feasible and meaningful as a basis for the development of a generic research approach to the design and control of infrastructural networks, in spite of the pronounced technical differences at the component level.

The second objective is to develop a structured research plan that will enable a deeper understanding of those generic factors and phenomena that are critical towards optimally exploiting infrastructures in a modern society.

The exploratory nature of this paper must be emphasized. This paper is an attempt to place the subject into perspective. It is concerned with the formulation of research questions rather than answers, as it marks the preparatory phase of a long term research project. Moreover, the above objectives are of a basic nature and need to be elaborated to arrive at an operational research strategy in the concluding section of this paper.

8. Working hypotheses

The following *working hypotheses* will be used in the sequel as the leading ideas for the formulation of a research approach:

1. Infrastructure systems constitute a backbone of the economy of modern societies, and determine to a large extent the level of economic prosperity, public health and environmental protection.
2. Different infrastructure systems have many aspects and functions in common, and show common behavior in many facets. They can be characterized individually by the specific values attributable to these aspects and functions.
3. The behavior and operational functionality of an infrastructural system can be captured in terms of a qualitative and quantitative model. Consequently, it can be systematically described in mathematical terms by a model having structure and quantifiable parameters.
4. The evolution of an infrastructure, i.e. its growth and decline, can be captured in a mathematical model.
5. A systems engineering approach provides a suitable way of arriving at a generic, abstract and model-based theory that supports the economically profitable design and operation of infrastructures, based on a sufficiently deep understanding of the underlying mechanisms.

These working hypotheses will be discussed, elaborated on and amplified in the sequel of the paper. Some of the direct consequences will be discussed in the next section.

9. Research implications of working hypotheses

The first working hypothesis (the backbone role of infrastructures) implies that the interaction between a specific infrastructure, its interaction with other sectors of infrastructure and the economy, the role of technological development and innovation, and the impact upon the society and the resulting governmental regulation and legislation must be taken into consideration in the research program. The research must be directed towards identifying and unraveling the underlying mechanisms and laws and must indicate, if possible quantitatively, how various subsystems interact with each other. The systems approach requires the definition of the system boundaries with the (physical, social and economic) environment, the definition of subsystems (with the system decomposition strategy depending on the research perspective) and system elements, and the relations governing the interactions between the subsystems within the system and between the system and its environment. To arrive at a proper definition of infrastructure systems at the generic level, a deeper and systematic understanding of the following aspects and phenomena will be required:

- The factors governing the origin, development, evolution and decline of an infrastructural system
- The handles and actuation mechanisms that allow actors in the infrastructure playing field to influence the evolution of an infrastructure system. Deeper understanding of these mechanisms ideally leads to a theory of synthesis (design and development strategies) for infrastructure systems. What are the functional specifications and operational performance criteria, and how can actors plan their decisions in order to arrive at an infrastructure system with properties perceived as optimal.
- The actuation mechanisms sought after are those that infrastructure system operators can employ to operate the system in a preferred fashion. Operational management largely determines the power and profit created by the utility character of an infrastructure system. A deep quantitative knowledge is required to establish a theory of optimal operation and control for infrastructures. The interaction of operational control (matching supply and demand, ensuring availability and quality of service) with evolutionary control (growth and decline of an infrastructure system in a lifespan perspective) must also be given ample attention.

- The possibilities of quantitative modeling of an infrastructural system in its operation and its evolution in time must be explored and realized. Synthesis and control strategies for infrastructure systems can be derived in a systematic way only if quantitative dynamic models are available describing and interrelating the main phenomena of interest.

The second hypothesis (analogies and commonalities between different infrastructures) requires that the assumed common phenomena existing in different infrastructural systems are identified. The study of a number of specific infrastructural systems provides a good approach. An overview of a number of sector-specific properties will be discussed in the next section. This discussion will affirmatively lead to the tentative conclusion that infrastructures exhibit many properties that can be considered as generic.

The third hypothesis (the behavior and operational functionality of infrastructure systems can be captured in a generically applicable mathematical model) needs a thorough investigation and validation. The assumption assumes that a more or less generic model for infrastructure systems can be derived. A specific case will result from this model by the proper selection of structural properties and parameter values within the generic model structure. The hypothesis assumes the existence of a model, which implies that the behavior and evolution of an infrastructure is governed by specific laws instead of by randomly occurring influences. The research to be developed will have as one of its main goals the establishment of the assumed generic model. This requires us to make decisions regarding model structure, interaction of the infrastructure model with its environment, and assumptions on the variables that represent the interaction. The autonomous growth and decay phenomena will also have to be established. It will be necessary to investigate the existence of constitutive equations for the behavior of network components in the generic infrastructure system. The existence of suitable replacements for the "conservation laws" governing the economic and market behavior of infrastructure and its environment must also be investigated. In addition, application to actual data from existing infrastructures will be needed to evaluate and validate the validity of the assumed model structure and to estimate parameters for a number of specific situations.

The fourth hypothesis (the evolution of infrastructural systems can be described by a generic evolutionary mathematical model) will require a careful analysis of the mechanisms that affect the evolution in time of the infrastructure. The interactions with the environment also will play a major role in these mechanisms. The time scale of evolution will be orders of magnitude different from the time scales involved in the operation of the infrastructure. Thus separate models for both phenomena may be expected, although the slower growth model may provide the parameters for the faster operational model. It will be necessary to fit the model to various sets of historical data and to evaluate the predicting properties of the model. Technological innovation and economic factors may be the key variables that determine growth and decay. The actual research efforts will have to make it clear whether this is true and what other factors are of major importance.

The fifth hypothesis (the effectiveness of a systems engineering approach as a research methodology) is a natural one in view of the preceding hypotheses. A systems approach [1] provides the language to formulate deterministic and stochastic models. Part of the models will be based on a priori formulated basic relations that describe certain "conservation law"-like phenomena. Another part of the modeling effort will have to rely on the fit to actual data. The combination of both models in terms of "gray box" models [2,3] may offer a suitable approach. A basis for a systems oriented modeling approach may be found in Forrester's industrial dynamics approach bridging the gap between economic behavior, dynamic systems, and engineering systems design [4,5,6]. The economic literature provides sufficient material to support the modeling of infrastructures from an economic point of view [7-16].

10. Research approach

From a scientific point of view, infrastructures are interesting as objects of study for several reasons. The study of infrastructures may generate a better understanding of how technological innovations influence society and the economy. It may reveal the mechanisms behind the evolution, growth and decline of large scale technical systems in society. It may show the influence of market forces and government regulations upon this evolution. It may provide insights into how technological innovations are selected and brought to large scale exploitation. From a scientific point of view, the study of infrastructures thus allows the study of vital interrelations between economy, technology, government and society.

Systematic knowledge must be available for the development of robust long term strategies by which the present and future behavior of infrastructures can be explained, thus also enabling other sectors of the economy to interact more proactively with the infrastructure sectors. The acquisition of such systematic knowledge requires scientific research into the generic properties of infrastructures. It is assumed that such generic properties exist and possibly can elucidate certain specific experiences with or properties of specific infrastructures. This motivates us to explore and investigate the generic and systematic properties associated with infrastructural systems.

A structured approach for the formulation of a research program for the analysis, modeling and synthesis of generic infrastructures will be discussed in this section. The main steps envisaged to constitute a research program for the study of infrastructures from a design and control engineering perspective are as follows:

1. The desire to be able to design and control infrastructures requires a deep model-based understanding of the behavior and evolution of infrastructural systems. Deriving models first requires the description of generic functions of an infrastructure. A preliminary attempt, without going to deep levels of subsystems and components, has been exercised in this paper on the basis of five sample infrastructures. This effort will be continued, especially in trying to make a step from the specific to the generic case. A description in terms of physical and economical functionality of a generic infrastructure may show the basic mechanisms, their spatial distribution, the interconnection with other phenomena in the society, and the possibilities to quantify the mechanisms and phenomena involved.
2. The next step must bring the physical and economic functions to an abstract level where functionality, variables and subsystems can be defined. This step also involves a generalization from the sample level to the generic level. It requires us to define the functionality of an infrastructure in its interaction with its environment. It requires the definition of system boundaries, subsystems, and interaction variables. It must allow the functionality to be broken down into simple subsystem functions, each to be characterized by sufficiently simple mathematical relations. Some of these relations can be determined on the basis of reasoning about the underlying physical and economical laws, others must be estimated empirically for each infrastructure individually. The role of investments, technological innovation, economic laws of the market, government regulation and legislation and public versus private business development must be taken into consideration. The study of these issues is assumed to lead to the understanding of many more factors that contribute in a significant manner to infrastructure behavior.
3. In the next step, the understanding of the basic mechanisms must be elaborated differently in two directions. One is the understanding and modeling of infrastructure behavior related to normal operation. How can operation be influenced? What are the actuating mechanisms? What makes operation successful and what are the performance indicators that should be optimized for optimal performance? Are the dynamics of the infrastructure chain important in looking for optimal operation? Is the spatial structure (network structure, interconnection structures, fractal substructures) an important issue, and how can understanding of this structure be exploited in operation? If we understand the role of these structural properties, what design decisions should be taken to improve the parameter values or properties and what are the limiting factors in this respect?

4. The second direction is to understand the mechanisms of growth and decay of an infrastructure. It is conceivable that the natural mechanisms are strongly dependent upon the interaction with other sectors of the economy, and thus are not autonomously evolving in time. It must become clear in what manner one can influence the evolution by changing the infrastructural properties in a systematic fashion. Synthesis of infrastructures should be directed towards this issue.
5. Research in this area must borrow strongly from various related fields. The systems approach as used to analyze management and production systems [1] can be of help. The 'industrial dynamics' approach [4,5] has recently been shown to be effective in understanding the dynamics of supply chains [17,18] which has similarities with certain aspects of infrastructures. The time evolution of infrastructures may be directed towards finding certain rules that determine growth or decay. The study of other systems that may show these phenomena may assist in developing the required models [19].

References

[1] J.P. Van Gigch, *System Design Modeling and Metamodeling*, Plenum Press, New York, 1991.

[2] T.Bohlin, Derivation of a "designer's guide" for interactive grey-box identification of nonlinear stochastic objects, *International Journal of Control* 59(1994), 6, pp. 1505-1524

[3] A.H.Kemna, D.A.Mellichamp, Identification of combined physical and empirical models using nonlinear a priori knowledge, *Control Engineering Practice*, 3(1995), 3, pp. 375-382

[4] J.W. Forrester, *Industrial Dynamics*. The MIT Press, Cambridge, MA, 1964.

[5] R.G.Coyle, On the scope and purpose of industrial dynamics, *International Journal of Systems Science*, 4(1973), 3, pp. 397-406

[6] R.L.Tummala, L.J.Connor, Mass-energy based economic models, *IEEE Transactions on Systems Man & Cybernetics*, 3(1973), 6, pp. 548-55

[7] A.P.Elhance, T.R.Lakshmanan, Infrastructure-Production System Dynamics in National and Regional Systems: An Econometric Study, *Regional Science & Urban Economics*, 18(1988), 4, pp. 511-531

[8] A.Shah, Dynamics of Public Infrastructure, Industrial Productivity and profitability,

Review of Economics & Statistics, 74(1992), 1, pp. 28-36

[9] G.Glomm, B.Ravikumar, Public investment in infrastructure in a simple growth model, *Journal of Economic Dynamics and Control*, 18(1994), 6, pp. 1173-1187

[10] D.Holtz-Eakin, A.E.Schwartz, Infrastructure in a structural model of economic growth, *Regional Science & Urban Economics*, 25(1995), 2, pp. 131-151

[11] J.R.Neill, Fueling the engine of growth with investment in infrastructure: A lesson from neoclassical growth theory, *Journal of Macroeconomics*, 18(1996), 3, pp. 521-529

[12] D.Holtz-Eakin, M.E.Lovely, Scale economies, returns to variety, and the productivity of public infrastructure, *Regional Science & Urban Economics*, 26(1996), 2, pp. 105-123

[13] S.H.P.Lau, C.Y.Sin, Public infrastructure and economic growth: Time-series properties and evidence, *Economic Record*, 73(1997), 221, pp. 125-135

[14] K.Conrad, H.Seitz, Infrastructure provision and international market share ivalry, *Regional Science & Urban Economics*, 27(1997), 6, pp. 715-734

[15] B.Sanchez-Robles, Infrastructure investment and growth: Some empirical evidence, *Contemporary Economic Policy*, 16(1998), 1, pp. 98-108

[16] F.Nourzad, Infrastructure capital and private sector productivity: A dynamic analysis. *Quarterly Journal of Business & Economics*, 37(1998), 1, pp. 13-25

[17] D.R.Towill, Industrial dynamics modelling of supply chains. *Int. J. of Physical Distribution & Logistics Management*, 26(1996), 2, pp. 23-42

[18] D.R.Towill, M.M.Naim, J.Wikner, Industrial dynamics simulation models in the design of supply chains. *Int. J. of Physical Distribution & Logistics Management*, 22(1992), 5, pp. 3-13

[19] F.E.Cellier, *Continuous System Modeling*. Chapter 10: Population Dynamics Modeling. Springer Verlag, Berlin, 1991.

[20] T.P. Hughes, *Networks of Power, Electrification in Western Society, 1880 - 1930*, The John Hopkins University Press, Baltimore and London, U.S.A., 1983.

Systems Engineering: The Telecommunications Infrastructure in 2030

Jens Arnbak

Netherlands Regulatory Authority for the Telecommunications and Postal sector (OPTA)

&

Delft University of Technology

Faculty of Information Technology and Systems

Abstract

The major driving forces in infrastructural network design are economies of scale and scope in the provision of services to the user community sharing the network. The present extremely rapid reductions of the cost of modern information and communication technologies (ICT) are causing fundamental shifts in the planning and development of the network architecture and service capabilities of future telecommunications. This is illustrated by a systems engineering approach.

1. Introduction: A definition of Systems Engineering

'Systems theory' and 'systems engineering' are widely used terms in science and engineering – so much so that it has become difficult to know what these terms actually mean in a given textbook. This discussion follows the authoritative *Encyclopædia Britannica,* which contains the lemma shown in condensed form in the box on the next page.

This lemma goes far beyond the conventional model of a communications system used in most undergraduate engineering courses, namely: "the chain of mechanisms required for (electromagnetic) transfer of messages from an information source to an information sink".

"**Systems engineering,** technique of using knowledge from various branches of engineering and science to introduce technological innovations into the planning and development stages of a system. Its first application as a specific discipline was in the organisation of commercial telephone systems in the 1920s and 1930s. The systems engineer is usually called upon to incorporate new technology into a system that is

- man-made;
- large and complex (where a change in one part, or subsystem, may affect many others);
- stochastic (subject to random, unscheduled changes).

After identifying the objective[s] of the current system, the system engineer adjusts the new technology to maintain that objective." (*Enc. Brit.*, 1995, Vol. 11, p. 472)

This definition corresponds to the simple representation of a communications 'system' as the isolated sequence of elements shown in Fig. 1. Such a simple model is useful for classification, description and analysis of given (prescribed) types of links for one-way transmission of information [1]. In the scientific theory of systems, such a generic model with minimal interaction with the rest of the world is known as a 'metasystem' [2].

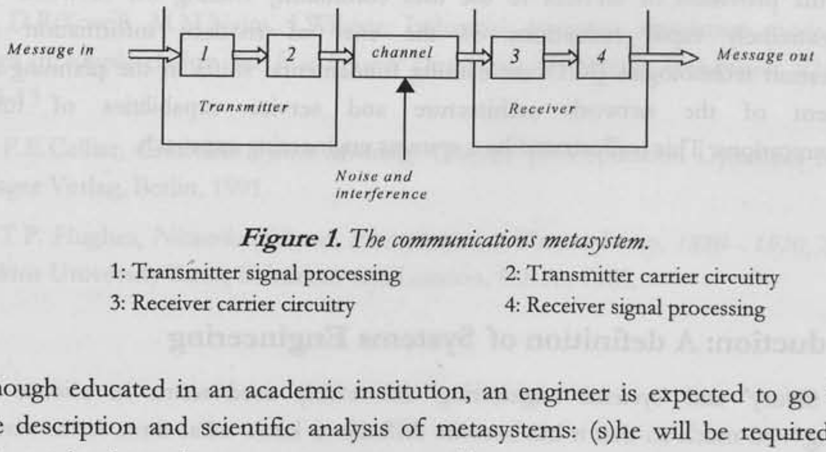


Figure 1. *The communications metasystem.*

1: Transmitter signal processing — 2: Transmitter carrier circuitry
3: Receiver carrier circuitry — 4: Receiver signal processing

Although educated in an academic institution, an engineer is expected to go beyond a pure description and scientific analysis of metasystems: (s)he will be required to plan, design, develop, integrate, operate and/or upgrade real-world systems. In telecommunications engineering, this synthetic creativity has to be exercised in response to external objectives and quality requirements, e.g. to provide a certain traffic capacity, signal quality and service reliability, and under various legal, administrative, and economic constraints. In general, the only viable way to provide communications services to a real society with limited resources, in terms of capital, manpower, frequencies, antenna locations or other real estate, etc., will be to share such resources or facilities among many users. Precisely this collective feature of facility sharing tends to result in very large systems: Clearly, individuals cannot afford to establish and maintain a permanent link to each potential communication partner in the World. In a dynamic environment, facility sharing includes traffic switching and adaptive control (network management) of the system resources to match the fluctuating demands of individual users and the collective resources of the system from time to time. It should be noted that the closed metasystem in Fig. 1 leaves aside the external demands which, in an open and unpredictable society, influence the design and operation of any useful communications system.

The simple reason for, despite this, studying closed systems in science and academic education is analytical convenience: Often, it proves sufficient to consider a very limited set of

1. natural laws, like Newton's laws or Maxwell's equations, and
2. fundamentals limits, say, the speed of light, or the Shannon bound on channel capacity

to obtain a mathematical description and broad physical understanding of closed systems. This approach is particularly useful in the study of idealised models, such as the elliptic orbits of individual planets around the Sun. In reality, however, each planets orbit is affected and perturbed by those of the other planets. Outside the closed academic Paradise of 'pure' idealised problems, a demanding world superimposes its complicated boundary conditions and 'dirty' interactions on to the general scientific relations taught in undergraduate, and surprisingly many graduate, engineering courses. Professional engineers will therefore also be confronted with

3. society norms and/or demands from competitive markets.

Accordingly, telecommunications engineers have to establish and meet functional specifications for engineering and design of open systems. Differently stated, they are asked to define and synthesise systems that satisfy both the application demands of specific users and the standards for interfacing with other systems. As known from the layered Open Systems Interconnection (OSI) network model [1, sect. C-3], open models tend to be much more complicated than typical models of closed systems based purely on knowledge of fundamental principles or present device capabilities. The complications arise because a proper perspective of any useful communications system requires a third dimension: systems engineering. This is the technical discipline concerned with the definition, selection and development of a complete solution to a functional problem defined in a real environment.

Symbolically embedding the closed communication system model given in Fig. 1 in a real environment results in the open system model shown in Fig. 2. This model [3] illustrates the inclusion of the extra perspectives not fully covered by the fundamental layer of the exact (β) sciences. More user-oriented, often empirical disciplines, such as α) social communications or media psychology,

γ) economics, including industrial organisation, or law

are frequently required to model complex man-made communication systems adequately. Nevertheless, such systems may still be treated by quantitative engineering methods, especially if the behaviour of the involved communities of users can be described by statistical methods. This is generally the case when the number of similar users is great, and they are sufficiently independent of each other. As known from teletraffic theory [4], the average inter-arrival time and holding time of telephone calls during the busy hour of the user community considered is sufficient to determine the overall grade of service (blocking rate). The publication in 1917 by A.K. Erlang of a pioneering theory for cost-effective dimensioning of large automatic telephone networks explains why 'systems engineering' emerged in the 1920s and 1930s as a new discipline for dimensioning and upgrading large systems.

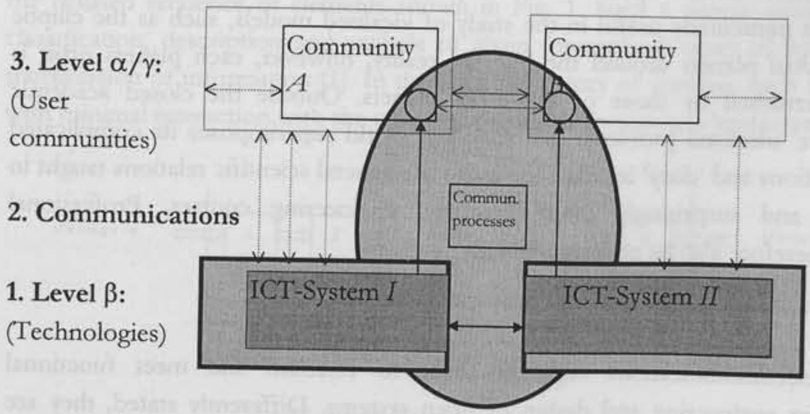


Figure 2. *Processes, objects and subjects of information & communications technology (ICT):*

- ICT systems: information & communication technologies (**1**), supporting
- open (tele-)communications processes (**2**), as determined by their
- ambient human, societal and economical (f)actors (**3**).

Legend:

full arrows: (tele-)communication processes, meeting the requirements of

a: an ensemble member of community *A*; user of/connected to ICT-System *I*

b: an ensemble member of community *B*; user of/connected to ICT-System *II*

interconnection links between ICT-systems (e.g. *I* & *II*).

dashed arrows: secondary interactive processes (for reference).

2. Outline of the Challenges of Telecommunications Systems Engineering

As defined in the box in section 1, system engineers are engaged in designing or improving large systems, which are stochastic, man-made (as distinct from biological or other natural systems, such as the weather) and include many interdependent sub-systems. An increasing number of possible physical transmission media - satellite and terrestrial radio relay, optical cables, copper wire pairs and wireless local loops, mobile links, broadcast transmitters, and so on - have become available and have to be considered as design options.

Concentration of traffic (in nodes or exchanges) may be attractive for sharing the more costly of these physical media. This trend, in turn, drives network topologies towards fewer and bigger nodes and exchanges. In this way, a reduction of the unit cost per circuit is achieved by adding circuits on each link, to form high-capacity ("multiplexed") links. The trend towards extremely high-capacity systems is now leading to international 'mega-carriers', i.e. joint ventures or mergers of national telecommunication firms with global ambitions. Examples are the AT&T alliance with British Telecom (BT) announced in the summer of 1998, Unisource (the alliance of KPN, Swiss Telecom and Telia of Sweden from 1991, recently disbanded), Global One (France Télécom, Deutsche Telekom and Sprint), and WorldCom's acquisition of MCI. In contrast, an opposite trend is observed locally, due to the lower price of single circuits to individual users made possible by modern radio technologies, such as wireless local loops and mobile cellular systems like GSM. This suggests that the unit cost per circuit is the key incentive to look for the most appropriate network architecture and business structure.

An important consequence of this is the likelihood of interconnection of several links to form large transmission systems. Traffic flow between two widely separated users may pass through various transmission sections connected in tandem; obviously, all these sections must have both compatible interfaces and sufficient signal quality to meet the end-to-end standards required for the particular circuit between any pair of communicating users considered. Note that the quality of an electronic circuit between two end-users – or between the end user and a distant ICT system, such as a data server in the World Wide Web – will be compared to 'live' interaction between the persons and entities concerned (see the dashed lines in Fig. 2). Not surprisingly, such a comparison is particularly complicated for international connections and interoperations of systems owned by different companies, e.g. computer communications over public data networks; calls from the fixed telephone network to mobile terminals. Most connections are not permanently established, but switched (on and off, by exchanges); thus it would be incorrect to compensate a poor quality of one section in another section, since the poorer section could subsequently be switched to different sections without adequate compensation. A better approach is therefore to line each individual transmission section up to jointly agreed performance standards, e.g. those recommended by international committees and organisations such as the ITU-T (formerly CCITT[1]), the ITU-R (formerly CCIR), and the European Telecommunication Standards Institute (ETSI).

Jointly agreed performance standards often relate the maximum tolerable level of signal degradations to a hypothetical reference circuit (HCR), an artificial model of a real long-distance circuit of defined length and with a specified number of intermediate equipment (multiplexers, modulators, etc.). The HRC is more specific than the metasystem shown in Fig.2, but it can embrace many individually designed systems of a particular type and so is more suited to describe the performance of communications circuits connected through such systems at random. Still, the planning and design of individual systems include many specific criteria that are not found in Fig. 2 or the HRC. Typical design criteria are listed in Table 1 and cover some of the conditions and requirements which are imposed on the system from outside by users, authorities, and co-operative or competing systems.

Table 1. *Typical categories of specifications for a communications system.*

DESIGN ASPECT	*SPECIFICATION/SYSTEM NORM*
1. Provisioning period	The period for which the system is designed for adequate service.
2. Route capacity	Provision of sufficient system resources (e.g. power, bandwidth, coverage area, hardware facilities) to meet the forecasted user traffic in the provisioning period (1), for each required link (route)
3. Expansion capability	Possibility of upgrading the system to meet higher traffic demands than forecasted for the provisioning period (1)
4. Connectivity	Provision of the traffic routes between the network locations forecasted in the provisioning period (1)
5. Flexibility	Ability to adapt the route capacity (2) and connectivity (3) to varying traffic patterns
6. Transmission quality	The signal norms required for adequate service
7. Availability	The percentage of time in the provisioning period (1) in which adequate route capacity (2) is provided with the prescribed transmission quality (6)
8. Equipment reliability	The contribution to availability (7) from adequate equipment design and redundancy
9. Working ('hot') standby capacity	The contribution to availability (7) from extra equipment with rapid cut-over into service.
10. Maintainability	The contribution to availability (7) from good repair possibilities, easy supervision, etc.
11. Redundant ('cold') sparing	The contribution to maintainability (10) from additional equipment, stored in reserve positions.
12. Route diversity	The contribution to availability (7) from meshed network lay-out

[1] See the list of abbreviations in the front-end pages of Couch [1].

DESIGN ASPECT	SPECIFICATION/SYSTEM NORM
	with alternate routings
13. Interfaces	The hardware and signal formats required for adequate interconnection with other systems and with terminals
14. Electromagnetic compatibility	Ability of the system to operate adequately without causing or suffering from electromagnetic interference.
15. Cost	The total cost of installing and operating the system in the provisioning period (1), using a prescribed accounting method.

It is the task of the communications systems engineer to establish the optimum system solution in a defined environment by rational methods. If the environment is not sufficiently specified for an optimum to be determined, this engineering task also includes the exploration and stipulation of additional criteria necessary to determine a technical solution in a non-arbitrary way. The next section shows the significance of imposing sufficient functional specifications on a communications (meta-)system to obtain a meaningful norm for an "optimum" (or at least a "useful") system design.

3. The Relevance of a Normative System Approach

To demonstrate that natural laws, fundamental limits and advanced device technology do not suffice to define and design a useful communications system, return for a moment to the metasystem given in Fig. 1. Let its overall end-to-end transmission quality (as defined in Table 1) be given in the generic form of a figure-of-merit, typically a signal-to-noise ratio link budget. By taking logarithms, this ratio is determined as a linear expression in decibels (dB)

$$S/N = \sum_{i=1}^{K} G_i(x_i) - \sum_{j=1}^{M} N_j(y_j) \quad . \tag{1}$$

Here, G_i is the gain contribution (in dB) determined by the independent system variable x_i, while $-N_j$ is the noise degradation (in dB) determined by the independent system variable y_j. There is a total of $K+M$ system variables; we shall assume these variables to be non-negative physical quantities. In a radio link, for example, x_1 and x_2 might represent RF transmit power and receive antenna diameter, respectively, whereas y_1 and y_2 might be the preamplifier noise temperature and the system noise bandwidth. See the various typical link budgets in [1, pp. 572- 580], which illustrate the very general nature of the link budget (1).

Note that it is the end user who will judge whether the overall transmission quality (1) is "adequate", generally based on a subjective comparison with a reference situation known from real life (see the various dashed arrows in Fig. 2, which show a variety of such reference situations: man-man, man-machine and machine-machine interactions). For example, the audio quality of a telephone circuit is designed to allow an untrained listener to recognise a speaker from his or her voice, which typically requires a level some 35 dB above the ambient noise level. High-fidelity ("HiFi") quality reproduces the subjective audio perception in a concert hall; this level is substantially (30-40 dB) above telephone quality. The determination of appropriate measures of quality, as perceived in test series by representative groups of end users, is an interesting psycho-physical engineering discipline of its own, based on statistical analysis of (long) series of subjective opinion scores. Clearly, this exercise is impossible in any closed system model (unless we could imprison the users inside the metasystem!). Here, we restrict ourselves to asking three seemingly simple questions and pursuing their answers.

Question 1: How should the set of *K*+*M* free system variables $\{x_i, y_j\}$ be selected to maximise the end user's S/N?

Answer: If (as is usual) G_i and N_j are monotonically increasing functions, this solution is obviously found in the limit determined by

$$x_i \to \infty \quad \text{for} \quad i = 1, 2, ..., K$$
$$y_j \to 0 \quad \text{for} \quad j = 1, 2, ..., M$$

In the simple example of a radio link given above in conjunction with (1), the transmit power and receive antenna diameter should be infinite large, whereas the preamplifier noise and (message!) bandwidth should be zero. Apparently, our specification has led to a very costly and entirely useless communications system. A wiser design question is the following:

Question 2: How should the free system variables $\{x_i, y_j\}$ be selected to reach a specified finite user $(S/N)_{sp}$?

Answer: There is <u>no</u> unique solution to this problem. We are free to select any *K*+*M*-*1* system variables, say x_i and $y_j (\forall i, \forall j \neq M_0)$, as long as the remaining single variable is determined by

$$N(j_{M_0}) = \sum_{i=1}^{K} G_i(x_i) - \sum_{\substack{j=1, \\ j \neq M_0}}^{M} N_j(y_j) - (S/N)_{sp} \quad . \tag{2}$$

Thus we can trade off freely between all system parameters except one, so this specification has not resulted in any particular technical design. In the above radio link example, X dB reduction of antenna gain would (of course) be offset by X dB increase of transmitter power, or by any other equivalent improvement of subsystem performance. This illustrates the embarrassing number of possibilities for designing sub-optimal systems by combining excellent and poor elements to meet a total specification. To avoid this, we finally pose the following

Question 3: How should we choose the free system variables $\{x_i, y_j\}$ to achieve a specified user signal-to-noise ratio $(S/N)_{sp}$ at minimum system cost?

Answer: Let the cost impact of each independent system variable be expressed by (known) non-negative functions $g_i(x_i)$ and $n_j(y_j)$. The total system cost to be minimised is then

$$C = C_o + \sum_{i=1}^{K} g_i(x_i) + \sum_{j=l}^{M} n_j(y_j)$$

where C_o is that part of the total system costs which is independent of the K+M free variables.

The Lagrange function to be minimised [5] is

$$L\{x_i, y_j\} = C + \lambda((S/N)_{sp} - S/N). \tag{3}$$

The constant λ multiplies a function which is zero if the system specification is met. A minimum-cost solution of (3) requires satisfaction of the associate K+M conditions

$$\frac{\partial L}{\partial x_i} = g_i^{'}(x_i) - \lambda G_i^{'}(x_i) = 0 \,; \quad i = 1, 2, ..., K$$

$$\frac{\partial L}{\partial y_j} = n_j^{'}(y_j) + \lambda N_j^{'}(y_j) = 0 \,; \; j = 1, 2, ..., M$$

A non-trivial solution of the above system of homogeneous equations requires the λ-value to be the same for all system variables

$$\lambda = \frac{g_i'(x_i)}{G_i'(x_i)} \equiv -\frac{n_j'(y_j)}{N_j'(y_j)}, \forall i, j. \tag{4}$$

This is a very significant result: The minimum-cost norm (3), which includes meeting the specified S/N, is satisfied only if the set of system variables $\{x_i, y_j\}$ is chosen such that the cost increase for an arbitrarily small performance improvement is the same for all system variables. This is known as balanced cost-effectiveness. In more popular terms: "A small performance improvement of, say, 0.1 dB should cost the same for all system elements". On hindsight, this result is hardly surprising: If indeed "cheaper decibels" could be found somewhere in the link budget, the designer should have chosen to exploit these in the design, which therefore cannot be optimal. Clearly, it is the cost norm (3) which prevents us from making the capital design error in answering Question 2 above: focussing only on certain system elements by specifying these tightly, while neglecting other elements when meeting the total system specification. Thus, use of an overall norm is the difference from answering Question 2 above.

To quantify the solution, the incremental cost-effectiveness figure, λ, should be determined by finding that particular set of solutions to (4), $\{x_i, y_j\}_{\lambda_{opt}}$, for which the overall specification is satisfied. When the functions in (4) are not simple analytical expressions, this may require iteration by graphical or computer methods. The total cost is finally obtained from (3).

The progression of answers to the above three questions demonstrate that a (meta-)system cannot be claimed to be optimum, unless a sufficient number of external norms and requirements are used to fix all its degrees of freedom. The systems engineer should be conscious of this fact: the nature of a system designer's task is different from both descriptive science and device expertise: Scientists explore general physical relations like (1), while technology specialists tend to focus their attention on only a few of the cost-effectiveness ratios in (4).

In systems engineering, the specialist's pleasures should always be tempered by the generalist's more sobering knowledge that the choice of state-of-the-art technology can prove costly (Question 2) or even silly (Question 1), if no functional norms have been formulated for optimality in the total context of the desired system. It is an important task for a systems engineer to determine the desired norms, if necessary by presenting clear options to the appropriate decision makers (the customers, the company management, or the politicians, as the case may be). For a professional expert accustomed to digging deeply into in a particularly challenging technical problem, keeping this cool distance often proves difficult. Nevertheless, it behoves anybody with a sense of academic freedom, objectivity and professional integrity to investigate and state the identifiable reasons for, and consequences of, the techniques and systems which (s)he has chosen to develop and recommend.

4. Introducing New Technology – When?

By quantifying the benefits and penalties of a particular technology in a given system context, a cost-benefit optimisation can be made along the lines sketched in Sect. 3. Such an optimisation, however is highly time-dependent for new ("state-of-the art") technologies. The strategy for introducing new technologies in large existing communications systems and services should be carefully considered. A main risk was illustrated in the previous section, namely "too expensive dB's" for some of the system variables in (4). In a large transmission system, this could happen for several reasons, such as:

- the introduction of technology which is still in rapid development, and thus has not yet reached the low, stable price levels of mature commercial products. Hence it might pay to wait.
- the forced amortisation (writing off) of a previous investment in an existing system, making early replacement more expensive than the pure technical cost of a new system. This consideration has been decisive in most existing telephone networks in Europe and the US. It explains why complete digitalisation is made much sooner in 'green-field' systems (completely new networks) and some special non-commercial services (for example diplomatic 'hot-lines', which require encryption for security reasons).
- the difficulty of forecasting traffic for new public services and systems (e.g. broadband networks for multimedia use). Such forecasts are speculative due to unknown customer judgements, whereas the known annual traffic evolution in traditional telephone systems can be factored into a more gradual strategy of system upgrading.
- the interface problems when adding an overlay system to an existing system (to accommodate traffic growth), if the two systems serve the same group of users ('community of interest'). Interface equipment might be avoided at a later stage of an optimum transition plan.

It follows that the shaping of an optimum strategy for modernisation of an *existing* telecommunications network is far more complex than a decision to introduce a *new* network. In the former case, the choice is between increasingly progressive strategies for transition based on:

1. continued expansion of the existing facilities and technology in the system

2. overlay with new facilities to accommodate the future growth of traffic in the system
3. gradual replacement by new facilities/technologies
4. immediate total replacement by new facilities/technologies

The proper choice or mixture of such strategies is an exercise in non-linear programming [6], which is beyond this discussion. Nevertheless, a rational system engineering method exists for finding the best time schedule. Enforced introduction of new technological solutions, simply because they are intellectually appealing to researchers, almost never maintains (or provides) the specified service at the lowest total cost. The normative systems engineer focuses on those bottlenecks needing improvement in time to allow smooth and efficient evolution of a complete system <u>in its particular environment</u>. This requires a good perception of both the existing technical approaches and their constraining influence on the development plans for new large communications systems, and of the dominant trends and costs of new technological opportunities. This is discussed briefly in the following section.

5. Technology Trends and Their Implications for Communication Systems Engineering

The following brief review of major market trends in information and communication technology (ICT), relevant for system engineers, illustrates some major technology forces and changing cost drivers of the rapid growth of the five ICT-based systems and services sectors shown in Table 2. Note that, roughly, the turn-over of the service sectors of ICT is twice that of the Dutch hardware sectors[2].

Table 2. *ICT-markets in the Netherlands, in million guilders. [1.7]*

SECTOR	*1994*	*1995*	*1996*	*1997*
IT hardware	6745	7498	8100	8683
Software products	3612	4008	4523	5095
IT (computer) services	5801	6162	6541	6948
Telecommunications hardware	2611	3513	3230	3496
Telecommunications services	11550	12734	14404	16346
TOTAL	30319	33915	36798	40568

[2] Figure 3 shows that this ratio is similar for the European internal market as a whole.

5.1 Capabilities and Trends of VLSI Technology

The powerful communication and computing functions and equipment embodied in Table 2 have become economically viable thanks to the microelectronic revolution. Integrated-circuit (IC) technology allows a huge number of electronic components to be etched into a minute 'chip' cut from a semiconductor material. The vast majority of VLSI chips in telecommunications still use cheap silicon wafers[3]. In 1994, the turn-over in microelectronics accounted for less than 10% of the hardware market, and 3% of the total turn-over in ICT – see Figure 3. On the one hand, this illustrates the extremely high - and still increasing - productive efficiencies of modern microelectronic manufacturing. On the other hand, the added value (which reflects employment) in the systems and service sectors is much higher. Modern handheld mobile phones are based on low-cost VLSI chips and work in the 900 or 1800 MHz bands (or even in both bands, in the new 'dual mode' terminals required for roaming between mobile networks with different frequency standards such as GSM and DCS1800). This is one of the chief reasons for the spectacular drop in the price of mobile radio terminals, which in turn has stimulated the new world mass market for mobile telephony.

VLSI technology evolves in accordance to a rule-of-thumb postulated in 1965 by Gordon Moore, one of the founders of the leading US chip manufacturer, Intel. Standard chips come in families, with the generations separated by a couple of years to benefit from the major economies of scale in industrial mass production[4]. Broadly stated, Moore's 'law' predicts that the maximum number of transistors or other components on a state-of-the-art VLSI chip doubles in 18 months, so the cost per component halves between successive mass-produced generations of chips. Indeed, the number of components on a VLSI chip has increased by a factor of 100 in less than 10 years! This has shown up in both prices and improved storage capacity of random-access memories (RAMs), used for instance in personal computers; but as mentioned above (see footnote 2), the maximum operating frequencies of silicon chips also continued to break through expected barriers. This increased the output signal frequency of electronic systems, and the internal 'clock' frequency by which transistors are switched, when working in the binary mode used in digital telephone exchanges and other computers. In 10 years, the clock frequency of microprocessors has gone up by a factor of about 40. As a result, the total processing power of a standard microprocessor chip has increased by some 100*40 = 4000 times per

---

[3]When the author (J.A.) studied engineering some 30 years ago, the opinion of leading semiconductor experts was that sophisticated and costly compound materials (such as gallium arsenide) would soon become necessary for high-frequency circuits, to meet the demands for smaller radio sets at microwave frequencies. However, the development of silicon technology extended its practical frequency ranges by an order of magnitude, while reducing the price.

[4]Well-known examples of large industrial families are Intel's 286, 386, 486 and 'Pentium' microprocessors, and the 640 kbyte and 1, 2, 4, 8 and 16 Mbyte memory chips used for RAMs in successive PC generations.

decade. Thus, the cost trend over the years has been an exponential drop in the price-performance ratio for electronics subsystems for telecommunications. Accordingly, some of the cost-effectiveness ratios in (4) changed exponentially, and thus gave, in the 1980's, a strong stimulus to the use of more and higher-graded microprocessors in telecommunications systems and terminals.

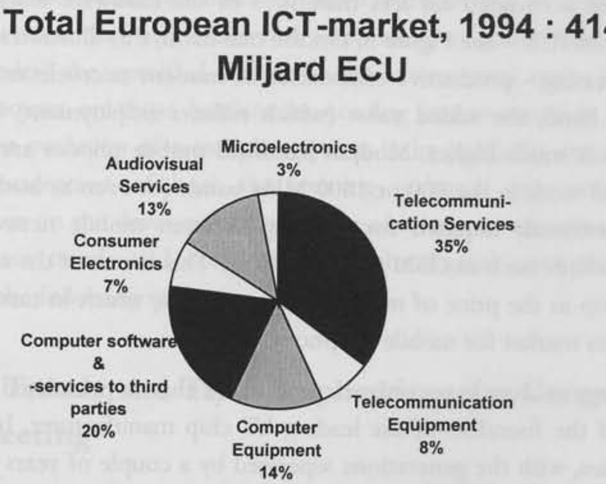


Figure 3. Distribution of total turn-over in the ICT-sectors in Europe (source: DG XIII, Commission of the EU)

Moreover, as illustrated in Figure 4, confluence of previously separate electronic industry branches has occurred in the past 10-15 years and extended a more professional performance to consumer equipment. Where such 'cross-overs' between industry branches has happened, the processing capacity of terminals and networks has risen even faster than predicted by Moore's law. In addition, labour costs - and, consequently, employment - have been reduced to a minor factor in modern manufacturing (mass production) of personal *digital* terminals. In 1994, leading manufacturers of mobile GSM terminals reported that the average human assembly and test time spent on a hand-held digital terminal had been reduced to 12 minutes, down from the 8 hours typically required for a first-generation (analogue) terminal produced in 1988. As previously experienced in manufacturing of radio and TV receivers, such VLSI-based mass production have brought affordable, yet highly advanced terminals within direct reach of most consumers. In the particular event of GSM introduction, it did not pay to move terminal production to other regions in the World with lower labour costs, as long as greater market demands for terminals existed in Europe. The exploding GSM market in Europe since 1992 and the (initial) hesitation to liberalise terminal markets in Asian regions with cheaper labour, made manufacturing close to the laboratories and growing markets in Europe advantageous, despite much higher labour costs. At present, the networks based on the internet protocols (IP) represent another cross-over path between classical industry branches.

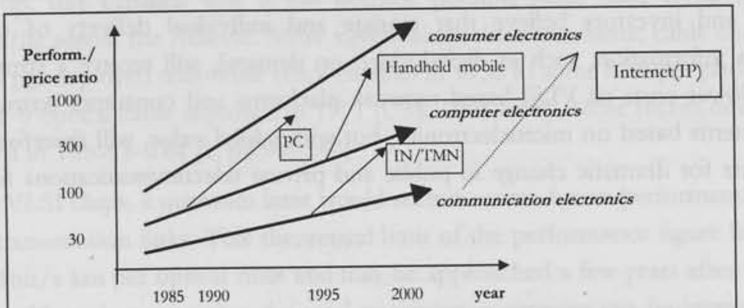


Figure 4. *Performance trajectories of ICT-product sectors, and recent cross-over paths.*

Will Moore's law continue to apply indefinitely in the future, and so produce ever increasing dynamic changes - if not turbulence - in the competitive markets for telecommunication terminals and services? Two probable limitations should be mentioned briefly.

The first, rather fundamental limit to further growth of VLSI capacity and functional capabilities arises from the fact that the transistors on present generations of silicon chips are already very closely packed. In the near future, the spacing between individual transistors on one chip will become so minute that the effects of quantum physics will start to affect individual electrons. These will probably either 'tunnel' through the walls separating individual transistors or be disturbed by individual atoms or impurities in the semiconductor material. In either event, the result will be less reliable transistor performance. With millions of even slightly unreliable transistors on a single VLSI chip, this would become useless in practice. Avoiding such problems already now leads to exponentially increasing factory costs for each new chip generation. This reduces the benefits of mass production, unless production volumes in each successive generation can also increase exponentially. Probably, the cost of a VLSI chip will, in the future, no longer be dominated by its *marginal* cost, but by the tremendous *fixed capital costs* of building each new chip factory, at present, several billion US$. This problem could be mitigated by co-operative teams (cartels?) of competitors, or by accepting longer lifetimes for each chip generation; in either case, the assumptions underlying Moore's law would cease to apply.

A second economic problem, related to the periodic microelectronics supply in separate chip generations, is the inherent creation of a commodity market, with cyclical fluctuations in price and demand as found in farming and for raw materials such as crude oil (OPEC). This could result in pleas for import regulations and national protection of producers[5] when prices drop. As is known from agricultural policy, this would lead to inefficient markets and, consequently, another departure from Moore's law.

Despite the changing seasons of microelectronics, its revolutionary crops are far from completely harvested by new ICT systems and service sectors. For instance, many observers and investors believe that storage and individual delivery of compressed multimedia information, such as digital video on demand, will receive a strong impulse from the lower costs of VLSI-based network platforms and consumer terminals in the future. Systems based on microelectronics, but with added value, will therefore continue to be agents for dramatic change in public and private telecommunications for years to come.

5.2 Capabilities and Limitations of Photonic Technology in Optical Transmission

The (re)invention of optical fibre technology in the 1960's[6] and the development of solid-state lasers generating infrared light have, in combination, probably already had a greater economic impact on telecommunications services than the electronic revolution discussed above. Specifically, this 'photonic' technology has enabled broadband transmission systems. In the area of long-distance high-capacity transmission, fibre-optical systems have completely out performed coaxial cables and permanent satellite links[7] within one decade.

A standard performance figure-of-merit for an optical fibre link is the product of its data transmission capacity (in Mbit/s) and the transmission distance bridged (in km). The first commercial systems in 1976 had figures-of-merit of about 20 (Mbit/s·km), corresponding to conveyance of 30 digital voice channels over a distance of 10 km. Since then, the performance figure of state-of-the-art optical links has continued to grow exponentially at

[5] "When the chips are down", Editorial, *The Economist*, March 23, 1996

[6] It is seldom recognised in the telecommunications field that optical fibres have been patented and used in medicine for visual probing inside blood vessels since the early 1950's. In 1966, Kao and Hockham (UK) on theoretical grounds proposed application in long-distance transmission of light signals.

[7] Satellites have made the World shrink since the Syncom-III TV distribution links to Western broadcasters from the Olympic Games in Japan in 1964; however, the most important satellite applications are now in support of mobile users, direct broadcasting, tailor-made links in unforeseen circumstances (e.g., disaster relief; CNN reports from 'hot spots'), and other thin-route traffic.

a rate of some 75% per annum, passing 2000 Mbit/s·km in 1984, and reaching 2 million Mbit/s·km in 1992.

The cumulative transport capacity of transatlantic submarine cable systems went up from 24 voice channels in 1956 (coaxial cable) to well above 100,000 channels in 1992, thanks to the dominance of optical technology from about 1987. This indicates why fixed satellite links can no longer compete on such high-density routes. By 1994, considerable spare capacity on the Atlantic cable routes between Western Europe and North America had built up; this explains why it has become possible since then to have so much Internet traffic across the Atlantic. More significantly, the transatlantic cable cost dropped from 7 M$ per deployed telephone voice channel in 1956 to some 6 k$ per phone channel in the TAT-9 optical cable deployed in 1991 (US$, 1991 level). More recent developments are included in Table 8-2 of [1, pp548-549]

Just as for VLSI chips, a quantum limit would seem to curse future performance jumps of photonic transmission links. This theoretical limit of the performance figure lies at about 1 billion Mbit/s km per optical fibre and may be approached a few years after the turn of the century. Note, however, that the total transmission capacity can be increased simply beyond the single-fibre limit, by including more fibres in one cable, and by adding light signals of different colours on one optical fibre. This hardly increases the laying costs of the cable. To understand the impact of opto-electronic transmission on long-distance telecommunications, consider the following realistic case of transoceanic conveyance costs.

<u>Problem</u>: *A submarine transatlantic (fibre-optical) cable was deployed by an international consortium in 1992 at a cost of 6000 US$ per telephone channel. How much does the corresponding transatlantic transmission cost contribute to the national tariff for an international telephone call between Europe and North America?*

To estimate the cost of transatlantic conveyance, we assume the following:

- *100,000 call minutes/year for each two-way circuit , corresponding to 5 busy hours/day*
- *The operational lifetime of the cable is 10 years (1993-2002)*
- *The cable investment was financed by a loan with an interest rate of 12% in 1992*
- *The annual operation and maintenance (O&M) costs are 25% of the initial capital*
- *The consortium wishes a return on investment (ROI) of 15% p.a.*

<u>Solution</u>: A 10-year annuity with present value 1 and interest rate 12% has equal instalments of 0,173. Hence, the total annual turnover of the system must be at least (17,3+25+15)% = 57,3% of the initial investment. Assume that this turnover would be earned during the busy hours only, i.e. during 100,000 fully-loaded minutes per year. The initial investment per (two-way) voice circuit being 2* 6,000 $, the required turn-over per transatlantic telephone circuit would be

12,000 $ * 0,573/100,000 minutes = 6,88 cent/minute. (5)

This allows a fair margin for further profit (above the 15% included above), even compared with the present prices for telephone calls across the Atlantic. *The Economist* has called this effect 'the death of distance'.

This crude estimate of 7 dollarcent per minute does not take some real risks of operators into account, such as competition from (newer) cables with even higher capacity and unused capacity during the busy hours. On the other hand, capacity will obviously also be reserved and/or used outside the peak hours. In particular, a very important transatlantic traffic was not foreseen in 1992, namely the popular use of Internet for World-Wide Web browsing, and this is not constrained by the narrow 5-hour time slots of the short joint business hours in North America and Europe.

Present transatlantic telephone tariffs are still much higher than the circuit revenue indicated in (5), plus the mark-up for the costs of delivery to end users through the national networks interconnected on both sides of the Atlantic Ocean. In other words, despite major recent international tariff cuts, high profit margins are still earned on such long-distance routes. This would explain why excess cable capacity can exist or even be further expanded without dropping below a reasonable financial return – such as the 15% p.a. assumed above. International competition, however fierce-looking, is not yet fully effective for telephone users, even on the most contested high-capacity routes. (This explains the computer freaks' early fascination with Internet telephony!) So long-distance telecommunications services still seem far away from the risks of commodity pricing, which have already affected the microelectronics and consumer electronics sectors seriously. Moreover, it should be emphasised that the cost of an optical subscriber line connecting an individual subscriber to the local telephone exchange does not reap any of the above economic benefits of massive collective use of each fibre. This explains the delays and difficulties in extending optical transmission to private users

5.3 Telecommunications Software: A Major Concern

Compared with VLSI and optical technology, the progress of software technology is much slower. This is a critical shortcoming, since the programming cost of software-controlled systems dominates the cost of all major network facilities except the fixed subscriber loops., The cost of switch software, especially, has now risen to a very substantial fraction of new telephone exchanges, typically 75-80% of their overall cost. Obviously, this lag is emphasised by the rapid performance/price improvements of electronic and photonic hardware discussed above. It reminds us of the fact that broad human creativity based on total *systems* understanding appears more difficult to muster than manufacturing efficiency and scientific expertise at the physical level of components and subsystems.

Moreover, it is often simpler to increase the reliability and availability of hardware systems by a suitable combination of the measures indicated in Table 1, items 7-12. One of the consequences of the crucial role of software in modern telecommunications systems is the increasing impact on society of inadvertent programming errors, occurring in exceptional operational situations almost never accounted for or even discovered in acceptance testing of new exchanges or other computer systems.

A modern intelligent network switch to be deployed in a node of a large network requires very complicated software. Checking the reliability of a telephone exchange under all feasible operational conditions is well-nigh impossible. Test programmes tend to concentrate on vital functions, plus more commonly observed errors and overload situations. Hence seemingly minor, but undiscovered software 'bugs' or viruses in telephone exchanges can lead to serious errors. These can propagate through the digital signalling system and its associated intelligent network (IN) and telecommunications management network (TMN) introduced in recent years (see Fig. 4). The impact on telephone and computer services and their underlying networks may be catastrophic. Such an error propagation occurred some years ago in New York, leaving all of Manhattan and Newark airport without telephone service for a considerable period. This resulted in a complete emergency, in which police, fire brigades and ambulances could not be reached, and incalculable financial losses on Wall Street and elsewhere occurred.

6. Conclusions

In this contribution, it has been argued that the major driving forces in public network designs for (tele-)communications are economies of scale and scope in the provision of services to the user community sharing a network. The individual user's behaviour proves to be of no importance, as long as it is not strongly correlated with that of other users. Scenario studies of user behaviour are therefore of relatively little use and can be limited to those particular trends which result in strong correlation of the collective behaviour of major user groups. The present extremely rapid reductions of the cost of modern information and communication technologies (ICT), such as VLSI-based microelectronics and optical cables, are causing the fundamental shifts in the planning and development of the network architecture and service capabilities of future telecommunications. This has been illustrated by a rational systems engineering approach which might, in principle, be extended to other network disciplines.

In practice, however, the cost drivers and performance features will of course differ between the different technologies of infrastructural networks. In particular, data transport in telecommunications networks enjoys a unique combination of performance features compared with other networks:

- transport and service delivery at the speed of light, unlike rail, road, or air
- transport over long distances without loss of quality, unlike electric power
- instant combination/copying/broadcasting/storage of information without loss of quality, unlike printing and publishing on paper
- cheap, lightweight terminals, unlike sea, air and rail transportation

It is clear that these performance features are driving the telecommunications markets towards global arrangements, not only of ICT applications - such as the Internet, intelligent networks (IN) and mobile GSM phones (see Fig. 4) - but increasingly of the institutional, commercial and regulatory principles for network operation.

References

[1] L.W. Couch II, *Digital and Analog Communication Systems* (5th ed.), Prentice Hall, 1997.

[2] A.F.G. Hanken & H.A. Reuver, *Inleiding tot de Systeemleer*, Stenfert Kroese , Leiden.

[3] J.C.Arnbak: "De Bremer Stadsmuzikanten. Een essay over informatie- en communicatie-technologie (ICT), in opdracht van de Overlegcommissie Verkenningen" (OCV), KNAW, 1995.

[4] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd ed., Addison-Wesley, 1994.

[5] B. de Pagter, Collegedictaat "Analyse II voor Elektrotechniek", TU Delft.

[6] L.G. Mason and J.P. Combot: "Optimal Modernization Policies for Telecommunications Facilities", IEEE Trans., COM-28, pp. 317-324, March 1980.

Power 4 Worlds

Scenario Analysis of the European Electricity Infrastructure

Rudi A. Hakvoort*, Selinde Biesheuvel*, Ivo Bouwmans*, Bert Enserink*, Sietske Herder*, Bob Paap**, Philip Vergragt*, Ad H.M. Verkooijen***

Delft University of Technology

* *Faculty of Technology, Policy and Management*

** *Faculty of Information Technology and Systems*

*** *Interfaculty Reactor Instiute*

Abstract

The generation of electricity is still almost entirely based on the burning of fossil fuels. Sustainable energy technologies (hydro, solar and wind power, biomass) play a minor role only. This study analyses several possible future scenarios. In two of them, fossil energy sources have mainly been substituted by renewable sources. The other major distinction between the scenarios is the level of decentralisation of the energy generation. In two scenarios the generation has been decentralised. Based on these extremes, three scenarios have been worked out. In a fourth scenario, a natural disaster largely disrupts the energy infrastructure.

In the first scenario, called Eden, the levels of both sustainability and decentralisation are high: companies and households generate their electricity by wind and solar power. In Oasis, solar panels in the Sahara desert are used for the central production of energy. In Cockaigne, the level of sustainability is as low as it is today and power generation is decentralised to a high degree: fossil fuels are burnt in small local generators. Finally, in Atlantis a large meteorite hits the planet and the resulting tsunamis wipe out the crucial facilities of the energy infrastructure.

The analysis shows that decisions affecting the technological infrastructural systems interact with the social, economic and political environments of these systems. Technological choices that work out well in one scenario may fail in another, while choices in the non-technical domain (such as liberalisation, globalisation and sustainable

development) may strongly influence the effectiveness of the technological options. Technological studies should therefore be complemented by social, economic and political analyses.

1. The European electricity system

April 1st, 2000. At 8.10 the largest nuclear unit of the Belgian power station at Doel is disconnected from the grid due to a transmission line failure. The pressurised water reactor unit is operating at full power and is feeding 1000 megawatt into the grid, when the automatic shut-down mechanism is activated. Suddenly, there is shortage of power in the European electricity network, at the moment Europe wakes up and electricity consumption rises fast. More power is consumed than is being generated, so the frequency in the network, normally stable at 50 cycles per second, goes down. This frequency change, however, is noticed at all power stations in Europe, and control mechanisms instantaneously increase the power output of all other power plants. A few seconds later the balance between production and consumption is restored.

Most of the power Belgium lacks due to the tripping of the Doel unit is now being imported from abroad. International agreements, however, state that each country should keep its energy supply and demand in balance, apart from contracted power and short-duration disturbances. Therefore, operators of Electrabel, the largest Belgian electricity generation company, immediately take action to restore the national power balance by increasing the generation of the other units in operation and starting backup gas turbines. Hardly anyone in Belgium notices that the entire power generation situation has changed in a few minutes. Maybe only a slight flicker was visible on television screens due to the switching actions in the high voltage network.

The electricity system of Western Europe may be viewed as a huge network of transmission lines with hundreds of generators located across the network and millions of customers. Practically all the electricity produced in the world today is derived from fossil fuels (natural gas, oil, coal), nuclear energy and hydropower. Renewable energy sources account for only a minor share, even though their contributions on a local level may be significant.

Almost all power units are connected to the electricity grid, which serves for transport of power from suppliers to customers. Some stations are built to generate power for a local installation and use the network only as backup. The network is generally divided into a 'transportation network', comprising the links with high voltages (up to 1.000.000 volts), the transmission lines and the lower voltage 'distribution network'. The transportation network serves as the highway for long-distance transport of power (up to several hundred kilometers), the transmission lines bring the power into the region with the energy demand, while the distribution network delivers power to the customers. Transformers are located between lines of different voltage levels to convert the power to the required voltage level. Both the transmission network and the major lines in the distribution networks have a ring or meshed shape, in order to enhance the security of supply.

In Europe, power is mainly generated by state-owned power companies, although more recently, private Independent Power Producers have mushroomed. Delivery of the power to households is taken care of by distribution companies, in some countries these are vertically integrated with the power generation industry, in other countries they are separate organisations. Power generation used to be perceived as a 'public service'. The power companies were held responsible for producing and delivering electricity at an affordable price, while guaranteeing security of supply and taking care of environmental constraints. Recently, cogeneration technology has opened the power generation business to industry. Small, decentral power units have been installed in large quantities which can cost-effectively supply local power needs at a lower cost

At present, the power market is liberalised in many countries. This calls for the 'unbundling' of transportation and supply. On the demand side, large consumers such as heavy industry have a special position in the market. They are able to negotiate special contracts, and sometimes even integrate power generation with their industrial processes. Except in Scandinavia, England and Wales, individual households are not allowed to choose their own power supplier. In order to monitor the tariffs these 'captives' pay, many countries have installed regulating authorities, to oversee the use of the transmission and distribution networks.

A central issue in the management of the physical infrastructure is the so-called technical dispatch. In Europe, the technical dispatch is presently performed per country in central control centres, where the balance between electricity demand and supply is monitored. Besides, actual control of the transmission grid is necessary to protect against overload, voltage and frequency control and to guarantee system stability. This implies the ability to maintain synchronism. The electricity grids of many European countries are interconnected under the Union pour la Coordination de la Production et du Transport de l'Electricité (UCPTE). All power production units in this system, approximately 250.000 megawatts, operate in synchronous mode, so that at each moment they run at exactly the same frequency. This enables easy power exchange between countries and enhances the stability of the system, since unexpected changes in energy demand or supply, e.g. due to incidents or line faults, are absorbed by the entire interconnected system. Mutual agreements decide on the amount of power each country should hold as backup for calamities.

To attract consumers, there is a world-wide tendency in the power industry to sell more than the simple product 'energy'. Companies add services to their product range, such as electricity carrying a green label, i.e. electricity derived from renewable resources, or electricity with supply guarantees. Additionally, the leasing of energy-related installations (e.g. boilers) is also becoming common practice, while some companies offer packages consisting of several types of public utilities, like electricity, gas, heat, water, telecom and waste services. All these novelties aim at increasing the added value of the simple product 'electricity' in order to bind customers to a company.

Presently, there certainly is a market demand for these packaged services, but the real incentive lies in recent EU legislation that forces the energy market to open up to competition. Liberalisation, and presumably also privatisation, of state-owned electricity companies will really reshape the playing field in the energy sector. Although those parts of the companies that exploit the transmission and distribution networks (a natural monopoly for the time being), will remain regulated by government, many countries have developed plans to leave the generation and sale of power to the laws of demand and supply in a free market. These restructuring initiatives will reshape the European energy market, and may influence future technological developments.

2. Technologies for the next decades

World energy demand is expected to grow at an annual average rate of between 2.4 and 3.1 percent up to 2010. Electricity demand in non-OECD countries will grow more than twice as fast as in OECD countries, although the total value will even then remain a factor lower than the energy consumption in the developed countries. Presently, more than 2 billion people do not have any form of commercial energy at their disposal. In all countries, the demand for electricity is growing relatively faster than the demand for energy. This increase is attributable to the versatility, transportability and controllability of electricity [1].

The main question is: how will the world cope with future electricity demand? What fuels will be used and which technologies will be applied? What will the system look like? Sustained increases in investments in new power generation will be required to meet the expected growth in electricity demand and the need to replace decommissioned plants. In order to explore future developments, it is good to glimpse first at projected technological developments already emerging in the present.

2.1 Advances in conventional electricity generation

Electricity generation, transmission and end use have benefited from substantial technology improvements. In the coming decades new technologies will be applied in power generation. Nevertheless, the search for cheap electricity needs to be reconciled with the desire for environmentally friendly and more sustainable ways of generating electricity. Emissions of greenhouse gases like carbon dioxide and air pollutants like nitrogen oxide and sulphur dioxide must be reduced significantly. New and improved technologies with a high potential for emission reduction and implementation times below 20 years are [2]:

- *clean coal technologies*, such as better coal cleaning technologies; advanced combustion processes giving higher thermal efficiencies, mostly requiring higher combustion temperatures and therefore the application of advanced materials for the boiler, improved post-combustion processes like advanced scrubbers, or even completely different conversion processes like fluidised-bed combustion, coal liquefaction or coal gasification that can be integrated with combined cycle gas turbine technology.
- *high-temperature gas turbine technology*, by converting aeroderivative turbines into heavy-duty industrial turbines. Advanced cycles will improve the efficiency of the conversion process up to 70%, while the promise of mass production of small turbines of 1 megawatt or less may enhance market penetration.
- *co-generation technology*, this is the simultaneous production of electrical and mechanical power and thermal energy from a single energy source. Cogeneration yields high total conversion efficiencies, but on the other hand needs heat or mechanical energy demand in the vicinity of the power generation site.
- *local power generation*, at the level of individual households or blocks of houses, e.g. based on small gas engines and integrated with the central heating system, may bring about a rearrangement of the complete electricity network. Individual homes may become self-supporting so that the local distribution networks may have a different function (local balancing of supply and demand only), while the large transmission lines may only be necessary for the transport of electricity during peak hours.

To cope with peak demand, either the electricity system must be dimensioned for the power peak with fast-starting peaking units available, or large-scale electricity storage must become economical (see below). Increased fuel diversification is needed to guarantee a higher level of security in the provision of electricity.

2.2 Advances in renewable energy technology

Renewable energy resources are not subject to depletion, because essentially the sun is the primary energy source: sunlight, winds, water flow and biomass. Most types of renewable energy can be applied at a local scale, so that decentralised energy production becomes a viable option. Since renewable energy is a low-flux resource, large-scale application is generally capital-intensive. The big advantage of renewable energy is that (apart from maintenance) variable costs are low, since no fossil fuels are needed.

The growth in the role of renewable energy will mainly depend on continuous technological development. Especially improvements in efficiency, cost and performance due to the application of new materials and mass production are foreseen. In principle, electricity generation based on renewable energy sources can be done in both grid-connected and stand-alone systems. In the long term, large-scale market penetration of renewable energy requires efficient energy storage facilities in the network. Some of the most promising technologies are [2-4]:

- *high-efficiency photovoltaic cells*, i.e. solid-state devices that convert sunlight directly into electricity. By application of new materials, conversion efficiencies may eventually reach 30 %, while cell lifetime goes up and costs go down. Solar cells can be utilised as small cells of just a few square centimeter or combined as modules into arrays. Photovoltaic energy is especially promising for remote power demands where grid connection is difficult (or expensive) to achieve.
- *liquid biofuels*, derived from biomass, e.g. methanol or ethanol, are suited as an alternative fuel for transportation. Many types of conversion processes are possible, ranging from biochemical processes like fermentation and anaerobic digestion, to thermochemical processes like pyrolysis and gasification.
- *hot dry rock*, which may yield a huge potential for geothermal energy. A pair of wells is drilled in artificially fractured rock deep below the earth surface. Water is circulated into the injection well and steam or hot water returns to the surface through the production well. In the longer term, it might even be technically possible to drill through the earth's crust to capture heat from magma bodies relatively near the surface.

Many innovative energy concepts are also being explored of course, such as wind energy systems at a high altitude, where the energy contents of the jet stream can be tapped, or installations running on ocean energy (tidal, wave or gulfstream energy).

2.3 Advances in transmission and storage technology

Several new types of technology may have a significant impact on electricity distribution and utilisation in the future. By decreasing line losses on the one hand and developing power storage systems on the other hand the limitations of electricity with respect to space and time disappear. Control and communication technology enhance the optimal allocation of transmission lines, improve the stability of the network, and may even open up electricity transmission to competition. Advanced technologies include [2,5-6]:

- *high-efficiency transformers, ultra-high-voltage transmission lines and eventually the application of superconducting materials*, in order to reduce power losses between the generation plant and the customer. Nowadays, these losses average at 10 %, but lower losses improve the operation of electricity grids. Ultra-high-voltage lines may even utilise 1.500.000 volts or higher.
- *long high-voltage direct-current (HVDC) links*, provide a more economic alternative to transmission of electricity over long distances, allowing power transport over distances up to a few thousand kilometers. Besides, they allow easy direct control of power over the line, while they block fault currents and the systems at both sides of the cable remain de-coupled in terms of frequency and stability.
- *electric storage*, which allows power to be generated and stored at a convenient moment, and can be released upon demand. In general, when storage is available, the operator may shift load from plants with high operating costs to those with lower costs. Apart from existing pumped hydro installations, electrical energy may be stored in batteries, superconducting magnetic devices (SMES) or compressed air systems [7].
- *advanced control technology*, e.g. with Flexible AC Transmission Systems (FACTS), that allow high-speed control, by which transmission lines may be better utilised without running the risk of stability problems. These systems even allow partial control of power flows in networks independent of transmission line impedances. Advanced power electronics are expected to enable precise control and tuning of all circuits, including gigawatt-scale power systems [6,8].

2.4 Coherent technological developments

Some of the technological developments mentioned have synergistic effects: The successful market introduction of one technology can be greatly enhanced by simultaneous application of some other technology. Some of these clusters are identified below:

- The implementation of long-distance power transmission by HVDC links or superconducting cables may stimulate power generation at locations where renewable resources are available or the highest efficiencies can be obtained, e.g. wind energy in offshore windparks, photovoltaic energy at sites with a high solar intensity and power generation from biomass at locations where biomass is abundant.
- The availability of facilities for power storage, either large-scale or small-scale, may greatly increase the market penetration of wind and solar power. Storage systems diminish the mismatch between power supply (dependent on e.g. the availability of wind and high solar intensity) and demand (principally dictated by the end-user characteristics).
- Finally, further decentralisation of the power supply may be enhanced by the development of small-scale cogeneration systems and improved electronics for distribution network control. Distributed generation is only a viable option when network stability can be safeguarded.

3. Power generation technology and the energy market

Power generation is a technical process that requires a physical infrastructure comprising power plants, transmission cables, distribution networks and control rooms. We define this to be the lowest system level. For the actual process it is required that the power plants are operated and the power flows are routed over the network in such a way that transmission losses are minimal and system stability optimal. These actions may be viewed as belonging to a second system level pertaining to the management of the physical infrastructure. On a third level, the market occupies the centre stage: products and services are offered for sale, traded and paid for. Actually, the market is not about joules or kilowatthours: people buy convenience in the form of light, propulsion power or heat. The third systems level therefore highlights products and services.

At all three levels the system interacts with the environment, i.e. the world, the government and the customers. Power generation on the physical infrastructure level consumes raw materials and yields power, heat, waste and emissions. Network operation and technical dispatch interact, for instance, with legislation and regulation, while at the level of products and services the energy system is influenced by market demand and market prices.

Parameters on all these levels eventually shape the form of the energy infrastructure. It is impossible to take into account all factors for the development of the scenarios, so it is best to limit them to the critical factors that will have the highest influence on the actual development of the future electricity infrastructure. These factors are limited to eight here for the sake of clarity and can be subdivided into three categories:

Category 'technology':

- *electricity storage:* when efficient electricity storage options are available, the number of alternatives for the future structure of the energy sector greatly increases (see section 2.4).
- *long-distance power transport:* as discussed above, economic and efficient power transport over longer distances may change the geographical distribution of power generation.
- *degree of decentralisation:* it is possible that economies of scale will play a decisive role in the development of new energy technologies. In this case, principally large installations will find their way onto the market: increasingly larger combined-cycle gasturbines, coal and biomass gasification units and large wind farms. The opposite force is decentralisation: efficient generation at a local level, either based on renewables (solar panels on roofs, local wind turbines, etc.) or caused by the further development of micro-cogeneration.

Category 'energy market':

- *energy price:* of course the energy price plays a central role in the development of the electricity infrastructure. Both the absolute value (e.g. the price of Brent crude) and the relative value (price of biomass compared to the price of natural gas) influence technological innovation and market penetration. It is expected that a structurally higher energy price leads to relatively more investments in energy systems with high capital cost and low variable cost (e.g. nuclear power and wind energy).
- *desired level of sustainability:* One of the uncertainties is the behaviour of the customer. Will a general demand for sustainable energy arise, or will the public get 'tired' of ecological and environmental arguments and simply choose the cheapest source of electricity?
- *energy policy and legislation:* Presently, there is a world-wide tendency towards liberalisation of public infrastructures. Corresponding legislation is being developed at national and European level and regulatory bodies are being installed. However, it is possible that in reaction to economic, political or environmental developments this tide will turn in a couple of decades.

Category 'external setting':

- *political stability:* This factor is closely related to the former criterion. The map of the world may change in 30 years, the balance between political powers may alter and new countries may rise to power and influence. Will world politics still be governed by Europe and North America in 2030? Or does Russia, the Middle East, China or South East Asia determine the political agenda? All these questions relate directly to the issue of political stability, a boundary condition for an efficient world energy market.
- *disasters:* Finally, environmental and technical disasters may reshape the energy sector. A rise in the sea level due to global warming or the approach of a new ice age, a nuclear accident or a meteorite impact may all have a dramatic influence on the world energy situation.

These eight criteria will serve as the basic parameters for the scenarios offered in the next paragraph. On the basis of these scenarios, possible future evolutions of the European electricity infrastructure will be explored. The aim is to get a realistic idea of the diverging options for the electricity system in several decades. In the scenarios the relation between technology and the political, economic and social environment will receive special attention.

4. Flashes of the future: four scenarios

Four environmental scenarios have been designed to investigate the electricity infrastructure in 2030: Eden, Oasis, Cockaigne and Atlantis. The parameters belonging to each scenario are shown in table 1. The scenarios are distinguished from the principally economic scenarios developed by e.g. the European Commission [9] in their technological focus.

Table 1. *The parameters for the four scenarios Eden, Oasis, Cockaigne and Atlantis. Parameters in italics form basic external factors influencing the described electricity infrastructure.*

Factor	Eden	Oasis	Cockaigne	Atlantis
Electricity storage	high	–	low	–
Long-distance power transport	low	*high*	low	–
Degree of decentralisation	high	low	*high*	high
Energy price	*high*	low	low	high
Desired level of sustainability	*high*	*high*	*low*	low
Energy policy and legislation	–	high	low	low
Political stability	–	high	–	*low*
Disasters	low	low	–	*high*

Eden describes Europe in 2030 with high energy prices and a strong demand for sustainable technology. It is further assumed that there is a high level of decentralisation (local power generation based on renewable resources) and that the technology for local power storage is economic and efficient.

Oasis describes an infrastructure in which efficient long-distance power transmission plays a crucial role: power can be generated at the economically optimal location, e.g. solar power in the Sahara and wind farms at offshore locations. Adequate large-scale storage facilities and, above all, sufficient political stability in all European and North African countries are necessary requirements for this scenario.

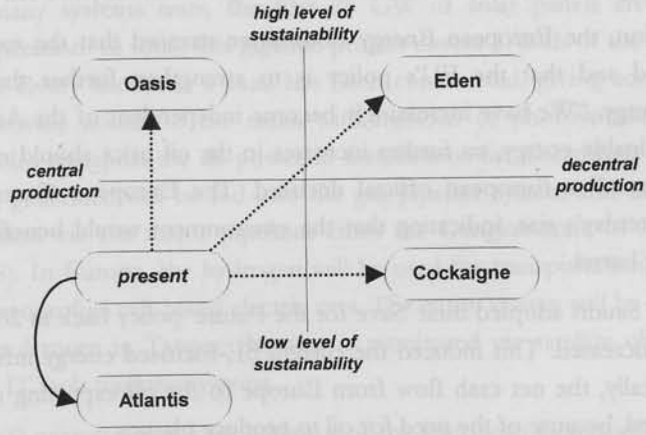


Figure 1. *Graphical representation of the four scenarios with respect to the parameters 'centralisation of production' (horizontal) and 'level of sustainability' (vertical). Atlantis represents a ghost scenario, and does not fit into this scheme.*

Cockaigne[1] describes a strongly decentralised, deregulated society with a low demand for sustainable technology. Since the energy prices are low, people do not mind wasting energy. Insufficient research on renewable energy and storage facilities has been done, so power generation is inefficient. Due to a lack of adequate governmental energy policy, people have invested in cheap, decentral power generation .

Finally, *Atlantis* shows the vulnerability of the European energy system to external factors. The scenario describes the consequences of a meteorite impact, knocking out most of Europe's energy infrastructure. This instantly transforms Europe into a politically

[1] The Land of Cockaigne is an imaginary Utopia in mediaeval legend where a life of luxury and idleness was possible. Cockaigne was a gourmand's paradise where the rivers flowed wine and the houses were made of cake and the pavements of pastry.

unstable continent on the brink of complete chaos. Consequently, everybody is thrown upon his own resources. This scenario, named after the sunk continent Atlantis, serves as a *ghost scenario*: although it is quite improbable, the effects would be immense.

4.1 Eden

The European Times, 29 February 2030:

The oil price again reached a record high yesterday, following the announcement of new export restrictions by the Saudi government. "This is completely in line with the policy we have followed since the turn of the millennium," a Saudi government spokesman said, "and there is no reason to assume we will change our plans." It is believed the other oil exporting countries will follow suit shortly.

In their reaction, the European Energy Committee stressed that the restrictions had been predicted and that the EU's policy is to strengthen further the position of sustainable energy. "We have increasingly become independent of the Arab oil supply by using sustainable energy, so further increases in the oil price should not worry any European citizen," a European official declared. The European Green Party even welcomed yesterday's rise, indicating that the environment would benefit from every Euro extra per barrel.

Ever since the Saudis adopted their 'Save for the Future' policy back in 2007, oil prices have steadily increased. This induced the current SE-focussed energy infrastructure in Europe. Ironically, the net cash flow from Europe to the oil exporting countries has hardly decreased, because of the need for oil to produce plastics.

Scientific American, February 2030

Europe's Energy Infrastructure: Divide & Rule

What causes the success of Europe's energy policy? The European Union has become almost independent of oil for energy purposes. Here in the US, however, companies are having a hard time dealing with their European competitors, because of the high energy prices. It is clear that the European strategy bears fruit. Jim Watson compares.

Travelling through Europe, Americans immediately see the differences. The huge number of windmills indicates that nearly every farm and every neighbourhood generates its own electricity. All roofs facing south are covered with solar cells or solar collectors. And there's more than meets the eye. Inside the houses, advanced mini cogeneration plants fill the energy needs of each household. Sophisticated energy storage apparatus is used everywhere.

All this has minimised the European dependence on oil. It was a controversial decision of the European Commission, prompted by the Saudi 'Save for the Future' policy, to bet on decentralised generation of sustainable energy. This SE policy now seems to have been the right choice. The European market, united since the Euro was introduced, has the extra advantage of independence. In this article, we will describe several of the technological breakthroughs that accommodated the transition.

4.2 Oasis

European Hydrogen Journal, October 2030

Sahara Desert Broods Europe's Energy

The Solar 1000 project in the Sahara desert finally reaches its completion. After performing many systems tests, the first 25 GW of solar panels are available for continuous operation. In total, this gigantic project covers 1/20th of the surface of the Sahara desert. Every kilometer a road has been constructed, giving access to smaller tracks that provide access to the racks with millions of photovoltaic solar panels. Electricity cables transport the dc power to installations for electrolysis of water. The hydrogen gas generated will be fed into the gas pipeline system and exported to the European Union via the major pipelines from the Conglomerate of North African States (CNAS). In Europe, the hydrogen will be used for transportation, principally in the new generation fuel cell-based electric cars. The entire system will be operated from a single control room in Tanger; the area is monitored via satellite observation and sophisticated IT fault tracking systems.

The Solar 1000 project was a joint initiative of the European Commission and several oil companies. The actual construction took more than 20 years. Due to the low energy prices and the low political stability in the CNAS countries, the project initially received a lot of opposition. The investment costs were high and mass production of solar panels remained too costly for a long time. Only in the '20s did the prospects improve. This was principally caused by the sky-rocketing demand for hydrogen as a source for clean energy, heavily stimulated by fiscal instruments and the ban on petrol-fuelled cars in many European cities. On the other hand, research into renewable energy systems was boosted due to the high ecotax, the green tax and the recently implemented water tax, to discourage the use of surface water for irrigation and cooling facilities. In the last few years, hydrogen-based energy systems have shot up like mushrooms, and last June, hydrogen replaced methane as the principal gas for energy application.

Solar 1000 has been built in the Sahara desert in North Africa, the only large region available with a high intensity of sunlight and a low population density. The few nomads living in the desert have been offered housing in villages at the border of the Solar 1000 system. Many of them are presently being trained as maintenance engineer for the gigantic field of solar panel arrays. Solar 1000 derives its name from the first large-scale solar power system, the Solar 1 built at Barstow, California in the early '80s of the last century. In contrast to the 10 MW of the Solar 1, the total power of the Solar 1000 complex with its 350 GWe equals the generated power of 1000 medium-scale power plants.

Advanced management systems have been developed for the operation of the gigantic system of power generation and conversion into hydrogen, storage, transportation and distribution. The entire system is controlled by 7 dialectic-logic transputer systems, each able to operate the entire system. Although the system is automatically operated, human supervision remains necessary, especially of the hydrolysis factories. Almost 10 % of the total investment went on guaranteeing safe transportation of the hydrogen. The system is highly protected against natural disasters and even terrorist attacks. Thanks to computer-controlled monitoring of the entire pipeline system, compartmentation and the network design of the arrays of pipelines, successful sabotage is nearly impossible. Moreover, political stability is safeguarded by involving local authorities in the project. Due to the project, employment prospects for the local population have increased considerably.

It is expected that finding new suitable areas for these large-scale solar energy installations will get more difficult. Environmental groups fiercely oppose any further plans for reshaping large areas with desert ecosystems. It is to be hoped that yellow energy will not come to a premature end in the same way that blue energy has: world-wide no more hydropower installations are being constructed, and existing complexes are being decommissioned in order to restore the former ecosystems. For the next 30 years, however, the Solar 1000 project is expected to guarantee Europe's energy supply.

4.3 Cockaigne

The European Times, June 26, '30, page 1

New Coal Pipeline for Dutch Dirty Power

Yesterday the new pipeline for the transport of cheap Russian coal to the Netherlands was put into operation. According to representatives of refineries and big chemical industries in the Netherlands, the main customers of the newly offered coal supply, new coal pipelines will be built the moment the Dutch government gives the green light. It is expected that the share of Russian coal on the Dutch energy market will increase rapidly to possibly over 50 percent within the next three years. British and Norwegian natural gas exporters are standing on the sideline disenchanted, gazing at their ever decreasing gas sales.

The European Times, June 26, '30, page 7

Coal success explained

The unexpected rise of coal springs from the liberalisation wave at the turn of the century. The most logical step following the liberalisation of the European electricity and telecom market was privatisation. The Dutch government even brought the national natural gas reserve onto the market. British and French power companies bought large parts of the Dutch energy market. Under market pressures, even the transport and distribution grids did not escape the liberalisation program of the progressive 'orange' government. After an apparently successful start it appeared that the maintenance and management of the distribution systems particularly should not have been entrusted to market parties. The government, however, was no longer able to turn the tide. International law did not permit further governmental intervention in the free market. The practical role of the government as investor in infrastructure was lost the moment the assets came into private hands.

As the service level of the distribution systems for the long- and medium-sized distances decreased, a demand for more trustworthy energy supplies developed. Eventually, this stimulated the breakthrough of independent decentral energy systems. Industry and large companies additionally chose to provide their own energy and chose their own fuel suppliers. As a hedging strategy, multi-fuel installations were commonly built. Due to the low price of coal the share for energy produced from coal has increased dramatically during the last twenty years. Meanwhile, new mining technologies have tripled the economically exploitable world stock of coal. Nowadays, cheap, high-sulphurous coal, worked in open-cast mining, is transported as slurry through pipelines to Western Europe by Russian and Ukrainian multinationals and forms an attractive source of energy. Even the petrochemical industry has started to switch from oil to coal as the raw material for their production processes.

Although it used to be an important political theme at the end of last century, consumers seem to have lost their involvement in the environment. The low energy prices in combination with the ample availability of fossil fuels, despite all predictions of depletion, have led to a substantial increase in the consumption of consumer goods. There no longer seems to be any justifiable need for energy savings, or investments in improving conversion efficiencies. Nor are these kinds of measures stimulated by government, since this would not fit in the modern, liberalised society anno 2030. Actually, the single remaining problem is the rising level of emissions from combustion processes, but it is expected that technological development will tackle that problem in due time.

4.4 Atlantis

CNN, February 1, 2030, 1 p.m. ET news service:

The Sky is falling! Two hours ago a meteorite with an estimated diameter of approximately three miles crashed into the Atlantic Ocean, 300 miles west of the African coast. Apocalyptic images come to mind when observing the fire and smoke emerging from the site of impact. Within a few days large parts of Europe will be covered by a dust and damp cloud, shadowing the vast dislocations caused by the collision.

Experts predict that tsunamis will cause havoc on the oceans and seas, flooding beaches, coastal settlements and port installations. Professor Atkinson of the Harvard Center of Natural Disaster Research said that in the hours and weeks ahead, all around the world the seas will behave like 'an inflatable pool housing a birthday party'. No pattern in the chaos can yet be predicted. Damage will be serious to extensive, depending on the local circumstances. Governments have started to evacuate their entire coastal populations.

Oil and Gas Report Daily, February 2, 2030:

The meteorite that plunged into the Atlantic yesterday, completely disrupted the global energy situation. The energy production capacity has suffered a severe blow. All offshore installations have been either destroyed or severely damaged.

The damage to the undersea pipeline systems is difficult to assess, but it is expected that many are ruptured. While landbased pipelines are mostly intact, the end-of-pipe refinery, handling and shipping infrastructure, generally located in coastal regions, is heavily damaged, especially in Europe and along the eastern coast of America. Contact with many high-capacity crude carriers on the oceans is lost. It is feared many will have shed their cargoes. The few carriers still intact have no place to go, since many ports are damaged as well. As a general picture, countries located near the shore are experiencing a disaster as has seldomly struck planet earth.

Oil and gas prices are soaring on the London International Petroleum Exchange. Since yesterday oil prices have increased tenfold. Analysts foresee that spotmarket prices will multiply by another factor of ten shortly. Consequences on the demand side are yet unclear. However, it is ironic that in the year the r/p-ratio of oil reserves reaches its historic maximum of 67 years, society must resort to the application of renewables like biomass and wind energy, thereby possibly returning to a form of individual energy supply as in the time before the industrial revolution.

The International Herald Tribune, February 7, 2030

The Greater Russian Republic has by law suspended all gas and oil export contracts, citing the internationally acknowledged *rebus sic stantibus* clause. All fossil fuels are declared strategic materials, thereby export is banned. According to officials, the dramatic change of circumstances caused by the meteorite hitting the earth, justifies the annulment of the contracts. It is yet unclear whether the contracts will be renegotiated or simply cease to exist. Compared to the coastal states, the energy reserves of the former Soviet Union are relatively untouched by the disaster. Since energy prices are expected to boom, the Russians seem determined to *make the most of it*, as a high-ranking Gazprom official noted.

The New York Times, February 7, 2030

The United States representative to the United Nations Security Council declared to be enraged by the Russian suspension of all energy contracts, which he labelled an *unparalleled act of egoism*. Any hope of facing this world-encompassing disaster in solidarity has been blown to pieces; the Russians have brought the world to the brink of a long period of political instability. `From now onward, every state is on its own, and only the very fittest will survive', according to the US representative. In a reaction the representative of the European Union endorsed these comments. She stated that the EU is perplexed. The Siberian-Rotterdam Interconnector, commissioned only two years ago, and the Moscow-Berlin Intertie, since five years the backbone in the Europe-Transural network, (a total investment of 60 billion Euro) were rendered worthless in one day.

5. Stepping back into the '90s

The scenarios sketched above are useful for formulating research questions. Strategic research on energy systems must cover all technologies necessary for the whole range of future worlds. It is short-sighted only to select the most desirable future, or the most likely one, and evaluate research programs according to that future. What is needed is research into the feasibility and impacts of alternative futures. A suitable method for performing this analysis may well be 'normative forecasting', sometimes called 'backcasting analysis'. Instead of trying to project present relationships forward into the future, backcasting attempts to assess the feasibility and impacts of different future scenarios and offers a method for exploring the implications of alternative development directions and their underlying values [10].

As a first conclusion in such an analysis, it can be observed that there is no one set of technologies that are robust in the sense of playing a significant role in each scenario. The major *power production technologies* in the scenarios range from large-scale conventional or renewable to extremely small-scale conventional or renewable. Parameters like energy prices, necessity or desirability of using sustainable energy resources and the developments in energy policy and liberalisation, may greatly influence the outcome of the evolution process. A related question concerns the function of the *power networks*, the transmission network and distribution grid. In Oasis the main function of these networks remains power transportation from the site of production to the customers, while in Eden and partly in Cockaigne, the networks principally serve to stabilise a highly decentralised system. In all scenarios, advanced *information and control technology* is required, although the exact functionality may differ in the various scenarios. However, further identification of the assumed social, economic and policy developments and necessary technology in the scenarios is required.

One of the central issues in all the described electricity infrastructures will be system control, comprising both technical dispatch and safeguarding system stability. In *Oasis* the main question is how the system will work, since power generated by solar energy depends heavily on the insolation. Actually, in this scenario the solar energy is used for hydrolysis, but when one chooses to feed the electric power directly into the grid, the question rises who offers the necessary spinning reserve for load following. A comparable issue plays a key role in evaluating *Eden*. The distributed utility concept completely undermines the current network philosophy with the automatic power-frequency control. Who is safeguarding system stability and network power balance? This problem culminates in *Cockaigne*, where nobody seems to be responsible for system services anymore. An interesting question is whether it is possible to replace active network operation with passive control systems based on power electronics and FACTS devices. At which stage of loosening control will the system collapse? Finally, *Atlantis* shows a world in which the European electricity system is completely smashed. The meteorite impact described is only one possible event causing the collapse. Many other environmental or industrial disasters may result in the same breaking up of the integrated European power system, while throwing back all cities on their own small-scale emergency systems.

A major drawback of the above analysis is that changes in electricity end-use have not been taken into account, although it can be expected that structural changes in the power sector will have a significant effect on power demand and may completely alter current network design strategies. Additional improvements in energy efficiency will decrease power demand, but new technologies in e.g. the transport sector (such as mass introduction of electric vehicles) may yield higher electricity consumption and a change of the load profile (e.g. flattening due to nightly charging of car batteries) [11]. It is sure that from the demand side perspective, the energy landscape will look completely different 30 years from now, which will impose other constraints on the electricity infrastructure.

It is likely that many infrastructural changes are driven by the tendency toward maximising individual freedom. At present, customers prefer to satisfy their own needs in a way that is defined by their local circumstances and preferences. Generally speaking, restraints imposed by a rigid infrastructure delivering only standardised goods in bulk quantities are no longer accepted without question. Technical, economic and social developments on the other hand stimulate innovation and the development of 'dedicated infrastructure'. A systematic characterisation of possible *product-market-technology combinations* may assist in analysing future infrastructural changes.

Each existing infrastructure, be it for power supply, transportation or telecommunication, has certain dynamics and stability characteristics. Stability comprises a so-called 'infrastructural lag', which means that due to its 'network externalities', the current infrastructure will persist for the next decades. On the other hand, on a micro-level a continuous stream of innovations is being injected into the current infrastructure, gradually changing the system characteristics and performance.

It is interesting to note that several infrastructures compete with each other. Electricity supply and gas supply are interchangeable: instead of gas, electricity may be applied for heating purposes; it is also possible to perform all necessary power functions with gas, either by a direct process (gas lamps, gas-driven refrigerator, etc.) or by local conversion of gas into electricity. It is possible that one infrastructure renders some other system redundant (the construction of the electricity grid led to the demise of pressurised-air networks), or that both remain in operation, each with its own advantages (like the rivalry between transport by train and car in the transport infrastructure, or the competition between existing telephone networks and telecommunication using television cables). Even 'infrastructure-less' systems are conceivable, such as completely decentralised (stand-alone) power generation, corresponding to wireless data transmission in the telecommunication sector. When one investigates the evolution of the electricity system, it is therefore necessary to glance at related energy infrastructures.

A central question is how one may influence this infrastructural evolution. Is it possible to direct the development of the energy infrastructure by promoting certain technologies? Or is infrastructure evolution mainly an autonomous process? In the first option, we may partly select our energy world of 2030; otherwise, we stand on the sideline only able to register the change in the energy landscape. But there is a lesson to be learnt even here. From our experience with the European power supply, criteria may be derived that enable us to design more efficient energy systems in regions where there is nowadays no adequate energy infrastructure. The lessons derived from the Western world should be translated into more efficient investments in energy systems in rural areas.

In conclusion, some issues that must be put on the research agenda are:

- Identification of the social, economic and policy changes underlying the sketched infrastructure evolution and the necessary technological developments.
- Characterisation of the range of functions power networks have in the four scenarios in combination with possible problems related to network control. Which innovative technologies might circumvent system collapse under the scenario conditions?
- Description of the relation between infrastructural innovation and current trends such as individualisation, liberalisation, globalisation and sustainable development. Which product-market-technology combinations have a high potential in each situation?
- Analysis of the interrelation between the electricity and gas infrastructures and their mutual interchangeability.
- Possibilities for influencing and directing infrastructural development. Which actors hold key positions, which instruments may be successfully applied, and which processes are completely autonomous?
- Development of criteria for designing an efficient energy infrastructure in regions presently without adequate energy supply, e.g. developing countries.

References

5. *World Energy Outlook 1996*, International Energy Agency, Paris, 1996.
6. *Energy Technologies for the 21st Century*, International Energy Agency, Paris, 1997.
7. T.B. Johansson, H. Kelly, A.K.N. Reddy, and R.H. Williams, *Renewable Energy, Sources for Fuels and Electricity*, Washington, 1993.
8. *The Future for Renewable Energy: Prospects and Directions*, EUREC Agency, London, 1996.
9. S. Linke and R.E. Schuler, *Electrical-Energy-Transmission Technology: The Key to Bulk-Power-Supply Policies*, Annu. Rev. Energy Environ, **13**, 1988, 23-45.
10. *Technologies for Tomorrow*, The EPRI Journal, **23** (1) 1998, 36-41.
11. S.M. Schoeng, J.M. Eyer, J.J. Ianucci, and S.A. Horgan, *Energy Storage for a Competitive Power Market*, Annu. Rev. Energy Environ., **21**, 1996, 347-370.
12. P. Moore and P. Ashmole, *Flexible AC Transmission Systems*, Power Engineering Journal, December 1995, 282-286.
13. *European Energy to 2020, A Scenario Approach*, European Commission, Directorate-General for Energy (DG XVII), Luxembourg, 1996.

14. J.B. Robinson, *Unlearning and Backcasting: Rethinking Some of the Questions We Ask about the Future*, Technological Forecasting and Social Change, **33**, 325-338, 1988.
15. A.B. Lovins, *Negawatts*, Energy Policy, **24**, 1996, 331-343.

14. L.B. Holtzman, [illegible] *the Future*, Technological Forecasting and Social Change, 50, [illegible]
15. A.B. Lovins, [illegible], Energy Policy 24, 1996, 331-343.

Scenarios for Waste Infrastructure in 2030

Gerard P.J. Dijkema*, Markus A. Reuter**, and Ewoud V. Verhoef**

Delft University of Technology

** Faculty of Technology, Policy and Management*

*** Faculty of Applied Earth Sciences*

Abstract

Scenario techniques were used to construct four images of a future waste-infrastructure. Combined with a systematic representation of current waste management and disposal techniques, these images were used to determine the scope of a decision tool for the design and management of waste infrastructures. This support tool must present transparent information on the infrastructure that is suitable for process engineers, technology managers, and policy-makers alike.

Apart from technological development, resource scarcity and final abatement of waste processing residues, public awareness and attitude were identified as the main parameters that determine the future context for waste infrastructure. 'Garbage land' represents a society where environmental issues have no priority, while in a 'Green Archipel' the environment is taken care of in everyday life. A society that resembles the 'Techno-Dream' is completely confident that a technical solution can be engineered for any problem. Finally, in 'Opportunia' problems and solutions come and go, and the people continue to behave opportunistically, no matter what happens.

The concept 'waste' was assessed, and redefined as 'an emerged quality of a substance that is qualified as waste if not used to its full potential. The material flows through production systems and waste management systems alike, cycles back and forth and forms the connection between industrial and waste management activities. The rearrangement and closing of material cycles opens the way to eliminate landfill of harmful residues and contributes to the conservation of resources. Therefore the material-cycle concept was adopted as the basis for system modelling. Material cycles represent a convenient method of abstraction to present system alternatives to decision-makers and demonstrate the interdependence between the availability and fate of all atomic elements.

The consideration, appreciation and adoption can be improved of less obvious but worthwhile options for improved process system design, by the apt supply of such information,. In addition, the fate of single materials becomes transparent, and the effects on future resource availability may be predicted. Ultimately, the objective is to improve the decision process by the supply of such information.

1. Introduction

The waste infrastructure is addressed in this paper. "Waste" commonly has a negative connotation: one thinks of garbage, rubbish, or maybe even dangerous or toxic material. Waste is a substance that one would like to dispose off, and one is prepared to pay some fee for the service. Apart from household garbage, there are many substances and objects that are considered to be waste, particularly in the process industry and manufacturing business. A substance, however, is a waste only when it is experienced as or labelled as waste. A producer, for example, may consider unwanted by-products 'prompt scrap' or 'production waste', whilst others regard these a potential resource, which is one of the economic bases of the recycling industry. Waste is a subjective concept, or rather a qualification of a particular substance or object, which does not vanish after disposal. The qualification, however, might change: what is considered waste today, can be a resource in the future. A more strategic notion, therefore, is that *a substance or object is qualified as waste when it is not used to its full potential.* Under this paradigm, any production process can be used for the transformation of waste, which vastly increases the alternatives for system design. In networks of industrial plants the waste of one plant can be the feedstock of another. Normally, in a transaction that concerns by-products neither of the two parties involved considers the substance flow a waste. If, however, the receiving party terminates its activity, the producer would immediately experience problems in disposing its by-products, and the substance would then be qualified as a waste product. *Waste, therefore, is an emerged quality of a substance or object.* Subsequent processing of any waste material, will cause the emerged quality to submerge again.

2. Waste management today – an infrastructure?

Generally, household waste is collected once a week in every municipality in Europe, and transported to some waste processing facility or landfill-site. Usually, these activities are the responsibility of the community authorities, and the general public expects that processing and possibly landfill occurs responsibly under the existing legal framework. As a consequence, a local tax is levied to balance the public expenses incurred. Companies that generate industrial waste also expect proper processing after they have paid some fee, be it to a privately owned waste management company, or special tax levied by the public authorities. In both cases, the service 'waste abatement' is considered to be a public good that must be available to all on equal terms. In this respect, the waste management sector resembles the electricity, natural gas, water, and telecom sector.

The public task of the waste management sector can be described as 'responsible waste abatement', which includes environmentally sound operations, preservation of fossil fuel, and recovery of valuable resources. A general definition of "infrastructure" usually refers to the underlying foundation or framework of basic services, facilities and institutions upon which the growth and development of an area, community or a system depend. Infrastructure, therefore, includes a broad spectrum of services, institutions and facilities that ranges from transportation systems and public utilities to finance systems, laws and law enforcement, and education and research (Larimer, 1994). It may be seen that the waste management sector falls into the category of public utilities that, apart from waste management, includes electric power, natural gas and water systems. Waste management is one of the public infrastructures that are based on a specific type of physical infrastructure to provide the goods or services.

A typical waste management system comprises collection, transportation, pre-treatment, processing, and final abatement of residues. Various types of waste can be collected separately (Figure 1). Transport can be to some local or regional pre-treatment facility, or directly to some regional or national processing facility, such as a waste incineration plant. Local or regional pre-treatment may include compressing, sorting, separation, drying, storage and so on.

As indicated above, the primary 'raison d'être' of a waste infrastructure is waste abatement, which can be characterised as a service to the general public or each individual household or company. Additional services include reuse, and recycling. In addition, waste processing presently yields products such as electric power, steam, distilled water and compost or synthetic crude oil (Sas et al., 1994). Unavoidably, however, part of the waste yields residue, which more often than not must be classified as 'hazardous waste'. Additional processing before final disposal of such residues is often mandatory, and transportation or handling restrictions may apply.

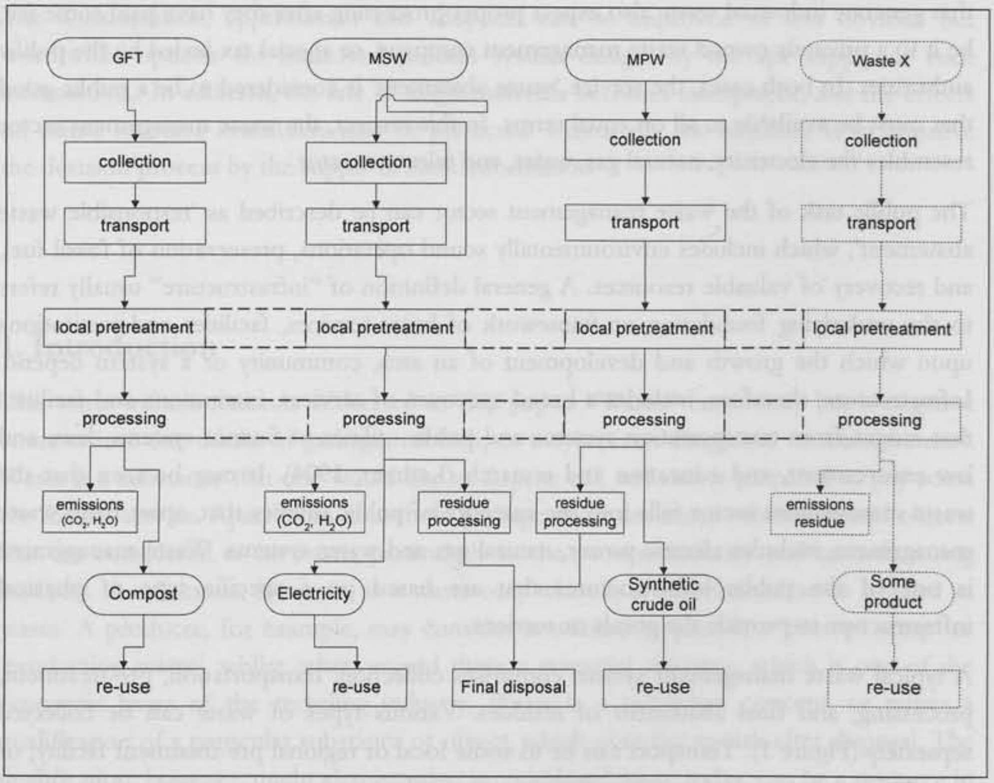


Figure 1. *A typical waste management system; Collection of organic waste (GFT), Municipal Solid Waste (MSW), and Mixed Plastic Waste (MPW); Other waste streams (Waste X).*

In the DIOC Scenario Workshop, apart from the physical infrastructure two additional system elements were identified, viz. products and services and management and control. Products and services determine the scope of the infrastructure, both on a practical and strategic level. The second element comprises the actors and systems that control and manage the infrastructure.

Table 1. *System elements waste infrastructure (DIOC Infrastructures, 1998).*

System	Surroundings
Products & services	
Municipal waste abatement	Supply
Electricity, hot water	Demand
Secondary materials	Market: Europe
Waste disposal	

System	Surroundings
Management & Process control	
Processing	Crime
Legislation, policies	"culture of mutual arrangement, consultation"
Collection	Hybridisation of products (e.g. micro-electronics in toys)
Preliminary treatment	Development of products
IT of detection and tracking	
Transport facilities	
Physical infrastructure	
Processing facilities	
Transport (tubes)	
Incineration	
Plants / facilities	
Link to emissions and sewer system	

The waste infrastructure can be defined as the underlying foundation on which the waste "market" and its development is based. Analogous to any other market, the waste market is greatly determined by the structure of supply and demand. The operators of waste reuse, waste recycle, waste incineration and waste disposal facilities create waste demand. Until recently, waste demand was not perceived as such, because waste management was completely organised as public utility with an annual fixed cost that was not correlated to the actual amount of waste processed. With private companies entering the market place, and publicly owned waste facilities operated as individual profit-centres, waste demand became apparent and competition started.

Figure 2. *Dutch Campaign Poster (Postbus 51). It contains the message that you must start improvement of the environment.*

In Germany, for example, in the early 90's incentives were created for the separate collection of plastic waste, the 'Duales System Deutschlands', or DSD; because the legislation and collection systems became effective prior to the processing facilities, stocks piled high, and the market was supply oriented, which led to very high processing fees. In the Netherlands, over-capacity recently threatened the MSW market, and the planned construction of a number of MSW incineration facilities was cancelled. When liberalisation of the waste market proceeds, one can expect competition by gate-fee, and competition between both public and private owners of waste processing facilities.

Waste supply includes waste generated by industry, households, the energy sector etc. Waste supply is largely considered to be an exogenous factor. It can, however, be influenced via life-styles and advertisement such as the Dutch 'Postbus 51' (P.O. Box 51) campaigns. An illustration is given in Figure 2, which is used in the Dutch Government's campaign for a better environment.

The change from a completely publicly owned sector to a public-private sector where privately owned companies carry out increasingly larger parts of the sector functions is an important similarity between the waste management sector and the other infrastructures. In addition, this sector is in transition with respect to its orientation and scale-of-operations, which is shifting from strictly regional or national to a truly international setting. In the Netherlands, for example, the Dutch Waste Management Council has recently changed its focus from Dutch waste management within the national borders to the realisation of proper waste abatement within the EU framework.

The Management & Process control element can be addressed at 4 levels: management and process control at local, national, European and world-wide level. Important actors in the waste infrastructure are:

- European Commission
- International Industrial corporations
- Waste management corporations
- Waste processing facilities
- Transport companies
- Green movements
- Consumer organisations
- National Governments
- AOO & comparable organisations
- Provincial Authorities
- Utility companies (electricity, heat)
- Municipal Authorities

The waste management sector differs from the other infrastructures in that it is not 'leiding-gebonden' (i.e. pipeline or grid based). Hence to some it seems that it does not present a clearly defined network; however common waste management systems can be represented as multi-node networked systems, where the nodes are clearly defined, because they consist of the physical processing facilities. In the Netherlands, the prime examples of these are the municipal waste incineration facilities, the dimensions of which are similar to central electric power plants. The links are the transportation routes, and collection stations or pre-separation or feed-preparation stations comprise sub-nodes. The network structure is not directly apparent because the common mode transportation is by road transport, often in combination with collection and pre-treatment. In designing a waste management system, degrees of freedom exist in the modes of transportation between processing facilities, in the logistics' organisation or (criminal) disorganisation, and in the selection and design of individual facilities.

Another unique characteristic of the waste management sector is that it deals with a great many different entities, whereas all other infrastructures must only deal with a single principal entity (electricity, natural gas, water, bytes). 'Waste' is an aggregate term for a large variety of materials.

In addition, waste processing includes chemical transformations, additions, and physical separation. In the electricity infrastructure, for example, only at the time of generation is there a chemical transformation of physical mass. In natural gas transportation and distribution, usually what goes in comes out. Only at the time of consumption, some emissions are generated by combustion. In telecom, as well as in the other infrastructures, there is a relatively small uptake of electricity to maintain the system's operation. In waste management, transformations can occur almost everywhere in the chain, and they are often unique to a particular type of waste. As a consequence these represent a broad spectrum of chemical and physical operations.

Finally, both the waste management and water sectors differ from the other infrastructures with respect to the publicly accessible good or service that they provide. In waste and water management, the physical entity presents stringent limitations on where the service commences all the way to where it ends.

3. Waste management in the European Community

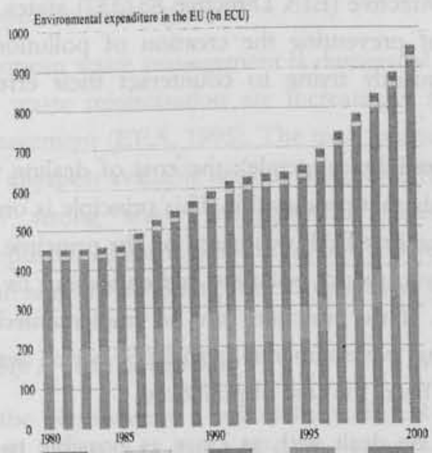


Figure 3. Development environmental expenditures in EU12 (Source EEA, 1995)

The total environmental expenditure in the EU12 was about 63 billion ECU in 1992 (see figure 3). Expenditure on waste water treatment measures accounts for the largest share of total environmental expenditure (approx. 50%), waste management accounts for second largest share (approx. 33%) (source: EEA). These figures give an indication of the importance of waste management in the EU environmental program.

3.1 Environmental policy

The first European environment program of the European Union (the

Environmental Action Program or EAP) was adopted in November 1973, as a follow-up to the 1972 Stockholm Conference. The EAP has been updated and extended every five years since, and the implementation of the fifth EAP for the EU is still ongoing (UNEP, 1997). The strategy in the Fifth EAP is twofold: high environmental standards set through regulations for almost all pollutant emissions, discharges and wastes are combined with positive incentives for industry (UNEP, 1997). In this way reduction of the impacts on the environment is not only achieved via end-of-pipe solutions, but also through the promotion and development of new and cleaner processes, products and techniques. Eastern and Central Europe followed the EU in 1993 (Lucerne Conference, Switzerland): 50 environmental ministers endorsed the (short-term) Environmental Action Program for Central and Eastern Europe (EAP/CEE). There are four key principles in Europe's common strategy for better waste management (EU DG XI fact sheet):

1. *Precautionary principle*: if there is a strong suspicion that a certain activity may have environmentally harmful consequences, it is better to act in a legal way before it is too late, rather than wait until scientific evidence is available (Borge, 1995[8]). In other words the industry (or any producing body or person) should prove that its products, wastes and processes are safe (i.e. non-hazardous to man and environment) before allowing releases of products and waste into the environment.

2. *Prevention principle*: prevention of waste generation by taking action at the source. The principle is not as far-reaching as the precautionary principle. The Third Environmental Action Program is focused strongly on the prevention principle and the Environmental Impact Assessment Directive (EIA Directive 85/337) states that "the best environment policy consists of preventing the creation of pollution or nuisances at source, rather than subsequently trying to counteract their effects" (Borge, 1995).

3. *"Polluter pays" principle* (or producer responsibility principle): the cost of dealing with waste should be met by the person or body that produced it. This principle is one of the cornerstones of environment policy: the first EAP was based on the principle that charging the polluters will encourage them to reduce pollution and endeavour to find less polluting products or technologies. This principle can be implemented by charging the polluters, and by imposing environmental standards, since setting standards helps to ensure that the polluter bears the cost of pollution.

4. *Proximity principle*: waste products should be dealt with as close as possible to the source. The "self sufficiency" principle - according to which the community as a whole and member states individually are self-sufficient in the disposal of waste rather than exporting it - is a derivative of the proximity principle.

Committed to these principles, to the concepts of sustainable use of resources and of minimisation of environmental damage the European Union developed an extensive range of legislative instruments intended to promote and harmonise the national legislation on waste (source: EEA).

3.2 Current status of waste management

National waste legislation is most developed in Western European countries such as the Netherlands, Germany and the Scandinavian countries. Some Central European countries are beginning to adopt similar approaches. A major incentive for these countries is to become accredited to enter European Union accession process. Waste legislation, however, is still poorly developed in most other Central and Eastern European Countries (CEEC) and in the New Independent States of the former Soviet Union (NIS) (source: EEA).

Dutch waste management

In Dutch environmental policy, Lansink's hierarchy (Lansinks Ladder) has been a leading principle (Wet Milieubeheer, art 10.1). This is a broadly recognised "rule of thumb" for proper waste management, also known as cascading waste management. The ranking prescribes that it is best to reduce the generation of waste at the source, to reuse what cannot be reduced, to recycle what can not be reused, to incinerate or compost with source recovery, and finally to landfill the remainder.

European waste management

European waste management is dominated by waste disposal, although waste prevention and waste minimisation are increasingly recognised as a preferable solution in waste management (EEA, 1995). The most prominent form of waste disposal is still the oldest and cheapest available option: landfilling. Recycling is increasing, particularly in countries with strong waste management infrastructures (EEA, 1995). Incineration (and composting/digesting) with resource recovery is also gaining ground. Reuse is limited to a small range of products, for example beer/soft drink bottles and computers.

A shift in waste management

In the Netherlands a successful shift has been made from Landfill to Incineration of MSW (Municipal Solid Waste): per 1 January 1996 Landfill was forbidden in the Netherlands. This was a result of a long process that started with the Lickebaert scandal: milk produced by cows grazing in this polder had a high dioxin-content, originating from a nearby MSW-incineration facility. As a consequence, all old incineration facilities in the

Netherlands came under suspicion, and they were rapidly shut down: because landfilling capacity at the time (mid 80's) was not sufficient, there was a great urgency for managed construction of new facilities. The Dutch Waste Management Council (AOO) was formed to combine the forces and negotiate between municipal, county, and national authorities. The planned construction of new facilities was completed in the early 90's. The capacity installed was based on supply-side projections that were reviewed on a yearly basis. At the same time, however, numerous initiatives were launched to increase recycling, the separate collection of paper and glass being the most successful programs to date. Waste prevention programs were also started, and a number of projects focused on the reuse of plastics. Finally, EU regulations forced the Dutch government to abandon the 'zelfvoorzieningsrichtlijn' (self-sufficiency principle). This implied that the Dutch government can no longer enforce treatment of Dutch MSW in Dutch facilities, but rather that Dutch waste management must compete in a European Market. This implies that some of the centrally planned incineration facilities may be faced with a shortage in supply.

4. Images of waste infrastructures in 2030

Kaleidoscopic development of waste management infrastructures throughout the world appears to be likely. Similar to other infrastructures, the scenario that materialises per nation, subcontinent or zone may well determine the welfare of such regions. In a scenario workshop (DIOC, 1998) the participants developed four scenarios to visualise possible realisation of the 2030' waste infrastructures (see Table 2).

A scenario is an image of a possible future, i.e. it is a prediction of a definite future with stated conditions, and it is neither branching nor does it have alternatives. In order to anticipate a future system or infrastructure successfully, one must compose a set of scenarios that covers all extremes of the range of possible scenarios. Whether any scenario become real depends on numerous factors and system parameters.

Table 2. *Scenarios for waste infrastructure (DIOC Infrastructures, 1998).*

Factor	Garbage land	Green Archipel	Techno Dream	Opportunia
Public awareness	No interest in a sustainable future, "that will not happen in my life, so.."	Sustainable, new age, back to nature	Aware of the environment and its interaction with industry	Opportunistic Free-market reigns
Green movements	Weak, no nonsense	Prominent	Prominent	Splintered, shivered
Moral code, standards,	Business as usual easy, supple,	Strong ethics	Ethics & no-nonsense	No-nonsense

Factor	Garbage land	Green Archipel	Techno Dream	Opportunia
Incentive–penalty structure	Flexible, marginal	Incentives offered, but no penalties	Stringent (both incentives and penalties)	Failing
Battling of criminal organisation	Lack of public perception	Virtually absent;	Continuous	Incidental
Welfare	Low	High	High	Large differences
Population size		Stable	Growth	
Calamities, accidents	Many	Regularly but few	Regularly and many	Many
Waste disposal technology	No development	Increase	Break through	Selective/per stream
Design & recycling/reusing	No development	Increase	Break through	Selective/per stream
Investments in waste sector	None	Strong increase	Extremely high	Selective/only profitable sectors

The four images must be considered extreme realisations of the development of waste management infrastructure. They range from an infrastructure for a society where the environment is not considered to be important (Garbage land) to a Green Archipel where sustainability is the leading principle; from a society where effective technology is complemented by an effective societal system (Techno Dream) to a completely opportunistic society (Opportunia) where a minimum of regulation is effective.

4.1 Garbage land

The Garbage land scenario can be found in those countries that have more urgent problems than environmental problems, such as war, famine etc. "A hungry man is an angry man" (Bob Marley, Them Belly Full [But We Hungry]). As a consequence little priority is given to environmental problems, and public interest in and awareness of environmental issues is low. In Garbage land, waste management, if there is any, is fully directed towards waste disposal techniques: landfill and land spreading of organic waste, and incineration. There is little or no development of new technologies because of the lack of interest, lack of priority and lack of funds. Investment in the sector is low, as any money available is allocated to more urgent problems. Waste prevention is virtually non-existent. Welfare in Garbage land must be low, when environmental problems are considered relatively unimportant or remain unrecognised. The people's non-sustainable behaviour results in the depletion of fossil fuel, scarce materials, wood and (unspoiled) nature. The resulting scarcity provide ample incentives for criminal organisations.

4.2 Green Archipel

Around the world various tribes still live in harmony with nature. In South Africa and Namibia, for example, the Khoisan, also known as Bushmen or Hottentots, form a small tribe. The Khoisan partly have preserved their hunter-gatherer tradition that stretches back thousands of years, and their cultural heritage. The Khoisan maintain a complex relationship with the environment.

At first sight these primitive people seem to have very little technology, they only use very simple tools. There are very few new developments. If there are any they are adaptations to a (partly) new environment due to calamities, changes in the climate (long periods of drought etc.) or migration. Waste disposal, design for recycle and investments are usually unknown concepts. One could argue, however, that their technology is all but primitive, it is perfectly adapted to earth assimilation capacities: all the tools they use and all the products they make are still part of natures metabolism. They have succeeded in creating a cyclic or sustainable society, so waste management, design for recycling and the large investments are not necessary.

The Green Archipel differs from the way tribal people live in that waste management and waste technology play a crucial role. Waste disposal technology and design for recycling are a constant focus, and investments in the waste management sector have top priority. Public demand has shifted from 'consumption-oriented' to 'needs-oriented' in harmony with nature. As a consequence, waste generation is low compared to the former consumption society, a trend that had been augmented by the stop in population-growth that resulted from the shift in public awareness. The development and influence of criminal organisations is negligible because of lack-of-incentive.

4.3 Techno Dream

The Techno Dream scenario resembles the Green Archipel scenario in appearance, because in the Techno Dream the people have achieved a high level of welfare, possess a high level of environmental awareness. In addition, the Green movement is prominently present, and priority is given to responsible waste disposal and design for environment.

In these two scenarios, however, people have a completely opposite attitude towards the earth. In the Techno Dream earth's capacities are adjusted to meet mankind's needs by means of technology, whereas in the Green Archipel people adjust to earth's capacities.

In Techno Dream the people are very much aware of environmental problems such as depletion of earth's natural resources and reaching earth's maximum waste and pollution assimilation capacity. At the same time, however, they are convinced that new technologies will solve these problems and will create a sustainable society.

4.4 Opportunia

"Opportunia" is a country in which the public perception is opportunistic: guided or influenced by the circumstances, it is not a land of opportunities (i.e. a land in which everything is possible). In describing Opportunia's public conscience, one could think of statements as "Carpe diem", "Don't do today what you can do tomorrow " or " We'll see how it goes". There is little or no structure; everything is taken as it comes. There is hardly any planning and anticipation. One could think of countries that are building up their economies and do not have a clear overview on the environmental problem, but it also resembles free-market economies where 'increasing shareholder value' is of major concern.

Welfare is not fairly divided over the population; there are large differences between the rich and the poor; because this is land of the opportunists, everybody makes the best of the situation, and cares little about others. Waste disposal is not such a big issue, due to the failing environmental framework, the splintered Green Movement and the opportunistic public mentality. Environmental problems only are dealt with at the moment they are encountered. The use of insecticides, for example, results in better crops without damage, and is thus regarded as a good opportunity. The moment that people get sick from eating the crops and fruit sprayed with insecticides, the public opinion changes and the problem is dealt with by some that regard this problem-solving as yet another opportunity.

In Opportunia the emphasis in waste disposal is on the removal of waste out of the public sight: waste is collected because of the stench and is processed via the cheapest available option. Problems such as the destruction of the ozone layer or the greenhouse effect get little or no attention. Design for recycling is only used if the problem can not be solved with waste disposal or if the manufacturing of a product causes short term environmental or health problems. In other words long term and non-visible problems are given little attention. Criminal organisations thrive, as there is ample opportunity for them, and there is hardly any systemic legal framework or enforcement.

4.5 Reflection on the scenarios

In the scenario-analysis in the workshop, the factors involved were ranked according to importance, their relevance to the development of the system, and their uncertainty. The rationale of this last criterion is that the more uncertain a parameter, the more likely it is its alternate developments will yield different development. Table 3 gives an overview of the results obtained.

Table 3. *Important factors and their relation to the infrastructure*

Factor	Relation with Waste infrastructure
Public mentality and awareness	definition of well being, environmental conscious and sustainable behaviour
Green movements	definition of waste
Paradigm	determines incentives & sanctions, chances for calculating actors
Welfare	proportional (?!) with waste production
Population size	proportional (?!) with waste production
Calamities, accidents	due to waste production and new (or not well understood or controlled) technology
waste disposal technology	more efficient and effective disposal
Design & recycling/reusing	less waste
Investments in waste sector	more capacity, diversity and quality

This set can be reduced to two main factors:

1. Public awareness and public perception
   Society and culture; Green movements, moral codes, standards and incentives, population size and even concepts of welfare and well being result from these.
2. Technology
   The development of technology, the economy and its influence on society.

5. Waste management technology development

In this context, waste management has been dominated by linear thinking: waste is an inevitable end product that has to be disposed of in such a manner that the impacts on the environment are minimised. Technology development for waste management has largely been focused on the transformation of particular types of waste. The Dutch Waste Management Council (AOO), for example, has presented a selection of promising waste disposal techniques (AOO, 1995).

***Table 4.** Selected waste disposal techniques, (AOO, 1995).*

Waste stream	Waste disposal technique	Costs; Remarks
	Roaster incineration	F 225,-/ton, - quality ashes
	Fluid bed incineration	F ??,-/ton, - limited scope waste
	Pyrolysis-incineration	F 300,-/ton, - quality of ashes, reliability
Combustible wastes	Pyrolysis-gasification	F 270-300,-/ton, - reliability
	Separation-composting-incineration	F 250,-/ton, - applicability streams
	(wet and dry) separation-digesting-incineration	F 250,-/ton (wet)
	Separation-digesting-pyrolysis	
	Separation-digesting-gasification	
	Separation-digesting-incineration in a cement plant	
	Selective separation-incineration	F 240,-/ton
Non-combustible wastes	Landfill	F 135,-/ton --> f 175,-/ton
Partially combustible waste streams		
	Pyrolysis and co-incineration in a coal power plant	
Wood (old and left-overs)	pyrolysis and co-incineration in a powdered coal power plant	
	Incineration in a fluid bed oven	F ??,-/ton, - quality ashes
	Gasification	F 30-40/ton, - cleaning flue gas ?
	Gasification	F 300,-/ton costs
Plastics	Feedstock Recycling	F 265,-/ton (200 +210 t/a vacuum residue)
Organic wastes	Composting	F 100,-/ton
	Digesting	F 120,-/ton

Such a listing conveys two messages to policy-makers:

1. Waste is an inevitable end product of industrial activities and consumption.
2. One can rest assured that a lot is already being done.

While the list appears to be rather extensive, it is of course far from complete, as in the selection presented two criteria were used:

- the techniques (are expected to) achieve improved performance compared to present waste disposal techniques (quality of emissions, residues, costs, and process reliability)

- the techniques will be commercially applicable within the next ten years.

The major categories are the options found for combustible waste, non-combustible waste etc. This illustrates the focus on single *technologies* rather than waste management *systems*. As a consequence, usually one waste problem is solved at a time, and more often than not a new waste problem emerges, because most treatment techniques yield residues. A municipal waste incinerator, for example, produces flue dusts that contain Zn, Pb, Hg, Cd, Na, K etc., which poses a problem. A similar situation exists in recycling technology, were recycling *systems* only have emerged recently. We conjecture therefore that a great many opportunities are yet to be explored in technology development for waste management. At present, integral resource management, for example, is addressed almost exclusively by academics (e.g. Reuter, 1998), while in policy development waste management per se prevails.

The idea of a linear economy (Figure 4) with waste as an end product is rather common, nevertheless it leads to depletion of our natural resources (extraction) and pollution of the environment (disposal).

Figure 4. *Schematic representation of a linear economy or extract and dump economy: Materials and energy are extracted, processed, used and dumped in a linear flow into, through and out of the economy (after Gertler, 1997).*

With time, however, such linear chains of activity will reach some limit, and eventually come to an end. In this view, waste management is reduced to an endeavour to minimise the environmental impact of the disposal of this useless end product.

As stated above waste must only be considered an emerging attribute of a resource. This is not an entirely new paradigm, as in many industries some processing of a waste becomes economically and ecologically feasible, and the substance or object previously labelled 'waste' changes into a 'resource'. In the development of the petrochemical industry, for example, this has been an ongoing process. The production of ethylene out of naphtha, for example, can only be profitable if the other products of this operation, notably propylene, C_4's and aromatics can also be sold at a reasonable price. Optimisation of naphtha crackers aims at maximising the yield of valuable products, while at the same time minimising the production of waste-gases that can only be used as fuel (Chauvel, 1989). Once one accepts the paradigm presented, the necessity of an integrated resource system is obvious. It immediately follows that under this paradigm, waste management and production form a single system. If decisions are made on either of them, the consequences for the complete system must be visualised and taken into account. The new paradigm thus forms the foundation of a movement towards a cyclic economy.

In a cyclic economy (Figure 5) the material cycles are closed. Individual processes are all connected: waste of one industrial activity can be the feedstock of another or a fuel for heating and energy generation or a new product.

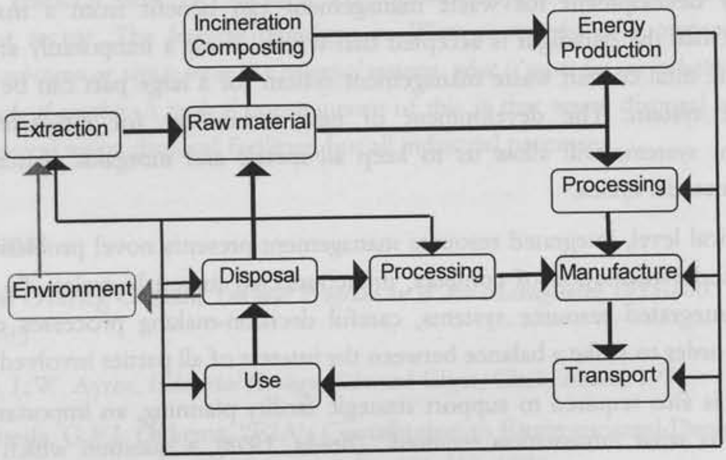


Figure 5. *Schematic representation of a cyclic economy.*

NOTE: A cyclic economy is theoretically an economy in which the material cycles are closed. This would imply that no material is released into the environment in any of the processes. In the real world such a system is unreachable and unfeasible. In case the materials released into the environment can be assimilated and somehow re-enter the cycle, however, without causing environmental damage, the system can also be regarded a cyclic economy.

This resembles the way nature's ecosystems are built up: a network of integrated processes that form cycles. The holistic approach towards industrial ecosystems (e.g. Ayres, 1996) and industrial ecology (Graedel and Allenby, 1995) use this analogy to capture and build a strategy for sustainable development. We use the concept of material cycles to effectuate this strategy for a waste infrastructure.

5.1 Management of material cycles

Waste management generally is treated separately from waste generation. In this respect, the waste management sector resembles the water sector, where fresh water supply has also been decoupled for a long time from wastewater treatment. In our present work we choose not to isolate the waste management infrastructure from the systems that *generate* waste. Rather we combine the two in the concept of material cycles. Our argument is that we consider 'waste' to be only an attribute of a physical resource similar to the attribute 'primary' of other resources. When we consider the complete network of systems that use primary and secondary resources (or waste), some problems in the management of material cycles that manifest themselves in the processing of waste can actually require a solution in the primary production system. Finally, the concept of material cycles offers a convenient method of abstraction to model the broad spectrum of technologies involved.

Technology development for waste management can benefit from a material cycle approach. Once the paradigm is accepted that waste is only a temporarily attribute of a resource, the total current waste management system for a large part can be labelled an end-of-pipe system. The development of new technology for integrated resource management systems will allow us to keep all metals and inorganic material in their respective material cycles.

On a technical level, integrated resource management presents novel problems as to the modelling and visualisation of complex, interconnected materials cycles. To implement prototype integrated resource systems, careful decision-making processes need to be designed in order to strike a balance between the interest of all parties involved.

As the tool is also required to support strategic facility planning, an important aspect is also "when is what information required" (Breda, 1998), a question which has to be narrowed in view of the current transition of the sector. The associated modelling requirements are not trivial. The industrial processes involved are usually interconnected, which results in very complicated and non-transparent networks of industrial activities; because of their complexity, these systems are usually modelled in several levels. An example of such a layered structure is depicted in Figure 6.

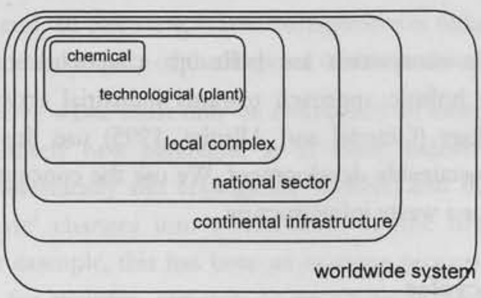


Figure 6. Different levels of the (chemical) industry

6. Conclusion

We conjecture that regarding technology development for waste management a great many opportunities are yet to be explored. At present, integral resource management, for example, is addressed almost exclusively by academics (e.g. Reuter, 1998), while in policy development waste management per se prevails.

Upcoming research will be focused on decision-support for a changing waste management sector. The leading paradigm is: *Waste is regarded as a temporarily emerging property of a substance or object: waste is a potential resource, what is waste for one industrial activity can be the feedstock of another.* A logical consequence of this is that waste disposal involves not only the classical waste disposal facilities, but all industrial processes.

References

AOO, Afval Overleg Orgaan, *Ten-year Program on Waste Management 1995-2005* (TJP.A-95), Utrecht, 1995

R.U. Ayres, L.W. Ayres, *Industrial Ecology*, Edward Elgar, Cheltenham, 1996

L.M. van Breda, G.P.J. Dijkema, "EIA's Contribution to Environmental Decision-Making on Large Chemical Plants", *EIA Review*, 18 (4), 391-410, 1998.

F. Borge, European Environmental Law, a Summary, University of Bergen, Norway, 1997 at http://www.uib.no/ped/humanities/law/Summary.txt)

A. Chauvel, G. Lefebvre, *Petrochemical Processes*, Part. II, Gulf Publ. Company, Houston, 1989

G.P.J. Dijkema and M.A. Reuter, Dealing with complexity in material cycle simulation and design, *Comp. Chem. Eng.*, 1999 (to appear).

DIOC Infrastructures, *Workshop Scenarios* 2030, May 19, Delft University of Technology, 1998

EEA, *Environment in the European Union 1995*, Report for the Review of the Fifth Environmental Action Program, 1995

EEC, *DG XI Environment Waste-Fact sheet, Environment, Nuclear Safety & Civil Protection, Waste: Facts and trends,* European Communities, 1998 at: http://www.europa.eu.int/geninfo/copyright_en.htm)

N. Gertler, *Developing sustainable industrial structures, Thesis Industrial Ecosystems,* Massachusetts Institute of Technology, Program on Technology, Business, & the Environment, at http://web.mit.edu/cptid/www/tbe/theses/gertler.htm)

T.E. Graedel, B.R. Allenby, *Industrial Ecology*, Prentice Hall, Englewood Cliffs, 1995

J. Larimer, *Infrastructures,* in: Int. Technology Res. Inst. Loyola College, *Display Technologies in Russia, Ukraine, and Belarus*, Maryland, U.S.A., ISBN 1-883712-36-X, 1994.

M.A. Reuter, The simulation of industrial ecosystems, *Min. Eng.* 11(10), 891-918, 1998.

H.J. Sas, J.P. van Soest, M.J. Zegwaard, G.P.J. Dijkema, *Disposal of plastic household waste: analysis of environmental impacts and costs,* Centrum voor Energiebesparing en Schone Technologie, 94.5251.022/e, Delft, 1994.

UNEP, *Global Environment Outlook-1*, United Nations Environment program, Global state of the environment report 1997 at: http://www.unep.org/unep/eia/geo1/ch/toc.htm

Abbreviations

AOO "Afval Overleg Orgaan"

The Dutch Waste Management Council was established in 1990 by the Dutch Ministry of Housing, Spatial Planning and the Environment (VROM), the Association of Provincial Authorities in Netherlands and the Association of Dutch Municipalities. The council's tasks are defined in the Agreement on Waste Disposal compiled by the parties involved. The AOO's primary task is to safeguard coherent collective waste management on a national scale.

AVI "Afval Verbrandings Installatie"

Dutch abbreviation for a Municipal Solid Waste Incineration facility

CEEC Central and Eastern European Countries

Economic association of countries in this region.

DIOC "Delfts Interfacultair Onderzoekscentrum"

In 1996 the Delfts Research Strategy committee advised the Board of the University to launch a number of strategic research initiatives on technology for the 21st Century. The DIOC Infrastructures is one of the programmes launched by January 1st 1998, which requires the input of multiple disciplines on a truly interdisciplinary research topic.

DSD "Duales System Deutschland"

The name of the law that effected a change of the German Waste Management System; it initiated separate collection of various waste fractions; collection, processing etc. had to be financed by fees paid for the use of "Der Grune Punkt" logo.

EAP Environmental Action Program

A series of policy programs developed by the EEC / EU.

EEC the European Economic Community

EU The European Union

EEA European Environment Agency

The EEA was launched by the European Union (EU) in 1993 with a mandate to orchestrate, crosscheck and put to strategic use information of relevance to the protection and improvement of Europe's environment. The Agency, based in Copenhagen, Denmark, has a mandate defined by Council Regulation (EEC) No. 1210/90 to ensure the supply of objective, reliable and comprehensive information at European level, enabling its member states to take the requisite measures to protect their environment, to assess the result of such measures and to insure that the public is properly informed about the state of the environment.

GFT "Groente, Fruit en Tuinafval"

Public name for the biomass-fraction of Dutch household waste, which is nowadays collected and processed separately in virtually all Dutch Municipalities.

MSW Municipal Solid Waste

This is the term generally used for waste collected at households.

MPW Mixed Plastic Waste

A term first used in Germany, where in the DSD-system all plastic waste are collected separately from households.

OECD Organisation for Economic Co-operation and Development

The OECD originally was set up as the Organisation for European Economic Co-operation (OEEC) in 1948 to administer Marshall Plan funding on the European side. In 1960, the Marshall Plan had completed its task and Member countries agreed to bring in the United States and Canada to form an organisation that would co-ordinate policy among the Western, industrialised countries. The new organisation was named the Organisation for Economic Co-operation and Development (OECD). Since 1960, a number of countries have joined the OECD, also outside Europe (including Japan, Finland, Australia, New Zealand, Mexico, the Czech Republic and Hungary). At OECD, representatives from Member countries meet to exchange information and harmonise policy with a view to maximising economic growth within Member countries and assisting non-member countries to develop more rapidly. (OECD at: http://www.OECD.org)

UNEP United Nations Environment Program

T.E. Graedel, B.R. Allenby, [illegible]

J. Lerner, Infrastructure for Int. Technology [illegible] in Russia, Ukraine and Belarus, Maryland, U.S.A., ISBN [illegible]

M.A. Reeves, The simulation of industrial ecosystems, [illegible]

H.J. Sas, J.P. van Soest, M.J. [illegible], C.P.J. [illegible], [illegible] Technologie, [illegible], Delft, 1994.

Abbreviations

AOO "Afval Overleg Orgaan"

The Dutch Waste Management Council was established in 1990 by the Dutch Ministry of Housing, Spatial Planning and the Environment (VROM), the Association of Provincial Authorities of Netherlands and the Association of Dutch Municipalities. The council's tasks are defined in the Agreement on [illegible] parties involved. The AOO [illegible] management on a national scale.

AVI "Afval Verbrandings Installatie"

Dutch abbreviation for a Municipal Solid Waste Incinerator [illegible]

CEEC Central and Eastern [illegible]

DIOC "Delftse Interfacultaire Onderzoekscentra"

In 1996 the Delft Research Strategy [illegible] advised the Board of the University to launch a number of strategic research initiatives on technology for the 21st Century. The DIOC Infrastructure is one of the [illegible] launched on January 1st 1998, which [illegible] as a truly interdisciplinary research [illegible].

[illegible]

The intake of [illegible] a change of the Dutch waste management system: a [illegible] collection of various waste materials [illegible], processing etc. had to be financed by fees paid for the use of [illegible]

EAP Environmental Action Program

Traffic in the 21st Century
A Scenario Analysis of the Traffic Market in 2030

Theo J.H. Schoemaker, Kaspar Koolstra, Piet H.L. Bovy

Delft University of Technology
Faculty of Civil Engineering and Geosciences
Transportation Planning and Traffic Engineering Section

Abstract

Scarcity of traffic infrastructures is increasingly seen as a fact that can not be resolved by simply building more roads, rails and airports. Scientific research can contribute to solving scarcity problems by systematically investigating possible strategies to deal with scarcity. The layers and markets of the transport system are defined in this paper and current technological developments are described. Traffic automation, the introduction of new signaling systems and new pricing mechanisms are expected to be important trends with respect to the possibilities for dealing with infrastructure scarcity. Since future developments are insecure, four scenarios are introduced to describe possible situations in 2030. The intensity of usage of new technologies and new pricing mechanisms varies among these scenarios. In the concluding section, we evaluate the merits and demerits of some of these technologies and pricing strategies.

1. Introduction

In many countries, increasing levels of car traffic, air traffic, and, to a less extent, rail traffic have induced scarcity problems on urban roads and freeways as well as on urban rail networks and at major airports. From the beginning, transportation science has aimed at optimizing the design of transport networks to meet transport demand, while also taking into account other aspects like safety, land use, investment costs and environmental costs, however, travel demand patterns are changing faster and are far more flexible than traffic infrastructure networks. For this reason, the management of traffic flows has gained the attention of transport planners and traffic engineers.

This paper reflects on future opportunities and threats with respect to the management of transport infrastructures that experience capacity scarcity problems. To this end, the multimodal transport system is considered to be composed of a number of interrelated functional layers between which markets of supply-demand interaction determine the outcomes in terms of volumes, qualities, travel times, prices, etc. The aim of this paper is to develop four scenarios to show the options for the future of transport systems. Developing coherent scenarios can help reveal the relation between technical and other developments.

2. The transport system: layers and markets

Within the transport system, three layers or system levels can be distinguished (see figure 1)[1, 2]:

- *transport patterns* of goods and persons
- *transport services*, implying traffic patterns of vehicles
- *traffic networks*, based on the physical transport infrastructure

2.1 Description of the layers

Transport patterns

Transport is necessary to combine activities at different locations and thus facilitates the spread of activities over time and space. The decision to make a trip depends on the extra utility of being at a destination, compared to that of staying at the point of origin, and the disutility (time, pain and money) of the transportation itself. Thus, transport behavior can be explained by assuming utility maximization behavior of individuals and organizations. Given the available time-budget and money-budget, individuals and organizations will seek to maximize the utility of their transport patterns.

Transport services

Transport services are necessary to convey goods or people and essential preconditions for transport services are a means of transport (vehicles and personnel) and access to

[1] The three-layer model was introduced at Delft University of Technology for education purposes some years ago. Schaafsma (1997) has published a four-layer model that distinguishes between passengers and goods services and the services of conveying trains. In the three-layer model both tasks are regarded as transport services, which includes also logistic activities.

[2] This model is also described (in Dutch) in Koolstra, 1998.

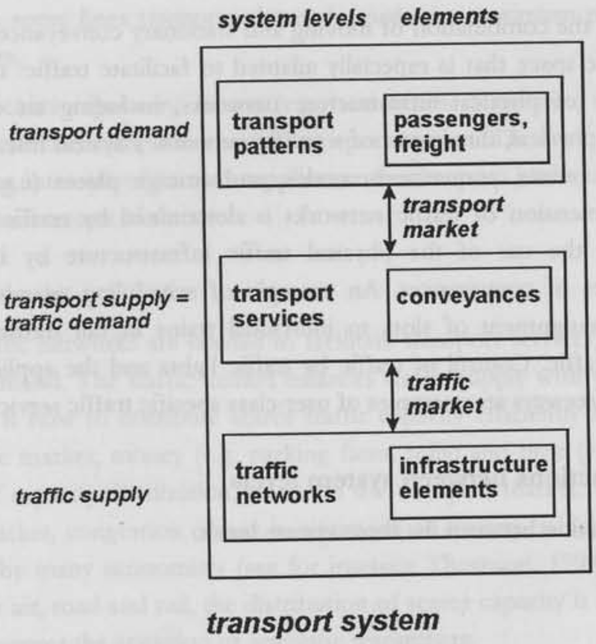


Figure 1. System layers and markets

traffic networks. In the case of self-organized transport, an individual organizes (but not necessarily performs) his own transport. If a third party organizes and performs the transport, the transport is called professional. The distinction between individual and collective transport is somewhat different. If people or goods with different origins or destinations are conveyed together ('bundled'), it is called collective transport. Individual transport implies the absence of bundling of different trips. Finally, there is a difference between private transport and public transport. The kinds of transport that every person or company can make use of (e.g. public trains and taxi's) are called public transport. Access to private transport is restricted to users who have private transport means at their disposal, or those who have permission to share a ride.

We refer to the transport services offered by professional transport companies as *service networks*. Service networks can be planned beforehand (scheduled transport) or be dependent partly or completely on actual demand (demand-responsive transport). Service networks have spatial as well as temporal dimensions. The spatial dimension includes the location of access points and the connections between these points. The time dimension relates to the timings of departures from and arrivals at the access points.

Traffic networks

Traffic can be defined as the combination of moving and stationary conveyances. Traffic occurs in a part of public space that is especially adapted to facilitate traffic: the traffic network. The availability of physical infrastructure networks, including air corridors, determines the static, or physical, dimension of a traffic network. Physical infrastructure networks include access points, connections, nodes, and storage places (e.g. parking places). The dynamic dimension of traffic networks is determined by traffic services. Traffic services regulate the use of the physical traffic infrastructure by individual conveyances or categories of conveyances. An example of scheduling, in advance, of individual paths, is the assignment of slots to individual trains in rail traffic, and to individual planes in air traffic. Control of traffic by traffic lights and the application of special purpose lanes on freeways are examples of user-class specific traffic services.

2.2 Markets: the interactions between system levels

Two types of markets function between the three system levels:

- transport market
- traffic market

The *market* concept is used to describe the interactions between the system levels. These markets act as regulators of supply and demand for transport and traffic. Requirements, for instance in terms of volumes or qualities, posed by the upper system levels determine the demand, and possibilities, in terms of capacities and costs, offered by the lower system levels determine the supply.

In general, however, these markets do not function as 'normal' or 'ideal' markets as described in economic theory. Supply and demand can be balanced in many ways other than using a market price, for instance by using a priority list (which user groups have priority) or by letting chance decide who will be supplied and who not. Of course, combinations are also feasible, for instance a reservation fee combined with a priority list and with first-in first-out queuing.

The transport market

The transport demands of persons and goods stem from the corresponding activity patterns. The supply of transport services by professionals, as well as self-supply of transport, enables persons and goods to be transported. The transport market balances supply and demand of transport. Basically, money and time (waiting) are used to distribute the available transport capacity in situations of scarcity. In many cases, it is possible to adjust vehicle capacities to meet transport demand. An example of an exception would be the crowded metropolitan railway systems of London and Paris, where on some lines transport demand equals the maximum passenger capacity during peak hours.

In some cases regulatory means (e.g., taxi licenses) are used to restrict the access of transport service companies to the transport market. The main reason behind this is to protect the interests of current suppliers and to keep up minimum standards of supply to protect the interests of the users.

The traffic market

Since traffic networks are needed to facilitate transport services, transport supply implies traffic demand. The traffic market balances traffic supply with traffic demand. The main question is how to distribute scarce traffic capacity efficiently among potential users. In the traffic market, money (e.g. parking fares, tolls) and time (e.g. queues) are important means of capacity distribution, just as in the transport market. Without regulation of the traffic market, congestion occurs in situations of scarcity, which is seen as a sub-optimal solution by many economists (see for instance Thomson, 1998). In all modal networks, especially air, road and rail, the distribution of scarce capacity is an important policy issue, which deserves the attention of scientific researchers.

3. Current developments

This section describes current developments, which are generally of technological nature. These developments have been initiated recently or even for some time, and probably will continue throughout the next years, however, neither the extent to which new technologies will be introduced by the year 2030, nor the possible outcomes of these trends, are discussed in this section; the extrapolation of trends is subject to the scenario analysis presented in section 5.

3.1 General developments

Technology push

During the development of the transport system, two key issues have demanded special attention, i.e. capacity and speed. Presumably, the increase in transport demand has led to greater capacities, while the need for shorter travel times has resulted in faster speeds. It could also be stated the other way around, i.e. greater capacities and faster speeds have facilitated more and faster transport. This *technology push* has been an important factor in transportation history. For instance, the introduction in London of metropolitan rail transport enabled suburbanization, because this new kind of transport was faster and could convey more people than existing modes. In the first decades of the nineteenth century, metropolitan railway companies could afford to extend their services beyond the edges of the London urban area, since it was expected that the newly served areas would become urbanized soon after the opening of the new line. In this situation, the new lines enabled people to combine suburban living with inner city jobs, something that was not previously feasible (Bayman & Connor, 1994).

New technologies of the late twentieth century include magnetic levitation (Maglev) combined with linear induction motors to attain even higher speeds than with conventional high speed trains, and automation of road traffic to enhance safety and increase the road capacity. The old technologies, however, will not be replaced unless the advantage of these new technologies is evident for both users and producers. Moreover, the functioning of the transport system does not depend only on the technology used, it also depends on the organization of the transport and traffic markets. This paper focuses mainly on possible technological and other developments that are related to the traffic market.

Automation and standardization

In general, there is a trend towards automation and standardization. Automatic coupling and uncoupling of railway cars, for instance, might enhance the competitiveness of rail cargo transport, while automation of public transport saves on driver costs. Standardization is also an important option, for reasons of cost reduction, for instance, standardization of load units enhances the efficiency of cargo handling.

The market oriented approach

Since higher speeds and larger traffic capacities have triggered increasing transport distances and volumes, planners have realized that it is not feasible to provide sufficient capacity to enable all trips. Especially in urban areas, traffic space is scarce and road congestion has become a structural problem. It should be noted, however, that in terms of opportunity costs, lack of traffic safety is still a far more important problem than road congestion, at least in the Netherlands (Ministerie van Verkeer etc., 1990). Nonetheless, road capacity has become scarce and is increasingly being regarded as such by traffic planners. The functioning of the traffic market is an important question for the future, and hence has been chosen as the main topic for the transport project of the Design and Management of Infrastructures research program of the Delft University of Technology. Improvement of the utilization of traffic capacity, given a range of different users with different needs and preferences, will be the main goal of this project. Since a comprehensive review of all (potential) developments in the transport sector would be either too complicated or unsatisfactory, the focus of this paper is on the developments that are especially relevant for the functioning of the traffic market.

3.2 Trends in traffic control

This section elaborates on the technological developments that are related to the functioning of the traffic market. Some developments are specific for one mode (rail, road, water or air), but others apply to several modes. New technologies in transportation are mainly related to the automation of traffic control. Among these are the automatic provision of information about routes and delays, the improvement of safety systems, for instance the introduction of moving blocks and the coordination of signaling installations, for instance traffic lights in urban areas. Other developments are related to the introduction of new pricing mechanisms in the traffic market. In some cases, this requires new technologies, for instance systems for automatic payment of tolls, but this is not always the case. In this case, the key innovation is the introduction of new economic mechanisms; the introduction of new technologies is only of secondary importance[3]. The next paragraphs focus on some developments that can have a substantial impact on the future of traffic.

Automated information

Especially on the road, where there are many independently operating users, providing information about incidents and delays may help to relieve congestion by optimizing the routing of vehicles. This will probably not be sufficient to change the routing to the system optimum, because it is not compulsory to follow the advised routing.

Automated traffic

In order to be able to offer frequent urban public transport at reasonable prices, some metro systems have introduced automated train operation, thus eliminating the necessity of at least one driver per car. Presently, this technology is only economical if demand is considerable and if high frequencies are considered necessary. It might also be feasible to automate inter-urban rail systems with grade crossings in the near future.

Automated traffic has also been introduced at some freight terminals at seaports to competitively handle the transshipment of containers. At Delft University of Technology, the possibilities of freight transport automation are being studied in the Freight Transport Automation and Multimodality (FTAM) interdisciplinary research program. Automation

[3] See for instance Schaafsma (1998) for a discussion of possible improvements of the functioning of rail traffic systems, including traffic networks, vehicle technologies and traffic market regulation.

of traffic control has also been introduced in air traffic for safety reasons. Automated flight control is now standard in many planes and it is used for the routine operations between the take-off and the landing.

For safety reasons, automation of traffic has been introduced first at systems without access of other vehicles or pedestrians (autonomous traffic). However, research is focussed increasingly on automation of vehicles in mixed traffic. Intelligent cruise control, designed to maintain a constant speed when possible and to maintain a safe distance from preceding cars, is an example of road traffic automation. Some researchers even envisage completely automated road traffic in the future, especially on freeways in urbanized areas.

Improvement of block systems

Block systems are used both in rail and air traffic for safety reasons. A block is a part of time-space that is uniquely reserved for one train or plane, thus ensuring a safe distance between vehicles. Static block systems, based on the brick-wall principle, are standard in current rail traffic. However, moving blocks are used for more advanced systems like the TGV in France. Though moving blocks basically maintain the same minimum safety distance, the actual minimum distance is smaller compared to the conventional block system, because moving blocks are not based on signaling beside the track but on in-vehicle signaling. A European standard system for mobile communications with trains (GSM-R) is being developed, which will enable further improvements of the system (Gibtner, 1998). For instance, if information about speed and acceleration of trains is transmitted to following trains, the 'brick wall' principle, i.e. the assumption that every train might come to a standstill in zero seconds, can be abolished, thus permitting even higher capacities.

Coordinated signaling

Operational traffic management can be improved in a number of ways. Advanced real-time management techniques are being developed to manage traffic flows more efficiently. Generally, these improvements imply the usage of more information for the optimization of the whole network instead of optimizing local situations. For instance, urban traffic lights can operate in a coordinated mode and respond to the traffic patterns measured in the whole city, instead of responding only to single vehicles arriving at the intersection (Van der Burgt & De Jong, 1998).

Inland navigation is little regulated by traffic lights, but it is especially dependent on the opening times of bridges and locks. Coordination of opening times and adjusting opening times to the general traffic pattern of the moment might reduce delays and thus enhance the position of inland navigation as an alternative to road and rail transport.

Demand-responsive reservation systems

Another option is the shift from advance reservations combined with operational traffic control to demand-responsive reservation systems. Instead of establishing timetables many months in advance, service companies are offered the possibility to reserve a slot well in advance or only a few minutes beforehand. In cases of delays or operational problems, reservations can be changed, provided of course that there is still room on the network. In this way, tactical planning of network usage and operational management would become integrated.

3.3 Regulation of the traffic market

Currently, policy makers are increasingly aware of the possibilities offered by the introduction of new pricing mechanisms to relief congestion problems. These strategies are founded in economic theory, but still suffer from some practical as well as theoretical problems. Some technical problems have already been overcome by the introduction of new technologies.

Road pricing

De Wit & Van Gent (1996) distinguish between two different pricing strategies, both relevant for future developments in the traffic market. *Cost pricing* aims at a better coverage of investment costs and maintenance costs of infrastructure, by demanding that the users pay. *Scarcity pricing*, however, primarily aims at a better distribution of the scarce available traffic capacity, which might have the consequence that the revenues are insufficient to cover all costs. Both principles are competitive and demand comparable types of technology. Cost pricing and scarcity pricing are both applicable in combination with automation of traffic control related measures.

In road traffic, two types of scarcity pricing have been proposed to reduce congestion costs. Congestion pricing aims at elimination of congestion by assigning dynamic tolls to routes. It is assumed that congestion will disappear if the right levels of tolls are chosen. However, offering a choice between paying and waiting might be preferred for several reasons. In this case, it would be an option to introduce paylanes, leaving the user the choice to pay nothing and wait in the other lane.

Free access to traffic markets

Since the European Union requires that railway companies from all the EU countries have access to the national railway networks, a strategy has to be developed to fairly assign railway capacity to competing users. One option is to use simple principles like 'international trains have priority over local trains', or 'first come, first served'. Another option is to choose the railway company that is willing to pay the most, i.e. a capacity auction. The same holds for air traffic. Instead of simple rules like 'home carriers have priority over other companies', some system of scarcity pricing can be used (see Duchemin, 1994).

4. External factors

The functioning of the traffic market can be judged by the efficiency of the transport system, e.g. in terms of time losses, but also in terms of the external merits and demerits. The transport system is also influenced by its environment. The influence of the environment on the traffic market is the subject of this section.

One of the external environment factors is the *technology push* factor. If new technologies are developed at a high pace, eagerly introduced and easily accepted by the public, the transport system will change considerably. It is thinkable, however, that only improvements of existing technologies will last and that completely new technologies will only have a marginal impact.

Another important external factor is the level of *environmental consciousness*. If both the public and policy makers are highly aware of the desirability of preserving the environment, transport policy will focus on an efficient usage of space and energy, and on a reduction of traffic emissions. In this case, road and air traffic capacity would become increasingly scarce. If, however, policy makers would focus on increasing accessibility and transportability, road and air traffic could be allowed to grow.

Related to this problem is the amount of *policy interest* in the traffic market, which is associated with the level of discomfort with problems such as traffic congestion and environmental damage caused by traffic that is experienced by the public, and moreover, by policy makers. Furthermore, both national and European 'liberalization' policies are the main trigger for the introduction of pricing mechanisms for infrastructure usage.

Safety consciousness is also related to environmental consciousness is. Since the economic impact of traffic safety still outweighs the economic impact of traffic congestion, and since traffic safety has also extensive social consequences, it is possible that this will become the main policy target in future decades. In this case, enhancing traffic safety would be the main criteria, while capacity enlargement and improvement of capacity scarcity management would only be secondary goals of technical innovations.

More generally, the functioning of the traffic market is influenced by economic and spatial developments. The level of economic welfare, for instance, influences transport demand and the availability of funds to build new infrastructures or to improve traffic control systems. The possibilities to build new infrastructures are also influenced by spatial developments, and the same holds for transport demand.

5. Scenarios

Four scenarios are introduced in this section that describe possible arrangements of the traffic market by the year 2030. Firstly, in the *technocracy* scenario the focus is on the introduction of automated control with centralized traffic management. Secondly, in the *collectivity* scenario, the absence of privatization is combined with the improvement of public person transport systems and the introduction of a new multimodal public transport system for goods. In the *marketing* scenario, the introduction of market mechanisms in the traffic market is the central theme. Privatization of infrastructures will have been the trigger for the pricing mechanisms in 2030. Finally, in the *first come, first served* scenario, pricing mechanisms are absent. In this scenario, the only structural policy interventions used to relieve congestion problems are enlargements of the physical infrastructure network.

The four scenarios can be related to the multi-layered system model (Figure 1) as follows. The *technocracy* scenario is focused on direct control of the traffic market by creating restricted individual traffic networks. In the *collectivity* scenario the main goal is to improve public transport services, mainly by investing in public transport infrastructures. The *marketing* scenario is, in this respect, the opposite of the collectivity scenario, because its main focus is on a better arrangement of the traffic market, without direct interventions in the transport service and traffic network layers. In the *first come, first served* scenario, the interventions in the service layer and the traffic market are kept to a minimum. The focus in this scenario is on building more infrastructures.

The level of usage of the technological options described earlier differs among the scenarios. For instance, automation of transport is a necessity for the centralized optimization scenario, but is irrelevant for the *first come, first served* scenario. The four scenarios are described in detail below.

5.1 Technocracy scenario

The main characteristic of the technocracy scenario is the massive use of new technologies to enable centralized traffic control. The main rationale behind this kind of public intervention is to reduce congestion, to prioritize 'economically important' and 'environmentally friendly' traffic, and to enhance traffic safety. In this section we sketch a traffic system for 2030 that will result from the technocracy scenario.

Physical infrastructures

Due to concern for the environment the possibilities to extent the traffic networks are limited. Thus, space efficient and energy efficient systems are preferred from a system point of view. However, there is a mismatch between the choices of an individual in a free-choice situation, and the system optimum. Since policy makers acknowledged the importance of public organizations playing an active role in the traffic market in time, massive developments have taken place between 2000 and 2030.

To prevent bureaucracy, independent organizations, not directly related to the government, manage the usage of infrastructures. The main task of these organizations is to assign individual paths ('slots') to network users. In this context, a user is not an end-user that consumes the transport service, but an individual or a company that offers a transport service. This corresponds with the 1998 situation with respect to the rail network and the airports. The previous high levels of congestion in road traffic have been reduced by the introduction of road infrastructure managers that have the same task, i.e. assigning slots to individual road users.

Slot reservation

The slot reservation procedure in road traffic works as follows. A path on the network can be reserved for a vehicle in advance, so activity patterns, production processes, and distribution patterns can be adjusted, without expensive delays during transport. The reservation system enables the policymakers to reserve sufficient capacity for privileged user groups, for instance public transport and international trade. A reservation fee is used as a mechanism to select 'economically important' traffic. Thus, the economic background of the system is scarcity pricing, combined with priorities to certain user groups. Nonetheless, the road system remains open to private cars without a reservation, provided that the admission of extra vehicles will not lead to congestion on the road. The reservation fee is used in this system as a form of marginal cost pricing. That is, it is not aimed at covering the expenses, involved in running the system, but at an efficient distribution of scarcely available traffic capacity.

Traffic control

Between 2000 and 2030, many new technologies have been implemented to be able to optimize the routing and timing of traffic, to discourage the use of environmentally unfriendly modes, and to enhance traffic safety. The following steps have taken place or are still in progress:

- introduction of signaling at nodes;
- complete signaling of the system;

- partial automation of traffic;
- automated traffic.

Though the usage of signaling at nodes was already widely used in the various traffic systems by the year 2000, only the rail system was completely signalized. The networks had to be signalized completely to allow control of all traffic markets. A further step has been taken in the form of a complete automatization of all rail systems and a partial automatization of road and air traffic.

The main rationale behind road traffic automation is the enhancement of traffic safety. A secondary benefit is a small increase in road capacity. A combination of automation of rail traffic and abandoning the 20th century fixed block system has led to lower operation costs and more than doubled rail capacities.

Public transport

Public transport companies are generally privately owned, but are dependent on public agencies for permits. The main task of these agencies is to prevent splintering of the supply of transport services. Due to the coordinating role of the national public transport agency, a single, multimodal tariff system has been introduced, using chip card technology. This means that the customer can pay with the same card, regardless of the company that offers the service and the kind of mode that is used. Public transport companies are allowed to compete with each other by offering lower or higher tariffs.

Public road transport has profited from the introduction of slot reservation systems on the road. Since the increase of traffic capacity supply has not kept pace with the increase in transport demand, the use of more efficient ways of transport has become a necessity. The usage of public transport, compared with private cars, has increased. Public transport, however, has not been able to replace the dominance of private cars completely. The main reason behind this is the rise in tariffs in the last decades, mainly induced by the abandonment of subsidies, and the fact that individual transport has also profited from the investments in the traffic system.

Information systems

Centralized optimization depends on new technologies, and on better information systems. Traffic monitoring systems can register deviations from the scheduled paths of vehicles, and respond by adjusting the routes of other vehicles. These information systems have enhanced the effectiveness of the control centers. If necessary, control centers can adjust the schedule to minimize the delays caused by incidents or can plan 'last minute' reservations of slots. The private transport companies also have control centers, which control the routing of the vehicles and adjust private schedules to delays and 'last minute' demands. The flexibility of the system is maintained via direct communication with the infrastructure controllers, despite of the generally rigid character of a slot reservation system.

Evaluation

The combination of traffic automation, improvement of signaling systems and the introduction of reservation systems, enables centralized optimization of the traffic system as a whole. Embedded in a democratic society, the centralized optimization scenario will have led to a well functioning transport system in a well functioning society. As long as the government strategy is focused on promoting public transport and efficient usage of vehicles, this system will result in less environmental damage. The absence of massive network extensions is favorable for the environment. However, the costs of these new systems will be high, and the definition of the optimum is largely political. An economic crisis after the year 2030 might reveal the weaknesses of the system.

5.2 Collectivity scenario

The application of new technologies is not the only way to induce a more efficient usage of traffic networks. If the public will accept the necessity for collectivization of transport, there are less technology-intensive ways to improve the traffic system. In this scenario, the main task of the government is to provide the necessary infrastructure to enable competitive public transport of goods and people. The awareness of the whole society of problems such as land scarcity and environmental pollution will have been the main rationale behind public policy. The situation in 2030, following the collectivity scenario is described below.

Public transport infrastructures

Public ownership is regarded the best way to defend the interests of the people. Thus, the infrastructure systems are managed by semi-public organizations, which are under the supervision of national or local governments. Due to environmental concerns, only limited extensions of the traffic network have taken place in the last decades. The government is mainly interested in investments in public transport. The public agencies have been able to provide the necessary infrastructures for an efficient public transport system by reserving exclusive lanes for buses on highways as well as in urban areas, and by building multimodal transfer stations, both for persons and goods. The acceptation of the necessity to use public transport, both of goods and of persons has gradually grown, and the present situation is that public transport is regarded as the backbone of the transport system, especially for long distance transport and commuting.

Public transport

The introduction of the just-in-time logistic concept at the end of the 20th century resulted in a distribution system that was characterized by a high traffic to transport ratio. Due to a lack of bundling and the high frequency of deliveries, mainly small trucks were used, often at less than 60% capacity. Government initiatives to promote public transport of goods have resulted in an efficient public transport system, that uses all kind of modes (rail, air, road and water) and offers highly frequent, flexible transport.

The public transport system for persons has developed to be the backbone of the transport system in urbanized areas. In rural areas, however, the system still functions mainly as a transport system for 'captives' (non-car owners), especially the young and the elderly.

Private companies operate the public transport systems; they are not allowed to compete for customers, but for permits only. The Ministry of Transport coordinates the public transport systems and subsidizes these systems for reasons of 'welfare economics' and to ensure that the whole population has access to transport.

Evaluation

The collectivity scenario shows that though using 'conventional' technologies, it is still possible to change the transport system. The boost for public transport in this scenario might prove to be an efficient way to reduce the external costs of the transport system. A prerequisite is acceptance by companies and individuals of the public transport systems. Lack of flexibility in the system and high costs, as well as diminishing acceptance by the public, might prove to be the main weaknesses of this scenario.

5.3 Marketing scenario

According to many economists, organizing well-working markets is the best way to deal with scarcities like traffic capacity. In a situation with many suppliers and demanders, a market price will become established which will balance supply and demand in what is supposed to be the most efficient way. In the marketing scenario, the government is not directly involved with the transport system and other areas of the economy, but it tries to arrange independently working markets. Both the transport companies are private and the infrastructure are privatized.

Infrastructure ownership

In this scenario, the ownership of transport infrastructures is completely privatized. Thus, the central government has no direct influence on the functioning of the traffic market.

Private companies build new roads and rails, maintain existing infrastructures and determine the user tariffs.

The key role of the government is to prevent the abuse of monopoly positions and to limit the environmental damage that is caused by traffic. Traffic safety and environment are not seen as issues of utmost importance in this scenario. The government uses a permit system to mitigate or compensate for the environmental damages. In this system, the amount to pay for a permit is equal to the amount of money needed to compensate the external costs of the transport system.

Road pricing

Both the government and the public regard traffic capacity as a scarcity that can be bought for money. Since private companies determine the toll level, the toll levels are at least as high as the scarcity price. However, since infrastructure maintenance and network extensions also have to be financed by toll incomes, tolls usually are higher than the scarcity price. This problem has partially been overcome by offering the possibility for frequent users to buy permits, however, the profitability of investments in the transport network is generally low, which has resulted in a pace of network extensions that is insufficient to keep up with the rise in transport demand. The rail network has even declined.

New techniques to control traffic have been introduced with the sole purpose of enabling efficient payment of tolls and detection of intruders in the system. Especially in road traffic, this has resulted in the introduction of some new technologies, but the technical situation in air traffic and rail traffic has remained fairly stable over the first decades of the 21st century.

Public transport

Many public transport companies are facing hard times. They have to compete for travelers on a free market, which is becoming increasingly hard because of a lack of investments in public transport systems during the last decades, except for some urban transport systems. Some companies have profited from their relative monopoly position in the first decades of the 21st century, and are now beginning to invest in the traffic system. Since the car system appears to have found its limits in metropolitan areas, the market share of the public transport system is again beginning to grow in these areas. However, public transport has completely vanished from the rural areas, except for some (generally expensive) taxi services.

Evaluation

A disadvantage of this scenario is that users with a low budget will have problems arranging for their transport. The kinds of transport with high fixed costs, like rail transport, will experience problems. Moreover, infrastructure owners might make large profits, due to a favorable monopoly position. Secondary effects will be that the activity space of individuals becomes more compact, and telecommunication will be used instead of transport.

This scenario might be advantageous to the environment, especially when the permit system works well. This depends, however, on the feasibility to internalize external costs of traffic.

The main advantage of this scenario is the introduction of markets with many players with private interests, combined with a system of compensation of externalities. This might enhance the flexibility of the transport system, though it will not make it less liable to an economic crisis, because of the lack of government investments.

5.4 First come, first served scenario

In the three previous scenarios, new technologies and new economic mechanisms were introduced to cope with scarcity. If, however, the development of new technologies stagnates, and the introduction of pricing mechanisms appears unfeasible because of objections of the public, it will become necessary to resort to a more simple approach. A sketch of the situation in 2030 that would result from a lack of innovations in the traffic market is given below.

Infrastructures and traffic control

The present traffic control systems are only slightly improved versions of the systems that were used in the nineties of the twentieth century. Like thirty years ago, simple criteria, for instance distance of travel, determine which train or plane gets priority in the timetable. The operational management mainly works with the rule 'first come, first served'. Congestion has increased, especially on urban roads, freeways and some waterways, which still do not have a system of scheduling in advance. New links and enlargement of the capacity of existing links partially compensate the lack of sophisticated methods to cope with capacity scarcity. Between 2000 and 2030, the area occupied by traffic infrastructures has grown by more than one third. To limit the negative external effects, many new urban roads and railroads have been built underground. Thus, compared with other scenarios, the savings in signaling systems and other technologies are more than compensated by the extra expenses on the infrastructure networks.

Evaluation

Though this scenario has the advantage of simplicity and does not depend on insecure technological developments, it seems to be the least positive scenario of all four. For instance, the space consumption of traffic is the greatest in this scenario. However, if traffic demand stabilizes in the near future, this might just be the most economical solution, provided that only limited extra investments in the traffic infrastructure network are needed.

6. Evaluation

The four scenarios presented in the previous section all have in common that they try to solve traffic problems, although they use different strategies and are based on different sets of priorities. The question now comes to mind which developments are desirable and which are not. For instance, is the increasing role of technology in the first scenario really desirable, and what are the dangers of privatization? In this section we will give our view on some of these questions.

Is technology the solution?

Some technological developments seem very desirable. For instance, rail traffic capacity can be enlarged by improvements of the signaling system, while maintaining the same level of security. According to some transport engineers, the same holds for road traffic automation. However, since this would imply the introduction of totally new technologies, the costs involved would be massive. These new technologies will probably need a long introductory phase before traffic safety and road capacity can be enhanced simultaneously.

Another question is the cost-benefit ratio of these new technologies. In some cases, the benefits of traffic automation might be marginal compared to more simple technologies, and may not outweigh the costs.

There are also examples of solutions with relatively low costs and high potential benefits. Better coordination of available paths on the level of scheduling, and better coordination of signaling on the operational level, are positive measures that need relatively low effort. For instance, providing 'freeways' for European cargo trains is necessary to compete with other modes. In the present situation, too much time is still lost at the internal European borders (UIC, 1998). The same holds for the coordination of urban traffic lights, which might reduce delays in urban road traffic.

Is pricing the solution?

From the point of view of an economist, the introduction of marginal cost pricing mechanisms at the traffic market seems the ideal thing to do. The user willing to pay the most would get priority, so the road would be free for those needing it most of all. Since the total budget differs among different user groups, however, it might turn out as well that the rich would get priority, regardless the importance of the trip. So, the introduction of scarcity pricing needs to be done with caution. Sometimes it might be better to grant priority based on objectively recognizable user groups.

Is privatization the solution?

The main advantage of privatization of infrastructures is that due to market processes, the infrastructure management would be forced to operate efficiently, provided that the users have an alternative. However, privatized infrastructure companies could possibly enjoy a monopoly position, thus enabling them to make profits at the expense of the rest of the economy. Thus, privatization is only feasible if the risk of misuse of monopoly positions is dealt with.

Conclusion

None of the scenarios presented is a doomsday scenario, but none is without threats. It is possible, however, to identify the main risks and disadvantages of various strategies. Theoretical, empirical and experimental research can be used to test hypothesis and identify chances and threats for the future. The Delft University of Technology interdisciplinary research program Design and Management of Infrastructures can contribute to indicate possible solutions for the future.

References

Bayman, B. & P. Connor (1994), *Underground: Official Handbook*. Harrow Weald: Capital Transport, 3rd edition.

De Wit, J. & H. van Gent (1996), *Economie en transport*. Utrecht: Lemma.

Duchemin, J. (1994), Réflexions sur l'attribution des créneaux horaires en aviation : enseignements pour le secteur ferroviaire. *Rail International* **25**/11 (Nov.), pp. 17-24.

Gibtner, H. (1998), Ausbaustrecke Berlin-Halle/Leipzig: Teststrecke für die europäische Funkzugbeeinflussung. *ETR* **47**/1, pp. 15 - 18.

Koolstra, K. (1998), Slottoedeling: het ontwerp van klasse-specifieke verkeersnetwerken. *Colloquium Vervoerplanologisch Speurwerk 1998: sturen met structuren*, deel 4, pp. 1543-1562. Delft: CVS.

Ministerie van Verkeer en Waterstaat & Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer (1990), *Tweede Structuurschema Verkeer en Vervoer; Deel D: regeringsbeslissing*. 's-Gravenhage: SDU.

Schaafsma, A.A.M. (1997), Liever het ontleedmes dan de botte bijl: radicale scheiding van infrastructuur en exploitatie niet doeltreffend voor concurrentie en innovatie in de spoorwegbranche. *Tijdschrift Vervoerswetenschap* **33**/3, pp. 237-258.

Schaafsma, A.A.M. (1998), Minder plannen, meer regelen: naar een betere performance van de spoorwegen met dynamisch railverkeersmanagement. *Tijdschrift Vervoerswetenschap* **34**/2, pp. 115-135.

Thomson (1998), Reflections on the Economics of Traffic Congestion. *Journal of Transport Economics and Policy* **32**/1, pp. 930112.

Union Internationale des Chemins de fer (1998), *Trans European Rail Freight Freeways: The view of the UIC Freight Commission*. Paris: UIC.

Van der Burgt, G. & P. de Jong (1998), Verkeerslichtenregelingen: van voertuigafhankelijk naar verkeersafhankelijk. *Verkeerskunde* **49**/7, pp. 14-19.

Wong, J.T., Basic concepts for a system for advance booking for highway use. *Transport Policy* **4**/2, pp. 109-114.

Water Infrasystem Design and Driving Forces for Change

Martine van der Ploeg, Machtelt H. Meijer

Delft University of Technology

Faculty of Technology, Policy and Management

Policy Analysis group

Abstract

The current organization of the water services industry is challenged by autonomous changes: increasing urbanization, globalization, and developments in the European Union leading towards liberalization of public services. Customer preference and government regulations change as society changes. This has consequences for the demand for products and services that are supported by water infrastructures: drinking water, water for industry and household use, water supply, wastewater collection and wastewater treatment. Water infrastructures, and the organizations that own, operate and maintain them, must be adjusted to meet the new demand characteristics of products and services.

Water works, or water infrastructures, and the organizations that own, maintain and operate them, can be considered as one system with specific boundaries. In this context, we ask ourselves what water infrasystem designs are best suited for a changing world? Which changes in socio-economic development, political climate or water availability may impact the functioning of water infrasystems? What would these impacts be? What designs are robust with respect to such changes? If we are able to predict the most important changes accurately, each possible design could be evaluated for its impacts under the predicted future circumstances and preferred designs could be identified. However, such prediction of the future is not possible, and it is unlikely that the uncertainties that confront infrasystem planners will be resolved over the course of time with new information. Therefore, infrasystems are at best adaptive, designed not to be optimal for a best-estimated future but robust across a range of possible futures. Methods for future capacity estimation and adaptive capacity management are important in the search for robust designs. With this paper, we take a first step in the search for robust water infrasystem designs.

1. Introduction

Water problems to be faced in the 21st century concern first and foremost the increasing demand for water for human consumption and use, and for irrigation and industry. Urban growth and industrialization put enormous pressures on the management of water supplies and the associated water infrastructures. Cities, citizens and their economic activities, compete for water supplies and draw on water resources located further and further away. The resulting depletion of water supplies, damage to ecosystems, and actual water scarcity put increasing limits on socio-economic development. Similarly, urban and industrial wastes are disposed into water systems that cannot absorb the environmental burden. This capacity to displace environmental burdens increases with wealth and the development of industry, centralized water supply, and sanitary systems. The immediate effect of the resulting water pollution is seen in the decrease of water available for drinking water preparation, and in the deterioration of public health and environmental quality within the city limits and in the surrounding areas. Water services organizations, responsible for water supply and public health, must find adequate answers to these threats at a time when water demand outruns water supplies (Kjellén and McGranahan, 1997).

The increasing scarcity of water and economic resources has led to a change of paradigm in water management from "supply management", which focused on meeting customer demand, to "demand management." Water policy now focuses on reducing water demand, e.g. by eliminating water spillage through leaks in infrastructure, rather than taking measures to increase water supplies to meet water demand. In water demand management, measures are geared towards improving the effectiveness, sustainability, and efficiency of water use. Such measures may be searched for and found in technology, the design and operation of water works and water appliances, in management and organization of the water services industry. Other important measures are those that lead to changes in customer behavior, e.g. the pricing of water services or information on ways to contribute voluntarily to a reduction in water consumption (UNESCO, 1998).

In urban developments, water works for water supply and wastewater treatment now are generally owned and operated by a central (public or private) administration. Industrial organizations may or may not make use of these centralized water services based on economic considerations, regulations, and the local conditions of the water system. New urban developments are also taking initiatives towards lessening the dependence on centralized water services. The collection of rainwater for use in office buildings or households and the application of small biological water treatment plants in neighborhoods, are but two examples of such initiatives.

These and other actions by industry, cities, and citizens challenge the task setting (water resources management, water supply, and public health) and organization of the water services industry. In addition, the current organization of the water services industry is challenged by autonomous changes like increasing urbanization, globalization, and developments in the European Union leading towards the liberalization of public services. How should water services organizations and their regulators respond to these changes at a time that, increasingly, water is considered a common property and a water crisis is not unthinkable? What technological, social, legal and economic considerations must be taken into account in the search for, and evaluation of, solutions?

Water works, or water infrastructures, and the organizations that own, maintain and operate them, can be considered as one system with specific boundaries. Hence, the term water infrasystem will be used in this paper (Weijnen and Bosgra, 1999).

Water infrasystems are complex systems. Our research is concerned with the behavior of such infrasystems and their responses to changes in the system surroundings. In this context, we ask ourselves what water infrasystem designs are best suited in a changing world. Which changes in socio-economic development, political climate or water availability might impact the functioning of water infrasystems? What would these impacts be? What designs are robust with respect to such changes? If we are able to predict the most important changes accurately, each possible design could be evaluated for its impacts under the predicted future circumstances and preferred designs could be identified. However, such prediction of the future is not possible, and it is unlikely that the uncertainties that confront infrasystem planners will be resolved over the course of time with new information. Therefore, infrasystems are at best adaptive, designed not to be optimal for a best-estimated future but robust across a range of possible futures (Walker et al., 1998).

We begin the search for robust water infrasystem designs with this paper. Since this is a first step in a large research project, we limit ourselves to water infrasystems of the Netherlands. Some of the characteristics influencing the design and operation of water infrasystems in this country are: high population density, high level of welfare and wealth, a long history of public water utility companies, and the uncertainties related to the integration within the European Union and environmental degradation. The scope of this paper is also limited to the small water cycle in urban areas. Waterworks for river management, like dams, dykes, or sluices, fall outside this scope. Instead, we focus on the delivery of water to households and industry, the collection of storm water and wastewater, and the treatment of wastewater. This allows comparison with other infrasystems, e.g. for energy supply, telecommunication, solid waste collection, and transportation, as described in this book.

The purpose of this paper is twofold. First, we want to give a general description of water infrastructure and, second, present a scope for research in the DIOC Infrastructure project. Therefore, the situation of Dutch water management is discussed briefly, followed by a conceptual model[1] for describing this water infrasystem. We specify the design variables for the infrastructure and institutional organization that we judge important for describing infrasystem behavior. This list of variables and the conceptualization represent our point of departure for research on water infrasystem design. These design variables, or variables of infrasystem behavior, change in response to changes in the system surroundings. We present a list of forces that drive these changes and influence the way a water infrasystem works. We then present three possible futures, or scenarios, based on this analysis of structural changes. In our conclusion we propose a focus for water infrasystem research and its potential contribution to research on infrastructure design and behavior.

2. Current situation of Dutch water infrasystems

The central organization of drinking water supply and sanitation has been developed in the Netherlands over a period of more than hundred years, starting in the late 1800's. A cholera epidemic in 1866 led to a public drinking water supply and sewer systems being installed. The current design of the Dutch water infrasystem reflects the political climate and economic developments of the past 50 years. Water is considered to be a public good and water services organizations in general belong to the public domain. The equity in access to water services is very high: 95 to 99% of the Dutch households are connected to water distribution and wastewater collection networks and pay for water services. Customers in rural areas receive the same quality and type of services and products as those in densely populated urban areas. The infrastructure hardware is well maintained but rates of replacement have been rather low in the past 30-50 years. The water services industry now faces major reconstruction of sewer and water distribution networks in the next 10-20 years.

The general public considers the quality of drinking water to be high and water services reliable. Prices for water services vary per geographic area. Explanatory variables for the differences in costs to consumers are the initial quality of water resources, population density of the service area, efficiency of operation, and past investments. The drinking water supply for about 5 million people relies on river water extracted from the Rhine or

[1] The vocabulary and modeling concepts we use in these conceptual models may differ from those used in other models for other infrastructures in this book. We have used jargon and modeling concepts of policy analysis and civil engineering. In the future, these will be adapted to vocabulary and concepts fit to describe infrastructures or infrasystems in general.

Maas, two major transboundary rivers. Water services to households are provided by public water utility companies. Industries buy water services from these organizations or build and exploit their own water supply and wastewater treatment plants. Changes in the claim on centralized water services by industry can have a large impact on the efficiency of infrastructure capacity management.

In the Netherlands, three different types of public institutions are involved in producing water services to households and industry. The division of tasks and responsibilities among these institutions is as follows:

- City councils own and manage sewer systems for the proper discharge of storm and wastewater. Construction and management of the sewer systems may be contracted out to private companies. (Outside the urban area, water boards are responsible for sewage and storm water collection and transportation.)
- Drinking water companies own and manage drinking water distribution networks. These companies hold the monopoly of delivering water to households. They also deliver water of different quality grades to industry. Drinking water companies are owned by public stakeholders: cities and/or governmental agencies that regulate the groundwater supplies. There are few exceptions to this delegated public ownership: two small companies are privately owned.
- Regional water boards own wastewater transportation networks and treatment plants. Some companies lease plants from American companies (cross-border lease) to reduce the costs of investment capital. Water boards are involved in wastewater treatment because of their public task to protect and manage the water quality of regional surface-waters.

The distribution of tasks over three public institutions may be understood from the historical perspective (Huisman et al., 1998). Drinking water companies are delegated public organizations, as they are owned by cities and provinces. Groundwater supplies are considered to be a public good and managed by the provincial councils. The cities and provinces are public administrations, as are the water boards, and elections are held every four years. In the past, these three organizations operated rather independently from each other. The need for collaboration is becoming more apparent, however, as the national water policy asks for sustainable water management. Closure of the small water cycle and sustainability of groundwater extraction are just two of the issues which require collaboration.

2.1 (Drinking) water preparation and distribution

The first reason for organizing central drinking water services is to secure public health. Water is one of life's necessities, a fact that is easily forgotten in a wealthy, water-rich country like the Netherlands. The Dutch government imposes strict rules, regarding the microbiological safety, pressure, taste, and color of drinking water. The Ministry of Housing, Spatial Planning and the Environment overseas the law on drinking water supply. This law sets norms for water quality and for ownership and management of water supply companies.

Fresh water resources in the Netherlands are relatively large because of the inflow of transboundary, snow and rain-fed rivers. Water scarcity should not be an issue in the Netherlands with respect to drinking water supplies, but the deterioration of water quality in rivers and lakes has prompted water companies to use groundwater resources. About two thirds of Dutch inhabitants drink water prepared from groundwater supplies located in the northern, eastern and southern provinces.

Table 1. *Client base, water resources and amount of drinking water produced in 1996 for the provinces of the Netherlands (adapted from: VEWIN Statistics, 1997)*

Province	Population in service area (10^6)	Yearly water production (10^3 m^3)	Ground water %	Surface-water %	Surface-water infiltrated artificially %	Reservoir-water %
Groningen	569	27	84	16		
Overijssel	1.124	76	91		9	
Gelderland	1.891	145	100			
Noord-Holland	2.385	181	19	28	53	
Zuid-Holland	3.301	275	18	53	29	
Noord-Brabant	2.256	213	99	1		
Limburg	1.134	80	100			
Drenthe	n.a.	58	100			
Zeeland	n.a.	14	7		16	77
Friesland	n.a.	45	100			
Utrecht	n.a.	79	100			
Flevoland	n.a.	15	100			
Netherlands	15.494	1.208	67	17	15	1

In the more populated western part of the country drinking water is prepared from river water (Table 1). In that part of the country groundwater contains large amounts of chlorides because of its proximity to the North Sea. The costs of transporting high volumes of water keep water companies dependent on nearby water sources. The maximum distance of purification plant to water source is 50 kilometer. The potential for the pollution of river water is a major concern in drinking water preparation, especially with regard to the potential for chronic toxicity from yet unknown substances. The precautionary principle, that is to prevent all recognized potential damage to public health, plays an important role in the infrastructure design for drinking water supply and distribution. Chances of risking public health must be kept as low as possible. River water is stored in reservoirs for settling or filtrated in coastal dune areas as a pre-treatment step. The supply of river water to the reservoirs and dunes can be closed in case of emergency; the reserves of these water supplies are designed to last for several weeks to months.

Industries use water in several different ways and the quality demands differ depending on the purpose of use: cleaning, cooling, or food and beverage preparation. For the latter purpose, water quality demands are much more stringent, sometimes even higher than for drinking water, as is the case for brewing. Drinking water supply companies can deliver these different products to industry, but some industries prepare water from groundwater resources that they exploit.

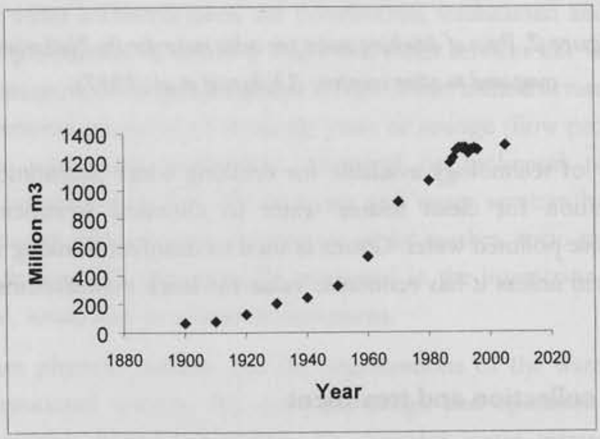


Figure 1. *Production public water supply in the Netherlands (Adapted from: VEWIN, 1997)*

After a sharp increase in 1960-1980, the demand for water in the Netherlands has stabilized in the past decade in spite of population, and economical growth (Figure 1). Household water consumption is decreasing because of the introduction of water saving appliances (mainly toilets and washing machines) and use of showers for personal hygiene rather than baths. Water prices for households range from one to two Euro per cubic meter, which is high compared to many industrialized countries with a similar quality service (Figure 2). The Dutch water price is determined on the principle of cost recovery and surcharged with a sales tax. This tax is 6 % as drinking water is considered a basic need. An increase to 17,5%, the tax rate for luxury goods, has been proposed by government as a measure to reduce water consumption further and protect water resources.

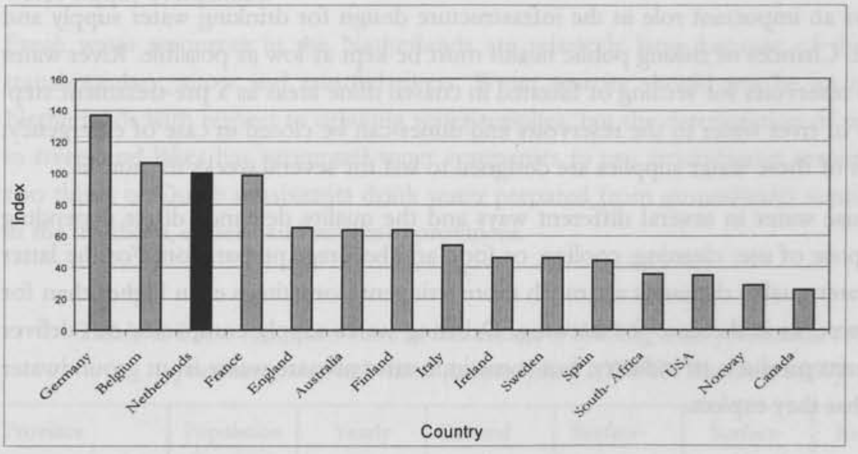


Figure 2. *Price of drinking water per cubic meter for the Netherlands compared to other countries (Dijkgraaf et al., 1997).*

There is a variety of technology available for drinking water preparation, ranging from simple sand filtration for clean source water to chemical treatment or membrane technology for more polluted water. Ozone is used to disinfect drinking water. Treatment sludge is incinerated unless it has economic value for brick manufacturing because of its iron content.

2.2 Wastewater collection and treatment

In the Netherlands about 97% of the domestic wastewater is collected and treated. Wastewater collection systems are designed to collect and discharge both storm water and wastewater. This type of sewer system is being replaced by separate sewers for storm water and domestic wastewater. The joint discharge system is cheaper than a separate one but cannot prevent pollution from sewer overflow during peak loads of storm water (e.g. during a heavy thunderstorm). Each year, as many as 16.000 sewer overflows may occur, 3.600 of which will cause severe problems like stench, pollution (both visual and chemical), fish kills etc. The separate system reduces chances of sewer overflow and increases possibilities for water infiltration and recharging local groundwater resources. Separation of sewage and storm water discharge also benefits the efficiency of operation and the capacity management of wastewater treatment plants.

The effluent of wastewater treatment plants is returned to surface-waters or directly to the sea. The norms for organic load, N and P loads are strict and in accordance with the Helsinki treaty for reduction of land-based emissions. Dutch wastewater treatment plants are upgrading and expanding their installations to satisfy the stricter norms for nitrogen loads. Effluent discharge to surface-waters by sewage treatment plants requires a permit from the national government and is taxed. Treatment sludge is incinerated at high cost.

Households and industry pay for wastewater collection and treatment. Industry pays according to the discharged waste load but households share costs. The 'users share cost' principle for sewage collection and treatment is maintained by both cities and water boards.

3. Water infrasystems

3.1 Water infrasystems and surroundings

Water works, or water infrastructures, are constructed, maintained and operated with a sole purpose: the production of drinking water and water services like water supply, water treatment and management of (ground)water levels. Water infrastructures are networks of pipes for unidirectional transport of drinking water or sewage (flow processes) and water treatment plants supporting mechanical, chemical or biological purification (point processes). The organizations that deliver water and water services have a structure of their own and operate within certain constraints set by market, socio-political culture and government regulations. We are especially interested in the interaction of infrastructures and their physical, social and economic environment.

Infrastructures are physical systems and the organizations of the water industry can be described as institutional systems. We view the design and operation of both types of systems as inextricably linked. Therefore, we consider water infrastructures and the organizations that own, maintain and/or operate them as one water infrasystem (Weijnen and Bosgra, 1999).

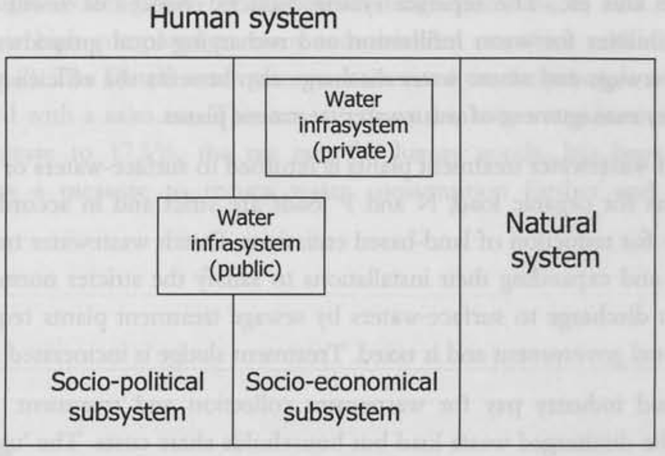


Figure 3. *Place of water infrasystems in regard to human and natural system.*

An integrated systems approach is warranted to understand the factors influencing the design of water infrastructure, the type and quality of products and services (see table 2 for *our* definition of these terms) which are in demand, and their pricing. From a systems perspective, the world can be viewed as composed of a human system and a natural system. Water infrasystems are man-made and thus part of the human system (Figure 3). The human system consists of two subsystems: the socio-political subsystem and the socio-economic subsystem. The socio-political subsystem regulates water infrasystems with regard to task setting, product and service specifications, management and control, and institutional design. The socio-economical subsystem influences water infrasystems through consumer demand, technology development, and competition for resources and clients. Water infrasystems are embedded in the socio-economic subsystem since they produce services that have social and economic value. The organization of a publicly owned water infrasystem is part of the socio-political system and therefore we see such infrasystems as part of both subsystems (Figure 3). The natural system supports water infrasystem activity. One could say that the water infrasystem consumes ecosystem services to produce water services. The dependence of the water infrasystem on the natural system is obvious: its potential for the water production and water treatment depends directly on the hydrology, water quality and epidemiology of the water systems it exploits (Van der Ploeg et al., 1997).

3.2 Water infrasystem design parameters

The design of water infrasystems relates to products and services, to infrastructure hardware and the processes they support, and to the institutional design for ownership, management, and operation. The criteria for a water infrasystem design are set by the consumer and regulatory demands for products and services on one hand, and regulatory demands for organizational design on the other (Table 2). The government may use different instruments to ensure that water companies safeguard public health and environmental quality and to protect equity of access to water services, a primary condition for urban living. In addition, government and market both can regulate the diversity of products and services that water companies deliver. Consumers set demands for the quality and affordability of products and services they pay for, as well as for the sustainability of the processes and hardware that are used to produce products and services. The criteria for design, which can be derived from consumer and regulatory demands, must be translated into specific choices for the design variables of the processes and hardware involved, and the organization which owns, maintains and/or operates the infrastructure.

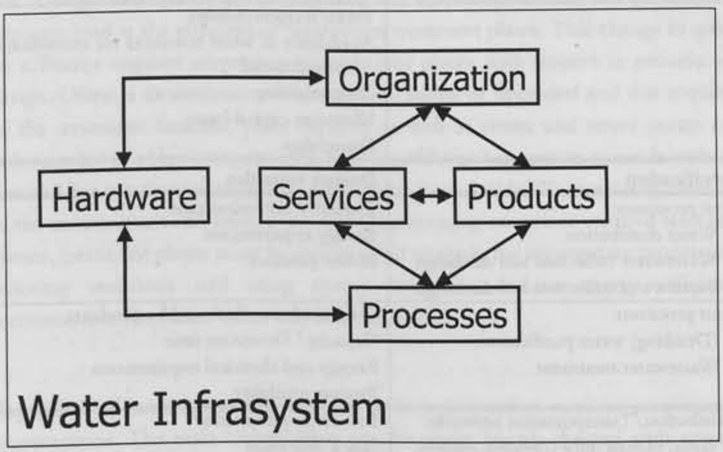


Figure 4. *Reciprocal influences on design characteristics among water infrasystem elements*

Table 2 is a list of variables and no attempt is made to show the complexity of the relationships among these variables. Figure 4 depicts how the different elements of the water infrasystem relate to each other with respect to their design characteristics. The designs of all elements are subject to influences from outside the water infrasystem. The availability of natural resources, new technology, social, human and economic capital have an impact on the type of hardware and processing technology available to a water company. The choice of a specific hardware design can have impact on organizational design. For instance, uncoupling of storm water and sewage collection in favor of separate sewer systems can only be done through new co-operative activities by cities and

Table 2. A specification of water infrasystem elements, criteria for design and design variables

Key word	Specification	Criteria for design which are determined by consumer demand & government regulations
Products	Urban water: Drinking water Household water Fire protection water Industry water: Process water Cooling water Food and beverage water	Quality Quantity Pricing
	Effluent Treatment sludge Odors and other emissions	Chemical composition Quantity Economic value
Services	Water supply	Public health Environmental quality
	Sewage treatment	Reliability/Ease of access Pricing
	(Ground)water level management	Risk (chance of flooding and potential for damage)
Organization		Tasks/Responsibilities Availability of water resources for extraction or for waste disposal Accountability Minimum capital-base Ownership
Key word	**Specification**	**Design variables**
Processes	Flow processes: Water distribution Wastewater collection and discharge Stormwater collection and discharge	Capacity/ detention time Energy requirements In-line pressure
	Point processes: (Drinking) water purification Wastewater treatment	Quality of raw, final, and by-products Capacity/ Detention time Energy and chemical requirements Process reliability
Hardware	Distribution/Transportation networks: Wells, pumps, pipes/sewers, meters, water appliances, etc.	Process requirements Space allocation Investments Service life Inertness of materials
	Treatment plants	Process requirements Space allocation Investments Service life
Organization	Scale	Size service area Extent of client base Turnover, sales
	Organizational design	Division of tasks (ownership, maintenance, operation) Composition of client base (households/industry) Nature of ownership (public/private/nationality) Cooperation/ Integration with water companies
	Economics	Amount of deferred investments Capital-base

water boards. Excess capacity for water distribution or water treatment may create opportunities to engage in private market activities without harming the execution of public tasks. The type and quality of the products and services offered by water companies depend, to a large extent, on the processing technology in place. Changes in product or services specification thus require changes in process design. Take the example of improving the flavor of drinking water: removal of odorous compounds may require the addition of active carbon filtration to the water preparation process. The mutual relationship between hardware and processing technology designs does not need explanation. Social, economical and political developments determine the demand for specific products and services and set boundaries for institutional design. The public character of water organizations in the Netherlands is a perfect example of this.

The following example shows the complexity of these relationships.

> The Helsinki accord has imposed strong restrictions on land-based emissions to the North Sea. The government of the Netherlands has signed this accord and enforces reductions of nitrogen-load in the effluents of wastewater treatment plants. This change in quality demands on effluents requires adaptation of treatment plants with respect to process and hardware design. Often, a denitrification step must be added or upgraded and this requires expansion of the treatment facilities, plant capacity as well as sewer and sewer pump capacity. This enforcement of more stringent standards for effluent happens to coincide with (1) a growing demand for wastewater treatment services by households, (2) increasing taxes on emissions to the environment and energy use, and (3) increasing economic value of land in urban areas. Hence, treatment plants must be (re)designed to apply the appropriate processing technology reducing emissions and using space, chemicals, and/or energy efficiently, and to accommodate for future water treatment demand.
>
> Expansion of existing treatment plants or (re)construction requires cooperation of several organizations. The main stakeholders are the water board, charged with sewage treatment and management of water quality, and city councils. City councils have a stake in keeping costs of sewage treatment as low as possible (on behalf of the consumers within the service area) but also in spatial planning and perhaps in selling land. The willingness and ability for cooperation between these organizations ultimately decides the criteria for design for treatment plant expansion and, inherently, the costs of water treatment services. The expansion of wastewater treatment services to households and industry may have consequences for the organization of the water infrasystem. For example, if expansion affects the budget of the water board, it may change the respective representation of taxpayer categories in the democratically organized water board.

In the Netherlands, the different water utility companies have a major impact on public health, but also on economic development. Society now demands high reliability and quality for a low price, but these demands change as society changes. The City of Amsterdam had a dual water system at the beginning of this century, supplying high quality dune water and lesser quality lake water. The dual system was dismantled after time but now, a mere hundred years later, the city is reconsidering installing a dual water system in new housing complexes. To better understand what design variables are the key variables for infrasystem design in the future, we need to analyze the driving forces for change that originate in the surroundings of the water infrasystem.

4. Forces driving structural changes in water infrasystems

Water infrasystems, their technical and institutional design and the range of products and services they supply, are subject to change. Even in the last decade we have seen major changes in water infrasystems, all induced by changes taking place in the system surroundings. The competition for groundwater resources has increased which has reduced groundwater levels to the point that nature areas are damaged beyond repair; new processing and information technologies have been developed and found application in water distribution and water treatment; investment policies have changed to allow for cross-border leasing of water treatment plants; the demand for bottled water is rising; consumers are more aware of global water problems and increasingly willing to change their own water consumption rate; industrial organizations reduce their dependency on central water production and wastewater treatment by investing and operating in groundwater extraction and water treatment plants; international policies have set stricter standards for land-based emissions to the North Sea and thus increased the need for wastewater treatment capacity.

Water infrasystems must respond to these changes in the natural system, the socio-economic subsystem and socio-political subsystem. But what is the best way to respond? Should the infrastructure design or the organizational design be altered or should measures try to affect the demand for water services? The following example gives an impression of the variety of measures that organizations in a water infrasystem can take in response to one particular change, namely the declining availability of groundwater for drinking water preparation. Obviously, there are more measures thinkable than those that are presented here. However, this case does show triggers (declining groundwater supply and a growing consumer demand for environmental friendly water services) and constraints (European and national quality standards) that influence water infrasystem behavior.

> Example 1: groundwater supplies in the Netherlands are declining *(natural system)* because of unsustainable extraction rates by the agricultural, industrial and drinking water sectors *(socio-economic subsystem)*. The Dutch national government *(socio-political subsystem)* now puts more stringent limits on the use of groundwater to ensure future water supplies, and to protect the ecosystems that depend on these water supplies also. The drinking water supply

companies *(part of the water infrasystem)* in the Netherlands are responding in several ways. First, drinking water companies put efforts into limiting water consumption by stimulating changes in consumer behavior. They have used advertisements and extension leaflets to stimulate consumers to install water saving appliances. This campaign has been very successful: drinking water consumption has dropped from 140 to 128 liter per capita per day. One water supply company has changed the billing system for its water services and now co-operates with the water board in charge of wastewater treatment. The two organizations hope to limit water consumption and wastewater production by linking the rate of water consumption directly to the charge for wastewater treatment.

Second, water companies shift away from using high quality groundwater to lower quality surface-water for drinking water preparation. This requires investments in new water treatment technology and increases costs of production. Consumer demand for an environment-friendly water supply is growing *(socio-economic subsystem)* and consumers question the use of high quality drinking water for household purposes. They are searching for ways to reduce costs of drinking water consumption to the environment and to their own pocket. In several cities, e.g. Amsterdam, consumers are asking for dual water systems that deliver household water (quality B water) and drinking water (quality A) separately. Drinking water companies *(water infrasystem)* are now participating in projects where dual water systems are implemented in new buildings. The companies invest in two separate water distribution systems and in new facilities for the preparation of household water. The 'users share costs' principle is being applied to the deferral of investments for these projects. Thus, the water price increases for all consumers in the service area regardless of their access to dual systems. The feasibility of installing dual water systems depends on the quality norms imposed by government *(socio-political subsystem)*. The Dutch government is developing quality norms for household water, taking into account the risks of dual water systems to public health and the developments in European policy. Drinking water companies *(water infrasystem)* are lobbying in the policy arena to influence the norm setting as a way to create opportunities for new products (household water) or to mitigate perceived threats to their status quo.

The above example describes past and current behavior of the Dutch water infrasystem in response to changes in groundwater availability. Can we also predict the future behavior of a water infrasystem in response to this or other changes? Before we can answer this question, we need to know which changes may alter the water infrasystem structure and cause it to behave differently. Potential changes of large magnitude and with important implications for the way a system works are called structural changes (RAND, 1997).

Table 3 lists forces which drive structural changes with implications for the way the water infrasystem works. We do not know the impacts of these *driving forces*, nor the rate at which they incur changes. For instance, we know that increasing competition for groundwater resources will have an impact on groundwater supplies, but we do not know how this will affect the availability of groundwater for drinking water preparation. Right now, drinking water companies are urged to reduce the use of groundwater but regulations could be changed in favor of water services and to the disadvantage of agriculture or industry. With other words: we lack knowledge about the future behavior of the natural system, the socio-economical subsystem, and the socio-political subsystem. The uncertainties that arise from this lack of knowledge are called structural uncertainties if they are variables from outside the water infrasystem, if their future value has consequences for infrasystem design, and if the realm of these consequences is unknown to water infrasystem planners.

Table 3. *Forces driving structural changes in the water infrasystem and the parameters they affect in unknown ways (structural uncertainties).*

Driving forces for change in the water infrasystem	Structural uncertainties
	Within the natural system
Increasing competition for groundwater resources	Groundwater availability
Deterioration of environmental quality. Calamities	Water quality of groundwater and surface-water
	Within the socio-economical subsystem
Individuation of consumer demand and behavior	Consumer acceptance of (health) risks Demand for type of water products and water services Water consumption rate/Wastewater production rate
Increasing competition for economic resources	Price of energy, interest rates, land Efficiency, safety and costs of (de)central water treatment
Technology development	Automation of water distribution and quality control
Urbanization	Population density patterns Space availability (above ground and underground)
Opening of European market for water services	Implementation rate of market changes
	Within the socio-political subsystem
Political values	Regulations
Equity in access to water services	Tax on water services.
Sovereignty in water management	Ownership of water services organizations
Public health	Licensing of water services organizations
Sustainability of water management	Permits for groundwater extraction Permits for disposal of effluent and sludge
European policy	Quality and access norms for water services Quality norms for waters receiving effluent

In table 3 we list the forces driving structural change in the water infrasystem and the key variables they affect: the structural uncertainties. We do not know the future values of these key variables but we do know that their future values will have consequences for water infrasystem design.

For instance: groundwater availability is a natural system variable which future value we cannot know since it is subject to so many factors (precipitation patterns, land development, surface-water quality, regulations for groundwater extraction, etc.). A decrease in groundwater availability may prompt managers of the water infrasystem to take either technical measures, changing processes or hardware (water preparation from surface-water, extracting water at greater depth or different watershed), managerial measures (public relations aiming at reducing water consumption), economical measures (increasing prices of groundwater to restrict water consumption), or organizational measures (a merger with a water company with access to groundwater).

There are many linkages between the structural uncertainties listed in table 3, the systems perspective given in figures 3 and figure 4, and the design variables listed in table 2.

1. Driving forces change the water infrasystem surroundings, causing a change in the variables indicated as "structural uncertainties."
2. These structural changes have an impact on the criteria for design of either water services, products, or organization.
3. As a consequence, the design for hardware, processes, or organization must be adapted to accommodate for these changes.
4. These adaptations in infrasystem design have a domino effect: a change in one aspect of the infrasystem design gives cause, reason, or possibilities for other changes.

The following example exemplifies these relationships between forces driving structural changes, structural uncertainties, and changes in infrasystem design. The numbers relate to the explanation above.

Example 2: A change in consumer acceptance of health risk (1) can have consequences for drinking water preparation and distribution. Customer preference (2) changes from tap water to bottled water to the extent that customers refuse to use tap water for drinking water. This opens up a large market for bottled water in the Netherlands. Who would be the major provider of drinking water (3) in such a situation? The existing drinking water companies or beverage companies who already have a network in place for distribution of bottled products? Is it possible that drinking water can be supplied (3) in gallon

drums and delivered from door to door, the way it is being done in the United States? What would this decentralized and possibly private water supply mean for public health and accountability regulations? Could such a change in consumer behavior lead to a change in task-setting, opening the way for delivering only quality B-water to households (4)?

Table 4. *Three possible futures that would affect water infrasystem behavior*

Possible futures (Scenarios) / Structural uncertainties	Customers first	Liberalization	Chemical disaster
Groundwater availability	↓	↓	↓↓
Water quality surface-water	↓		↓↓
Consumer acceptance of (health) risks	↓	↑	↓
Demand for type of water products and water services	↑↑		
Water consumption rate	↑		↓
Competition for economic resources.	↑	↑↑	↑
Technology development		↓	↑
Number of regulations		↑↑	
Equity in access to water services		↓	

The driving forces listed in table 3 may occur simultaneously and the impacts can counteract or reinforce each other. Our future is shaped by the combinations in which these forces occur and the direction of the changes. Thus: the identified structural uncertainties can be used to develop scenarios for possible futures. We have drawn up three examples of possible futures based on specific (and not unlikely) combinations of structural uncertainties and have indicated how these uncertainties change. In *"Customers first"* the major forces driving change are the decrease of safe drinking water supplies and environmental quality on one hand, and the increased concern for health risks on the other. The scenario *"Liberalization"* sketches a possible future after opening up the European market, including the market for water services. Competition for economic resources increases, the amount of regulation dictated by the European Union increases, and technology development goes slowly. In the third possible future, *"Chemical disaster"* a chemical spill occurs in one of the riparian states located in the upstream section of a major river basin. The chemical spill threatens the source of drinking water for 5 million people, at a time that groundwater availability decreases and technology develops rapidly.

The descriptions of these possible futures concern elements of the water infrasystem surroundings only. How will the water infrasystem behave under these circumstances? The development of scenarios is meant to aid in the design of strategies to deal pro-actively with the uncertainty of the future. We cannot predict how water infrasystem managers should or might adapt its operation or design in the futures sketched in table 4 or any other possible futures. What are robust water infrasystem designs, designs that are suited for many possible futures?

5. Conclusion

Infrasystems operate in a dynamic society and the social, economical, political and cultural aspects of human activity change in ways we cannot predict. Neither can we predict the consequences of these changes for the operation and design of infrasystems. With regard to water infrasystems, the interdependence of human activity and the state of the natural environment are especially important. Water infrasystem design differs according to the availability and characteristics of local water resources and determines the sustainability of these resources. There are many potential forces that drive structural changes in a water infrasystem, such as individuation of society, urbanization and liberalization of the European market. We have listed those that we know but, inherent to the topic of futures, this list cannot but remain incomplete.

A most appealing question for research on infrasystems is the treatment of uncertainty in design, and especially the design of the physical infrastructure. The expected life span of the major physical infrastructure of water infrasystems is 30-50 years. The robustness of the hardware design is thus crucial to the robustness of the overall infrasystem design. Hardware design determines the capacity for water production and water services. Changes in the demand for these services affect the demand for capacity, and alter the effectivity of the hardware design. Water infrasystems can share capacity or lease capacity to each other with long term contracts. A change in the organizational scale of a water infrasystem also has consequences for capacity management, and the possibilities for capacity expansion or contraction determine the sensitivity of a water company to changes in the demand for its services. The treatment of uncertainty for water infrastructure design requires methods for capacity estimation and methods for adaptive capacity management. What adaptations in process, hardware, products, services, and organizational design can or should be made to meet future capacity needs? Research on this topic will yield information, which can be applied to future water infrasystem designs. We anticipate that such research can be transferred to other physical infrastructures.

References

Dijkgraaf, E. et al. (1997), *Mogelijkheden tot Marktwerking in de Nederlandse watersector*, Oc*f*eb, Erasmus Universiteit, Rotterdam, the Netherlands (in Dutch).

Huisman, P., W. Cramer, G. van Ee, J.C. Hooghart, H. Salz, and F.C. Zuidema (1998), *Water in the Netherlands*, Netherlands Hydrological Society, Delft, the Netherlands.

Kjellén, M., and G. McGranahan (1997), Urban water- towards health and sustainability, in: *Comprehensive assessment of the freshwater resources of the world: Background documents*, Stockholm, Sweden: Stockholm Environment Institute.

Ministry of Transport, Public Works and Water Management (1999), *Fourth National Policy Documents on Water Management. Summary*, Sdu Uitgevers, The Hague.

van der Ploeg, M. and P.W.G. Bots (1997), *A conceptual framework for modeling environmental problems in river basins from a sustainability perspective*, Proceedings of the International Association for Impact Assessment Conference, New Orleans

RAND/European-American Center for Policy Analysis (1997), *Scenarios for examining civil aviation infrastructure options in the Netherlands*, DRU-1513-VW/VROM/EZ, Delft: EAC.

Rees, J.A. (1998), Regulation and private participation in the water and sanitation sector, *Tac background papers no.1*, Stockholm, Sweden: Global Water Partnership.

UNESCO (1998), *Summary and recommendations of the international conference on world water resources at the beginning of the 21st century*, Paris, France: UNESCO.

VEWIN (1997), *Statistics of drinking water supply in the Netherlands 1996*, Rijswijk: VEWIN (in Dutch).

Walker, W.E., S.A. Rahman, and J. Cave (1998), *Adaptive policies, policy analysis, and policymaking*, Leiden, the Netherlands: Rand Europe.

Weijnen, M.P.C. and Bosgra, O.H. (1999), *An Engineering Perspective to the Design and Control of Infrastructures - Explorations into a Generic Approach to Infrastructure Scenario Analysis* (this volume).

Final discussion

At the end of the First Annual Symposium held in Noordwijk, November 19, 1998, the papers presented were discussed in a plenary session and the participants were invited to give their views and inputs to the research program of the Delft Interfaculty Research Center for the Design and Management of Infrastructures. The discussion was structured to address four key issues.

1. Priority goals

As the liberalization of the utility markets is one of the main driving forces for the research, Künneke (TU Delft) posed the question: 'Can we trust the market'? He clarified this as a normative statement, i.e.: Do we want to trust it? Whereas both bread and drinking water are basic necessities, the market is trusted for the distribution of bread, but it is not trusted (yet) for the distribution of water. He furthermore emphasized the importance of transaction economics, suggesting that changing transaction costs in some infrastructure sectors may have a larger impact on the evolution of infrastructures than technological developments. In his opinion, this underlines the importance of the interfaculty research program, as technology obviously does not stand alone as a driving force for change in infrastructure sectors. Arnbak (OPTA, TU Delft) agreed it is not only institutional change that changes infrastructure sectors. He showed a prescription for a light(er) regulatory regime, showing that public (government owned and controlled) utility service should only be applied in case of market failure that can neither be redressed/repaired by the general rule of law, nor removed by ex ante sector-specific rules and regulations on relevant players in the market. Braga (World Bank) added the question: 'Can you trust the government'? In his opinion there are many, particularly Third World countries, where the answer would be negative. Ketting (Senator) reacted that the question of whether or not the market can be trusted is a non-question, as the market has no morals. The market forces are erratic and many players use their powers to manipulate regulators.

The market is only a means to an end, i.e. the goal of achieving high quality utility supply in the most cost-effective way. An assessment of the effectiveness of liberalized markets in achieving this goal, should be one of the research questions to be tackled in the research program.

2. Operational targets

According to Chamoux (Université du Havre) the operational target of the research should be to achieve results that are valuable to users, irrespective of a free market or a regulated market being in effect. For that reason, the waste sector approach presented by Reuter appealed to him. As today's investors need to view their strategies in a cross-sectoral perspective, the cross-sectoral approach of the research program is eminently suitable. He added that the Delft Interfaculty Research Center should also contribute to the education of engineering graduates with the ability to work across sectors, as engineers with cross-sectoral abilities are in large demand. Dijkema (TU Delft) expressed his agreement with Chamoux on the need to break through the existing regulatory system boundaries. The infrastructure models to be built in the course of the program should indeed be applicable anywhere, irrespective of the regulatory environment. The question remains if it is realistic to expect that such flexible models can be made without the cost of loosing all content value.

Do not tailor the research to the existing regulatory frameworks, but make it your objective to produce robust results that will still be valuable when the regulatory system changes. Generic, cross-sectoral research is in great demand e.g., by investors. Make sure that the research experiences and results are reflected in the engineering education at TU Delft.

3. Interfaces between the disciplines

Ehrenfeld (M.I.T.) pointed out the need for systems analysis. The question should not be where to put up boundaries, but how to make boundaries between disciplines disappear. Although the researchers need to set boundaries to their research projects, the major risk at this stage in the program is that boundaries will be set too soon. For the moment his advice is 'to let a thousand flowers bloom'. Braga (World Bank) strongly agreed with Ehrenfeld, in view of the multi-disciplinary nature of the research program. He advised though to make absolutely clear why and how this research is going to be unique. The difficult road of multi-disciplinary research is not followed for the fun of working together, but for the unique goal of acquiring fundamental understanding of the behavior of infrastructures, from a generic perspective. There are obvious results to be gained from a generic research perspective, but it is not yet clear that the generic perspective will yield meaningful results in all areas of the research program.

Be aware of non-constructive boundaries between disciplines and research projects. The multi-disciplinarity of the research is an essential condition to achieve the research goals. There is, however, no standard recipe for finding the balance between mono- and multi-disciplinarity. You will have to learn by doing.

4. Sector-specific versus generic research

As the most important generic research question, Ketting (Senator) proposed: 'How to design robust infrastructures, e.g., how to achieve robustness by flexibility'? He agreed with Chamoux that the research should not be conditioned by the current institutional setting in the Netherlands, but prove its worth irrespective of a changing institutional context. Ketting, however, also advised the researchers not to pursue the generic approach too far. Ideally, an early evaluation should be made, if possible, to see where a generic approach can be effective and where it will fail. His concern focused on the risk of losing content value in the generic projects. Berkhout (Vice President Research, TU Delft) urged the research team to demonstrate once and for all that multi-disciplinary research in a broad subject area as in the Design and Management of Infrastructures research program, does not necessarily yield shallow results. Given a common framework, e.g., for communication and model sharing purposes, each specialist can contribute his or her piece to the puzzle, in such a way that each piece is worthwhile and innovative, even from the mono-disciplinary point of view. He agrees with the need to change from ad hoc decisions to a more structural, systematic approach to the questions faced in infrastructure design and management. He supports the multi-disciplinary approach needed to achieve this target. If the risk of superficiality can be eliminated successfully, large economic and societal benefits can be obtained. This is precisely why the Delft University of Technology has embarked on the course of establishing interfaculty research centers, and has allocated Dfl 80 million to the execution of this and similar multi-disciplinary research programs.

Demonstrate that multi-disciplinary research is not doomed to be shallow. Develop a common framework and a common vocabulary to facilitate cross-sectoral and cross-disciplinary communication and model sharing. Do not base the research upon the current situation, e.g., the current institutional setting, but think at least 10-20 years ahead.

TU Delft

Delft University Press

ISBN 90-407-188[illegible]