

Infrastructural Innovation. Flexibility, Generativity and the Mobile Internet

Ole Hanseth
Department of Informatics, University of Oslo, Norway
oleha@ifi.uio.no

Petter Nielsen
Telenor R&I, Norway
petter.nielsen@telenor.com

Abstract

This paper addresses issues related to how to enable broadest possible innovative activities by infrastructural technology design. In this paper, we focus on the development of high level services based on mobile telecommunication technologies which we for matters of simplicity term the development of a Mobile Internet. The focus of our analysis is how features of the technology itself enable or constrain innovations. We do so by looking at a few embryos of the Mobile Internet (primarily the Norwegian CPA platform, but also two pre-CPA platforms in Norway and Japan's i-mode) through the concepts of *end-to-end* architecture, *programmability of terminals* and *generativity*.

This analysis illustrates that the change from closed infrastructures like MobilInfo and SMSInfo to more open ones like CPA and i-mode increased the speed and range of innovations substantially. At the same time the differences between CPA and i-mode regarding programmability of terminals, and the billing service provided by the CPA network enabling the billing of individual transactions, also contributed to, at least, basically the same speed and range of innovations around CPA as i-mode in spite of the huge differences in investments into the networks made by the owners. But the analysis also points out important differences between the Internet and the existing Mobile Internet regarding in particular technological constrains on innovations. It also points out important ways in which powerful actors' strategies inhibit innovations and how they embed their strategies into the technology and, accordingly, creates technological barriers for innovation. These issues are first of all linked to the programmability of terminals. Therefore, we conclude that the by far most important issue regarding the enabling of more innovations related to the Mobile Internet is to improve the terminals programmability.

1. Introduction

The Internet as we know it is a standard and infrastructure spurring innovation and fuelling entrepreneurship in an unprecedented fashion. Its counterpart for mobile phones – the Mobile Internet – has however only modestly, at its best, become such an arena for entrepreneurship. The success of the Internet has been explained in different ways, but a central factor has been the flexibility of its design. With an outset in the concept of loose coupling from software engineering and the argument for locating intelligence in the fringes (the so called end-to-end argument) in network architecture design, Jonathan Zittrain (2006) have ventured more deeply into these matters and coined the key success factor of Internet as its *generative* capacity. More particularly, Zittrain define generativity as a technology's capacity for leverage across a range of tasks, adaptability to a range of different tasks, ease of mastery, and accessibility. This conceptual framework illuminate that the essential flexibility of Internet as a infrastructure is not limited to its modularity and decentralized network architecture, but also the way in which it enable and leverage innovation performed by third-party contributors. Thus, infrastructures like the Internet require a different kind of flexibility and a central question become how this can be accommodated.

In this paper, we discuss two 'instances' of the Mobile Internet, namely the Japanese i-mode and the Norwegian CPA. By applying the concept of generativity, our aim is to disclose and to better understand the success and failure of attempts to create the Mobile Internet. We will in this paper present some proposed solutions to some such needs based on 'the Internet experience', and we will then discuss how these flexibility needs and solutions apply to the mobile Internet. By comparison with the Internet, we discern what generative capacities these platforms offer, how the capacities have developed and the motivations behind their development. Our contribution thus lies in a deeper insight in the successfulness of different approaches to the Mobile Internet in the way in which they support innovation. Theoretically, we also contribute to the literature on standardisation with an extended perspective on flexibility, namely generativity.

The rest of this paper is designed like this. In section 2, we frame the concept of standardisation in the broader literature on standards and flexibility. In section 3, we describe our research methodology and approach. In section 4, we introduce the CPA and i-mode in brief, before we in section 5 compare and discuss their generative capacities. In the last section 6, we draw implications related to the further developments of the Mobile Internet as well as reflect on the applicability of the concept of generativity.

2. A changing 'World of Standards' and the new needs for flexibility

The research presented in this paper is part of a growing interest in research on infrastructure standardisation in general and standardisation and flexibility within ICT in particular. This increasing interest is a result of the growth in the number and importance of standards as well as the transformation of the 'world of standards' that is a part of the so-called convergence of telecommunications and information technologies (Brunsson and Jacobsson 2002). The mobile Internet is among the newcomers that populates this 'world'.

The necessity of flexibility, as well as the contradiction between standards and flexibility has been a central topic in the literature on standardisation. In this section, we describe how changes in information and telecommunication technologies (ICTs) have brought new requirements for flexibility. From a perspective on flexibility as the seamlessly fit of objects into larger systems as well as the standards adaptability to contextual changes, the very nature of today's ICTs also require them to be flexible in the sense of enabling and promoting innovation, and in particular innovation by third party contributors. Namely, they must be *generative* to endure and grow.

Technological changes within telecommunications and ICT have brought many new actors into this field. Telecommunication standardisation used to be taken care of by (a limited number of) service providers and equipment manufacturers. With the digitalisation of telecom, computer manufacturers and software

companies also got involved. This technological change opened up possibilities for a broad range of new services. The development of such services involved even more actors – even users (big and small companies, professionals like medical doctors, etc.) (Jakobs 2000). These services also implied a need for new kinds of standards which raised new challenges. Some of these new and hard challenges were related to the fact that the standards for high level services needed to satisfy much more complex user practices (in particular compared to the simple ones supported by traditional telecommunications which just enabled users to dial a number, talk, and hang up.) (Bowker and Star 1999; Foray 1994; Hanseth and Monteiro 1997; Jakobs 2000). The ongoing ‘convergence’ of the ICT and the media sectors further increases the current technological and institutional complexity and variety as well as increases the speed of change. These changes partly triggered, and were partly taking place in parallel, with the deregulation of the telecommunication sector. The deregulation increased competition, which again brought more actors into the picture at the same time as it changed the relations between the actors involved.

Standardised systems such as large scale ICT solutions and infrastructures tend to become accumulatively change resistant as they grow and diffuse (Egyedi 2002; Hanseth et al. 1996). Thus, to endure, these systems have to be prepared for change to avoid becoming obsolete (Tassej 2000). Standards must allow for growth and change through various means of flexibility to avoid this. Flexibility of standards and infrastructures has become increasingly important as the “world of standards” has changed as described above. The fundamental principle for making technological systems flexible is modularization, allowing some components to be kept stable while others are changed without implications for the rest of the system. This also applies to standards and infrastructures (Hanseth et al., 1996).

2.1. Flexibility and generativity

The success of the Internet has triggered many discussions about lessons to be learned for how to

develop infrastructures, and in particular large scale ones. In the later years, the 'essence' of the Internet has been much in focus among some 'cyberlaw' scholars discussing regulation of cyberspace. An issue of interest for some of these has been how to regulate cyberspace or the Internet so that unwanted use (for instance distribution of child pornography and music, and film and software piracy) is constrained at the same time as the qualities of the Internet that has made it so successful is maintained. The qualities of the Internet that these scholars have identified as important to maintain is the speed and scope of innovations that the Internet has allowed and triggered – regarding both the Internet itself and its use. We will here introduce three “cyberlaw” scholars and the aspects of the Internet they have highlighted and discussed.

Lawrence Lessig (2001) has stressed the importance of location of functions close to the application that uses the function, the so-called end-to-end architecture, originally proposed by Saltzer et al. (1984). This is a central principle to provide flexibility by systems design. The point this principle is making is that functionality in communication networks only can be appropriately implemented if based on knowledge that only exists close to the applications standing at the endpoints of a communication system. Thus, the network should not control how it grows, the applications should. Both Lessig (2001) and David (David 2005) exemplifies this argument by illustrating the Internet as a network where intelligence is in the fringes. Since the network is not optimised for any application but open for and inviting the unexpected and surprising, innovations can flourish without changes in standards or the network itself. The important role of the end-to-end architecture in the success of the Internet is also underscored by historian Janet Abbate (1999) in her analysis of the history of the Internet. In particular, she demonstrated the substantial difference this end-to-end architecture represented in relation to traditional telecommunication (where all functionality is in the network and not in the ends (i.e. the telephones)) and argues convincingly that this was an important explanation why the Internet won the 'war' against the ISO/OSI standards (Abbate 1996).

Yochai Benkler (2006) develops this end-to-end argument one step further by underscoring the mutual dependence of the end-to-end architecture of the network and (easily) programmable terminals in terms of general purpose computers. Benkler base his argument on contrasting programmable computers and appliances. An appliance is a device with a limited and well defined set of functions which (normally) cannot be modified after the users have bought it. Typical examples include washing machines, radios and telephones (traditional ones, at least). By large, such devices have computers inside, but their software cannot (normally at least) be modified by its users. Benkler is worried that several proposals for increasing security and stop harmful use of the Internet, i.e. cyberspace regulation, will constrain the Internet users ability to program their computers, i.e. turn them into appliances. An example of this is found in the proposed 'trusted computing' technology and how this may be implemented and ways in which it might be enforced by law.¹

Jonathan Zittrain (2006) develops this argument yet another step by means of the concept of *generative technology*. He argues that the success of the Internet is closely linked to its generativity, and that regulation of cyberspace must carefully avoid doing harm to this. Generativity is "the essential quality animating the trajectory of information technology innovation." (ibid., p. 1980). It "denotes a technology's overall capacity to produce unprompted change driven by large, varied, and uncoordinated audiences" (ibid.). Zittrain argues that the grid of PCs connected by the Internet has developed in such a way that it is exceptionally generative. Zittrain defines generativity more detailed as a function of a technology's capacity for leverage across a range of tasks, adaptability to a range of different tasks, ease of mastery, and accessibility. This also makes it clear that generativity includes but also extends the matters of flexibility to include more than modularity and end-to-end architecture. *Leverage* describes the extent to which these objects enable valuable accomplishments that otherwise would be either impossible or not worth the effort to achieve. *Adaptability* refers to the breath of a technology's use without change and

¹ This argument is spelled out in detail by Benkler (2006), see in particular pages 409 - 10

the readiness with which it might be modified to broaden its range of uses. A technology's *ease of mastery* reflects how easy it is for broad audiences to adopt and adapt it: how much skill is necessary to make use of its leverage for tasks they care about, regardless of whether the technology was designed with those tasks in mind. *Accessibility* – the more readily people can come to use and control a technology, along with what information might be required to master it, the more accessible the technology is.

In the remainder of this paper we will discuss two examples of Mobile Internet infrastructures and discuss the importance of generativity in relation to such infrastructures and how generativity can be promoted.

3. Research methodology

Standards are widely accepted as being of strategic value, thus standards develop through a process where multiple actors pursue their strategies and agendas. Our research approach is based on an understanding of the processes of standard making as being open and situated as well as being understood differently by the various actors involved. Our empirical data is based on an empirical case study of the CPA, as well secondary data such as books and papers related to i-mode. Inspired by Star (Star 1999), our 'reading' of how CPA emerged revealed a highly complex process that was not primarily network operator driven. Further insights were gained into local contingencies, the properties of the standard and the achievements of those engaged in developing the standard.

The research presented here started in 2002 and continued until late 2004. Since CPA appeared as inseparable from its context, a case study approach was adopted (Yin 1994), following an interpretative perspective (Klein and Myers 1999; Orlikowski and Baroudi 1991; Walsham 1993; Walsham 1995). We found our role as researchers to involve describing, interpreting, analysing and understanding the social world of the involved actors (Klein and Myers 1999; Orlikowski and Baroudi 1991).

Starting out by interviewing the manager of the CPA within one of the network operator, our attention was directed towards how close the standard was interrelated with other (internal) technical platforms as well as actors within the business sector. We also found the appearance of the relationship between the various actors and their coordination interesting, which further guided us also to study how CPA was initially conceived and implemented. Thus, to understand the standard, the study reached both back in time towards the predecessors of CPA, out into the business sector as well as out into the more 'global' setting by studying internationalisation attempts.

A total of 39 formal interviews were conducted with managers, heads of sales and system developers in a total of 23 different organisations, official of government agencies and forums (listed in Table 1), including the two Norwegian network operators. The interviews lasted typically 45 minutes to an hour; they were all recorded, transcribed and notes were taken. The interviews did not follow a strict interview guide, but focused on discussing the very nature of CPA, its development and operation. As the interviews progressed, certain issues were also identified and focused on. In addition to the interviews, data was also collected from studying standard documents and specifications, websites and the trade press.

Table 1 Interviews

<u>Type organisation</u>	<u>No. interviews</u>
Network operator	18
Aggregator	6
Small content provider	5
Integrator	2
Forum/consortia	2
Government	4
Content producers	2
Total	39

While giving a broad understanding of the standard as well as its context, this approach came with certain challenges. Since we did not operate within the borders of one or a few organisations, we had to

negotiate access and justify the participation of the interviewees in a variety of different organisations, ranging from 5-men businesses to network operators with 20,000 employees. While this required different approaches to gain access, maintaining access to all these organisations was not feasible. Another challenge was to identify the important actors related to CPA, both historically and related to the business sector. To access these 'hard-to-reach' populations, a snowball strategy (Vogt 1999) was used.

The data analysis was interpretive and based on our capacity to conceptualise the essential topics in our data. In our analysis, we broadly focused on the industry's market structure, the nature of the services and the standard to include a broad context of influential factors as the actors' aims, institutions and organisations and their strategies. During the transcription of the interviews, the key themes were identified. The themes subsequently acted as input to discussions and guided the further analysis of the transcripts as well as the topic for new interviews. In parallel with this, the research has been guided by presentations and discussions at several seminars, workshops and conferences.

4. The CPA and i-mode as instances of the Mobile Internet

In this section we describe the Norwegian CPA and the Japanese i-mode infrastructure as instances of the mobile Internet. Since our empirical research is limited to the CPA infrastructure, we use considerable more space to introduce this infrastructure than i-mode. I-mode is also a better known case, and we limit our focus to its core, its idiosyncrasies and where it differs from the CPA.

4.1. The predecessors of CPA: Mobilinfo and SMSinfo

In 1997, both Norwegian mobile phone network operators introduced platforms for mobile content services, respectively Mobilinfo and SMSinfo. Inspired by information services on wired phones, the operators started to use SMS as a channel for information services. The technical platform of Mobilinfo and SMSinfo were built more or less ad-hoc and largely independent from the existing internal systems

of the network operators. Even if introduced at the same time, these platforms were outcomes of independent initiatives and launched for the purpose of differentiation. As “walled garden” approaches, it was the network operators that proactively took initiative and invited different information providers to “mobilise” their content. The network operators took the advantage of existing information feeds from providers such as newspapers, traffic information centres and the meteorological institute, and turned them into mobile information services. In addition, the network operators also offered a few non-utility services, such as jokes and bible citations. All in all, it was the network operators that identified potential services, developed the necessary interfaces, took responsibility for the marketing, and controlled the whole value chain of these services.

Even if Mobilinfo and SMSinfo certainly influenced their successor (CPA), they did not create much of a success. First, the network operators had very limited resources to invest in service innovation and development, not to say marketing activities. Since they were the only one in this game, little was done to develop and promote the services. Second, providing non-utility services was seen as a risk of jeopardizing the brands of the operators. Therefore, few new services were introduced. Third, the services could not be premium charged (i.e. SMS messages charged to the user at a rate higher than a standard SMS), rendering any business model impossible. Thus, the stake of the operators as well as the content providers in the platforms never became revenue. As a result, the service offerings did not develop beyond simple information services and fairly prudent entertainment under the control of the network operators.

4.2. The case of the Norwegian CPA

The CPA is an infrastructure for value added content services for mobile phones in Norway. The CPA infrastructure basically supports three tasks; the production, preparation and marketing of mobile content services; transportation (requests and deliveries) of services between producers and consumers; and handling the involved billing transactions. As described above, this service sector was

in Norway up until 1999 based on the network operators providing separate infrastructures and network operators controlled the whole value chain. The introduction of CPA broke up this vertical integration into functional domains, enabling and requiring a range of new roles and actors. The provision of services will in the case of CPA usually involve:

- content producers (producing ringtones, jokes, news, weather information);
- content providers (preparing content for mobile phones and developing new service concepts);
- aggregators collecting a rich variety of content and possibly integrating these in larger service concepts;
- media windows (i.e. newspapers, magazines, TV-broadcasters, etc.) providing space for marketing; and
- network operators providing transportation and billing services.

Linked to the core of its business idea, CPA is a joint undertaking by both the three Norwegian mobile network operators. On the one hand, they provide the same set of functions and a common service level (i.e. functionality, level of capacity, etc.), but not a single technical interface towards content providers. On the other, they provide a common user interface for content service consumers. Thus, CPA enables the consumers to acquire content services through some simple and standardised steps.

A key element of the CPA is that service usage is billed over the regular mobile phone bill. Since the consumer is already registered with one of the network operators, there is no need for cumbersome registration and confirmation of personal data, credit card number etc. Services are requested with an SMS sent to a short number (four digit number such as 2004), and when the network operator to which the consumer subscribe receives the SMS at its SMSC (message centre), the message is recognized as a service request and the CPA platform forwards it to the appropriate content provider over a TCP/IP connection (as illustrated in figure 1).

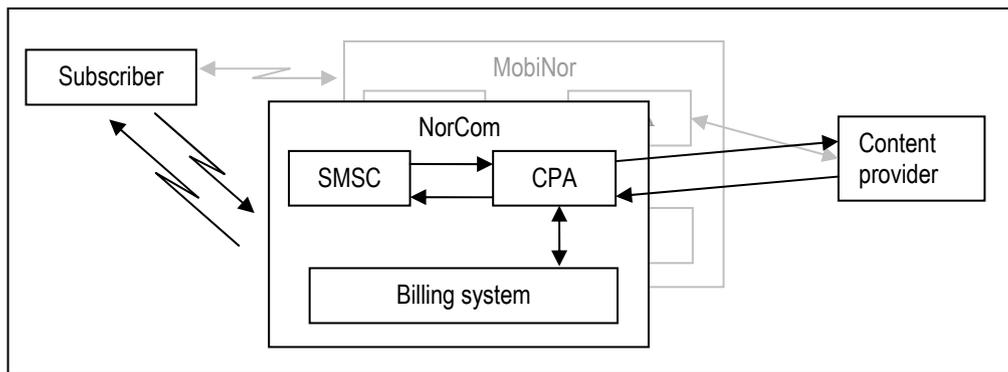


Figure 1 Content services transactions on the CPA

When the content provider receives the request, they produce and return the requested content back to the network operator. In addition, the content provider also specifies the rating class of the service, i.e. the cost which the consumer is to be charged. The network operator requests their billing system with a CDR (Call Data Record) to handle the request according to the rating class before returning the content to the customer. Finally, when the subscriber pays his mobile phone bill, the revenue is split based on a standardised split model between the network operator and the content provider. The content of these transactions are not monitored by the network operators. However, it is to their discretion to react to complaints and close down services they find violating Norwegian law and consumer rights.

One prominent aspect with the CPA is that it is based on 'premium Mobile Terminated (MT) billing'. This means that the return message originating from the content provider is premium rated, i.e. charges the receiver for more than the cost of a regular SMS. Giving the content providers the possibility to charge several times for one request, this further enables subscription or push services as well as services that are requested from other sources than an SMS, in particular the Internet.

Of course, as owners of the underlying mobile telecommunication infrastructures, including the billing systems, the network operators were central actors in the establishment of the CPA infrastructure. But their recent efforts have been modest and catered primarily to increasing traffic. At the same time, several application houses are active in building add-ons to the underlying infrastructure to enable new services and service concepts. Examples include software to collect votes, and produce and visually

present numbers and statistics on ballots in relation to TV-shows, as well as software which presents on-screen comments and questions posted by SMS to discussions/talk-shows.

To summarise, the key components of the CPA are:

- *Business model and revenue sharing model.*

Network operators provide a standardised business model for premium rated content services to the content providers. Operators allow any content provider to distribute their content to all subscribers, offering public market access as well as economies of scale in billing. In addition, they offer a standardized revenue sharing model (i.e. it is non-negotiable). The maximum charge is 60 NOK (approx Euro 7.50) and the predefined revenue split favours the network operator from 54 to 29 percent.

- *Equivalent functionality, architecture and service level*

The content providers are offered basically the same functionally and service level, even if the interfaces to the network operators' implementation of CPA platforms differs. The services are provided by means of an infrastructure based on a common architecture. This infrastructure is, however, implemented differently by the different network operators. Both interfaces are based on content providers initiating a TCP/IP connection to the respective CPA platforms. At the same time, aggregators provide interfaces which hide the differences between the operators' implementations of CPA for the majority of the content providers. This reduces time-to-market for content providers and lowers the barrier of up-front investments to connect to the CPA. Further, it also lessens the administrative burden of network operators as smaller content providers find it more convenient to connect through the aggregators.

- *Administration and use of rating classes and short numbers*

Based on their public market approach, network operators have also standardised their administration and use of short numbers and rating classes. This adds to the transparency of

the market by being the basis for a standardised way of marketing the services.

- *Guidelines for consumer protection*

Further, in order to reduce the risk of 'offensive' services being provided and marketed or marketed fallaciously, the network operators have standardised guidelines describing which services cannot be provided over CPA as well as how to market the services in a consumer friendly manner.

- *Interface for service acquisition*

By providing a standardised interface for service acquisition – the user interface, every mobile phone user in Norway has easy and transparent access to content services. Independent of which operator they subscribe to as well as the type of subscription and calling plan, subscribers can access the same services, from the same short number and for the same price. This also makes the marketing of services simpler as instructions for service requests become easier to read for the consumers.

So, why did the network operators choose this particular approach in Norway? In short, the developers and the promoters of CPA had the experiences with Mobilinfo and SMSinfo and were operating with scarce resources and did not have a strong and convincing business case. Thus, they had to work without the costs of the usual grand marketing campaigns, they had to circumvent the need to change the billing system (as it would have been too cumbersome and time consuming, if possible at all) and they managed to postpone technical systematisation and documentation. Therefore, CPA was developed in a bottom-up fashion where only a few enthusiasts, working for the operators and a couple content providers, set up a pilot version of the infrastructure and a few pilot services using it. The successful demonstration of these attracted more content providers and other actors. As the use of the infrastructure expanded, it was polished and extended and the standard defining it worked out. Rather than a traditional telecommunication standardization model, i.e. a formal top-down process focusing on

formal and detailed technical specifications, the standardization model was more driven by 'rough consensus and running code' – i.e. in line with the slogan describing the Internet standardization approach.

While the network operators implemented technical CPA platforms, content providers were similarly important in their persistent belief and pursuit for its realisation. In this process, aggregators found their role in providing support for smaller content providers where the standard did not suffice. In addition, and perhaps more important, they developed and introduced add-ons and extensions to the platform, enabling new services and service concepts. At least partially resulting from these circumstances, the cost of implementing and operating CPA platforms became marginal for the network operators. The costs and further the risks involved primarily rested with the content providers.

4.3. The case of the Japanese i-mode

Japan enjoys the highest diffusion rate of the Mobile Internet, and the major network operator NTT DoCoMo with its i-mode standard offers a range of services of which the most popular are travelling information, SMS equivalent e-mail services, weather and news, music, games and entertainment (Ishii 2004). NTT provides the content providers a business model where they can charge for monthly subscriptions or for packet transmission. The users' acceptance of NTT DoCoMo's i-mode has been found remarkably high, and i-mode have been identified as a unique success case incomparable to other mobile content services (presented by e.g. MacDonald 2003).

While CPA has had little strategic value for the operators, i-mode is based on NTT exploiting content services for strategic purposes. Where CPA is public and transparent across mobile networks, i-mode is used by NTT for differentiation purposes to attract and retain customers. Because of its strategic importance, NTT has spent significantly more resources on, but also limit the range of services offered to their own users with a 'walled garden' approach. Where the providers of the CPA delegates the

responsibility and the related costs of designing, introducing and administering services to other actors in the market, NTT not only take editorial responsibility, but has also introduced a large bureaucratic organization to administer their service portfolio. Changes and extensions to the i-mode is the responsibility of the network operator, and where innovation is pursued by content providers it is under NTT's central scrutiny. Thus, only if a new service fit the service portfolio and is perceived to have the potential to create revenue, it is accepted. In the case of CPA, anyone can add new components and services on the fly without prior evaluations and approval. Thus, where a 'walled garden' approach leaves the network operators with considerable risk, CPA redistributes most of the risk among a large number of other actors.

When it comes to penetration rate and the number of users, the i-mode story is quite different compared to the CPA. In Japan, the Mobile Internet are in general (including the competing Mobile Internet standards Sky web and EZ-web) available for 36 percent of the total population (Ishii 2004). In the case of CPA, the penetration of CPA is following the mobile phone penetration, which is more or less saturated. While 82 percent of the mobile phone users subscribed to mobile internet services in Japan, only 53 percent of the mobile phone users actually use the services. i-mode users also have to register for the service and have a certain i-mode phone.

5. Discussion: I-mode, CPA and generativity

We will now discuss how infrastructures supporting distribution of information to mobile devices (phones) may enable or constrain innovation and development of new information distribution services.

We will do so by exploring how the various features of the Internet that have been proposed as explanations of its success (in terms of rapid diffusion and development of new services and capabilities). Based on this understanding, our aim is to shed light on how infrastructure design enable and constrain innovative development of new mobile content services. We will first discuss the role of end-to-end architectures, and then the programmability terminals, and finally generativity.

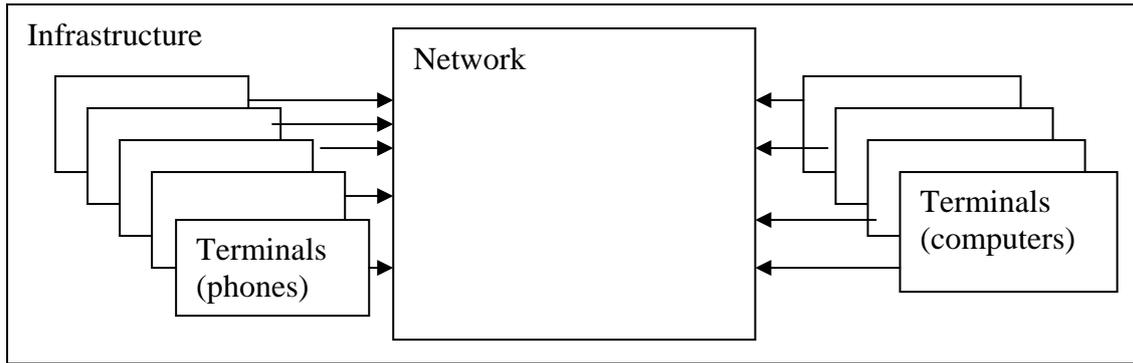


Figure 2. Basic architectural and conceptual framework

Our discussion draws upon the network architecture and concepts illustrated in figure 2. We split the collection of computing and communication technology involved in the requesting and delivering content into network and terminals. The terminals will normally be of two kinds: computers (run by content providers) and (subscribers') handsets. We denote the total collection of technology an infrastructure.

5.1. Mobile Internet and innovations

Early platforms for distributions of content to mobile phones (like MobilInfo and SMSinfo) represented substantial innovations. However, the actual services were not new. From a service perspective it was all about copying just a few very simple services from the Internet domain into the mobile domain. And the success in terms of number of users and usage was modest. With the emergence of CPA and i-mode the speed of innovations, in terms establishing successful services in the mobile domain, increased dramatically. But most services established were still basically copied from the Internet domain. Genuinely new services are still only modestly present.

We see the speed and degree of innovations related to CPA and i-mode as basically on the same level and so also their usage. However, we see the range of services based on CPA as a bit beyond that of i-mode and so also usage compared to the size of the populations and markets in Norway and Japan. But there is a huge difference we would like to point at: the success of the CPA was achieved without any significant investments from the operators' sides, while DoCoMo made huge investments in i-mode.

When comparing innovations related to CPA and i-mode to the Internet there is also a huge difference in terms of speed of innovations and the innovations of brand new services (like file sharing networks and services, community services like you-tube and Facebook, blogging technologies and services, etc.).

5.2. End-to-end architecture

We will now look at the architecture of the various networks and how they relate to the end-to-end principle.

5.1.1 Mobilinfo and SMSinfo

The architecture of Mobilinfo and SMSinfo are definitively *not* end-to-end. The operators asked potential content providers for relevant content. When agreement was reached, the operators received the actual information in one form or another and prepared it for distribution to their subscribers and made it available on their own in-house servers. Content providers were not linked to the platform as such at all, even though real time information like stock exchange rates was received through a computer communication link. So in this case, all information and the technology used to transfer it to users were located in the network and full controlled by the network operators.

5.1.2 The CPA

The CPA platform represented a significant change in architecture compared to the previous proprietary platforms - basically towards an end-to-end architecture where more of the overall functionality was located in the fringes. In this case, what was 'taken out' of the network, was primarily moved to the content providers' computers. But more functionality was also put into the subscribers' handsets.

The CPA network included two main functions. First, it transported SMS messages between content providers' computers and subscribers' handsets. Second, the network included the billing systems so that the transactions could be billed. This network enabled any content provider to establish any kind of

service that could be supported by the SMS transport services and that could be billed by the billing system. That means that the content providers could develop any software that on the bases of received SMS messages generated SMS messages. The content sent to subscribers' handsets may also include WAP-push messages and Java software to run in the handsets that have such a capability.

However, the CPA infrastructure is still not fully based on a pure end-to-end architecture. Substantial – and important – ‘intelligence’ still remains located in the network. This includes first of all the billing systems. Their complexity, and accordingly their ‘resistance’ to change, implies that they constrain innovation. For instance, subscription services cannot be implemented if the billing system is based on so-called MO billing². The billing systems are controlled by the telecom operators. This gives them opportunities to block innovations by others if they want to do so for strategic or other purposes³. But also the billing function may be extended by adding functionality in the terminals and without modifying the (billing system running in the) network. This happened on one occasion when it was discovered that the operators' billing system did not have the capacity to bill all transactions. One of the first Idol competitions on television where the viewers could vote by sending SMS resulted in a very high number of messages sent to the TV channel over the CPA platform within a short period of time. These messages generated more billing transactions than what the billing systems were able to process. Accordingly, the overall billing system had to be changed. In this case, the modifications required to bill all transactions was actually done to the software running in the “end,” i.e. on the TV channel's computers. The modified version of the software of the TV channels stored the information required to bill the transactions and transferred them to the operators in chunks of a size and at a frequency adapted to the billing systems' processing capacity.

5.1.3 i-mode's architecture

² Mo billing means that billing is “mobile originated,” i.e. the billing takes place when the mobile phone sends an SMS. The alternative billing strategy is MT, i.e. “mobile terminated” billing which means that billing takes place when the mobile phone receives the message. See also section 4.2.

³ Examples include content of pornographic or racist nature and services that violate consumer rights

i-mode's architecture is close to identical to that of the CPA platform. It includes the same two basic functions – transfer of requests for content and content plus billing. However, both transfer and billing functionality are different. Requests for content and content transfer are based on the HTTP protocol (requests for the content identified by a url and transferring content as C-HTML). Billing of subscribers is based on monthly subscription rates, not individual transactions. But information about individual transactions is stored in order to calculate the 'sales' and income of each content provider.

5.1.4. Mobile platforms vs. Internet

The CPA and i-mode infrastructures' architecture is rather similar. The Internet does not include billing system and the data transfer protocols (TCP/IP) provide a bit more general data transfer service the CPA and i-mode. In addition the Internet is more symmetrical that the terminals connected are of the same kind (general computers) while CPA and i-mode aim at enabling communication between mobile phones and computers. Even though (more recent models) mobile phones run TCP/IP and may be connected to the Internet, so far neither CPA nor i-mode facilitates arbitrary TCP/IP based communication between mobile phones.

5.1.5 Summary

This discussion shows that there are substantial differences regarding the various mobile content infrastructures' architecture. The architectures of CPA and i-mode are close to the same (related to end-to-end). Their architectures are close to end-to-end. Mobilinfo and SMSinfo, however, are almost as far from end-to-end as you may get. And we find strong evidence support the assumption that these differences in architectures also explain a significant part of the differences when it comes to speed and range of innovations. But the architectures of CPA and i-mode do not conform as closely to the end-to-end principle as the Internet. And we do see this difference as part of the explanation of the differences that still exists between the mobile and the Internet domains when it comes to range and speed of innovations.

5.3. Programmability of terminals

The ultimate programmable device is the modern computer. We will first discuss briefly the features making the computer a powerful programmable device, then the programmability of mobile phone technologies and finally the CPA and i-mode platforms respectively. Early proprietary platforms like Mobilinfo and SMSinfo did not offer any possibilities for programming of terminals at all.

5.2.1 Defining programmability

What first of all makes a device programmable is the possibility of controlling it by means of software. But it is still easier to write software for some devices than others. This depends on the power of the available programming language and the environment supporting the software development activity (i.e. tools supporting the writing, debugging and testing the software). Second, the programmability of a device depends on its basic functionality – usually made available through an operating system. This may include functions like file system, multitasking, support for real time applications, network and communication capabilities, etc. Third, programmability depends on the device's capacity in terms of screen size and quality, keyboard and other input devices available, and the capacity of the processor, storage, and networking technologies. Programmability may also be indirectly affected by other issues like degree of standardization.

As already mentioned, computers are in general the ultimate programmable device. Over the years its programmability has increased as the computers capacity and to some extent also its functionality has increased. For instance, the availability of the Internet has increased the possibilities of developing distributed systems of all kinds. However, recent suggestions for how to solve various security problems (hacking, distribution of viruses, illegal file sharing, etc.), in particular the ideas of 'trusted computing', have made some worried that PC's programmability will be seriously reduced and that they may be turned into mere appliances (Benckler 2006, Zittrain 2007).

5.2.2 Programmability of mobile phones

Originally mobile phones were appliances with no capabilities for being (re-)programmed after they had left the manufacturers. Over the years, however, they have evolved rapidly towards generalized computers. And the most advanced (3G) phones seem to have all capabilities of modern computers – except screen size and keyboard. But the toolkits – like powerful DBMS systems and libraries for writing various kind of software – seem to be limited in number and maturity in comparison. But the most significant difference seems to be the number of ways in which both handset manufacturers and telecom operators in various ways block possibilities for changing programs that are initially put into the handsets and in that way reduced the handsets' programmability for third parties (see for instance Benckler 2006). So, even though the programmability of mobile phones over the years has increased, there is still an enormous gap between mobile phones and PC's in this regard.

5.2.3 CPA vs. i-mode

The basic computational model of i-mode is a transaction model, or more precisely send requests for web-pages and then receiving those pages. And i-mode is linked to the specific i-mode phone which cannot be programmed. So i-mode only allows programming of the terminals by the content providers. And these terminals can most easily be programmed by formatting content in C-HTML or more generally by writing software that by receiving HTTP requests as input produces C-HTML pages.

The computational model of the CPA infrastructure is basically the same as for i-mode: it allows phone users sending SMS messages to request some content service and when receiving an SMS a content provider may produce and return any kind of information that the requesting user's phone is capable of receiving (SMS, ring tones, music, java programs, etc., depending on the model). And the programmability of content providers' terminals is, then, beyond that of i-mode. The CPA platform works for all mobile phones. That means that the power of the programming language of the users' terminals varies. But many of the latest models can run, for instance, java code.

5.2.4 Summary programmability

There is a substantial and significant difference regarding programmability between closed and proprietary infrastructures like Mobilinfo and SMSinfo on the one hand and i-mode and CPA on the other. But the programmability of the CPA infrastructure's terminals also goes significant beyond that of i-mode. But at the same time, there are also significant and substantial differences between CPA and i-mode on the one hand and the Internet on the other. And these differences are crucial parts of the explanation of differences in range and speed of innovations.

5.4. Generativity

We will now turn to the concept of generativity. We will discuss the generativity of the various platforms (PC & Internet, CPA and i-mode) by discussing how they relate to each of the five aspects of generativity included in Zittrain's definition.

5.3.1 Allowing unprompted change by a large, varied, and uncoordinated audience

The extent unprompted change by a large, varied, and uncoordinated audience is allowed is to a large extent determined by an infrastructure's architecture (how close it is to end-to-end) and the programmability of the terminals. Proprietary and closed networks like Mobilinfo and SMSinfo, because of their lack of both end-to-end architecture and programmability of terminals, do not allow or enable any change by anybody except by the network operators themselves. But this issue goes beyond these two technological features. For instance, network operators' policies and the conditions they set for using their network also have impact on to what extent "unprompted change by a large, varied, and uncoordinated audience" is allowed or possible.

CPA clearly appears as an infrastructure where those who control the network do not involve

themselves in the innovative activities. New services are developed and added on the fly, without any evaluation or notification of the network operators. As long as new services do not violate Norwegian laws and a few formal guidelines, anything goes. At the same time, aggregators and other facilitators have taken the role of coordinating the relationship with and technical interfaces to the different operators. They do neither control new services.

NTT DoCoMo has a very different approach to control and coordination. Even if their strategy have changed over the years, the business model and payment system only apply for the official content. Official content is content that have been considered and evaluated as suitable by the content specialists of DoCoMo. As a result, innovation is not only a time consuming and bureaucratic process, but also risky since it is the employees of DoCoMo and not the users to decide its appropriateness. At the same time, DoCoMo argue that this process is highly appreciated by their customers as they can be sure that the content is of high quality.

5.3.2 Capacity of leverage across a range of tasks

Capacity of leverage are largely determined by the terminals programmability, i.e. what kind of useful services that can be developed based on the programming tools available in the terminals. But it also derives from the services offered by the network. And the CPA and i-mode offer one service of particular relevance: billing. The availability of this kind of service enable business models required for the successful development of certain services. And in this respect these two networks contains one form of capacity for leverage not available on the Internet.

Both the CPA and i-mode are infrastructures that enable the transportation and billing of mobile content services. Both infrastructures provide third-party contributors an attractive business model, business models that make it worthwhile producing and offering content. At the same time, direct invoicing of

users or using alternatives such as credit cards is not justified because of the relatively small transactions involved. While CPA billing is on a per transaction basis (also enabling subscription services), i-mode only offers monthly subscriptions. The i-mode approach is therefore less attractive when it comes to offer services which trigger immediate consumption as it requires signing up for a subscription first.

Offering billing services to the content providers, the network operators offers their capability of billing and the content providers can leverage on the operators existing customer relationships. The cost of adding the billing of content services to existing phone bills is marginal compared to the cost of content providers formalising new billing relationships and perform billing of small transactions.

While the business models are attractive, the CPA has been criticized for favouring the network operators. Beyond the economical argument, the content providers are concerned that the cost of using the infrastructure only allows for low-cost services. Providing services which are more costly in terms of investments in technology and production is thus rendered impossible.

While CPA is less attractive revenue share wise, it is more so market wise. While i-mode only is accessible for NTT DoCoMo customers and i-mode users, CPA is available for any Norwegian mobile phone user. While the size of NTT's customer base still make providing content worthwhile, the limited Norwegian market renders provision of services to only a section of the market much less attractive.

The capacity of leverage represented by the billing systems points to another interesting and important conclusion: the limits of the end-to-end architecture. Its existence breaks with the end-to-end principle at the same time as it enables the provision of certain services that would otherwise be almost impossible to develop.

5.3.3 Adaptability to a range of different tasks

Adaptability is also closely linked to programmability. In fact, we have problems in seen that adaptability goes beyond the issue of programmability, at least in relation to the technologies considered in this article.

5.3.4 Ease of mastery

Both CPA and i-mode appear as relatively simple infrastructures for content providers. In the case of the CPA, only a minimal set of functionality is available, meaning that it easy to understand and to use or implement, and it is cheap and easy to provide new services based on it. It is also easy to *change* when new requirements are revealed. Actors such as aggregators and integrators and the services and interfaces they provide also add to the simplicity. Further, the completeness in the sense of a mixture of marketing, use of short numbers, rating classes, etc. also make CPA easier to master. In the case of i-mode, basic knowledge in C-HTML is the primary requirement to provide services.

Network operators are challenged by the complexity and inflexibility of billing systems. Both in the case of the CPA and i-mode, there is no need for the content providers to be involved in and be knowledgeable in the underlying billing systems.

On a general level, the nature of programming language appears as a contradiction between ease of mastery and programmability as a more task specific (and what is often called a more high level) programming language is easier to learn while a more general and abstract one has wider applicability.

Ease of mastery is also influenced by non-technical issues like availability of text books, degree of standardization, number of knowledgeable persons around, etc. For instance, the huge number of text books available around the Internet (from "Internet for dummies" like books to books on narrow and specialized issues) and the widely distributed knowledge of the Internet make the Internet easier to

master for both users and developers than CPA and i-mode.

5.3.5 Accessibility

I-mode appears with limited accessibility both for end-users and for content providers. On the one side, services are only accessible for i-mode customers, on the other, the content for the official sites are filtered by NTT. In this sense, i-mode is not a network that anyone can join – a strategic choice by NTT to make its network more attractive than other networks. In comparison to this ‘walled garden’ approach, CPA appears more like an “open garden” where services are accessible for any user, and any service can be provided.

Where the threshold for providing i-mode services primarily lies in the filtering process of NTT, the substantial costs of connecting to the CPA infrastructures and the revenue share model may hamper content providers launching new services using the CPA.

5.3.6 Summary

The concept of generativity can to a large extent be seen as the sum of end-to-end and programmability of terminals. Accordingly, the analysis of the various infrastructures’ generativity brings us to basically the same conclusion as the discussion of their architectures and programmability. However, the analysis adds a few important results. First of all, we see the concept of generativity to be broader in the sense that it also includes aspects going beyond the purely technological. An example of this is the role of the operators’ (and other actors’) strategies and policies regarding issues like openness. Generativity is also a more holistic than the sum of end-to-end and programmability. This is illustrated by the billing system example. The fact that this service is provided by the network, contradictory to the end-to-end principle, increases the programmability of the terminals in certain ways. Services provided by the network also matters – not only that everything should be in the end-points.

6. Conclusion

We will now conclude our analysis of how to enable or constrain infrastructural innovations in the Mobile Internet domain by means of the concepts of end-to-end architecture, programmability of terminals and generativity.

This analysis illustrates that the change from closed infrastructures like MobillInfo and SMSInfo to more open ones like CPA and i-mode increased the speed and range of innovations substantially. At the same time the differences between CPA and i-mode regarding programmability of terminals, and the billing service provided by the CPA enabling the billing of individual transactions, also contributed to, at least, basically the same speed and range of innovations around CPA as i-mode in spite of the huge differences in investments into the networks made by the owners. But the analysis also points out important differences between the Internet and the existing Mobile Internet regarding in particular technological constraints on innovations. It also points out important ways in which powerful actors' strategies inhibit innovations and how they embed their strategies into the technology and, accordingly, creates technological barriers for innovation. These issues are first of all linked to the programmability of terminals. So we will conclude that the by far most important issue regarding the enabling of more innovations related to the Mobile Internet is to improve the terminals programmability. In this, the concept of generativity with its focus on the grid – that is the mobile phone network and terminals, has shown as fruitful.

References

- Abbate, J. 1999. *Inventing the Internet*. Cambridge, Mass, MIT Press.
- Abbate, J. *The Internet Challenge: Conflict and Compromise in Computer Networking*. In J. Summerton (ed.), *Changing large technical systems*, Westview Press, 1994, pp. 193-210.
- Benkler, Y. 2006. *The Wealth of Networks. How Social Production Transforms Markets and Freedom*. Yale University Press, 2006.
- Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting things out: Classification and Its Consequences*. Cambridge, Massachusetts: The MIT Press.
- Brunsson, Nils, and Bengt Jacobsson. 2002. "The Contemporary Expansion of Standardization." in *A World of*

- Standards, edited by Nils Brunsson, Bengt Jacobsson and Associates. New York: Oxford University Press.
- David, Paul A. 2005. "The Beginnings and Prospective Ending of "End-to-End": An Evolutionary Perspective on the Internet's Architecture." *Industrial Organization 0502012*, Economics Working Paper Archive EconWPA.
- Egyedi, Tineke M. 2002. "Standards Enhance System Flexibility? Mapping Compatibility Strategies Onto Flexibility Objectives." in *EASST 2002 Conference*. University of York, UK.
- Foray, Dominique. 1994. "Users, Standards and the Economics of Coalitions and Committees." *Information Economics and Policy* 6: 269-293.
- Funk, Jeffrey Lee. 2001. *The Mobile Internet: How Japan dialled up and the West disconnected*. Kent, UK: ISI Publications.
- Hanseth, Ole, and Eric Monteiro. 1997. *Understanding Information Infrastructures*. Unpublished Manuscript, <http://heim.ifi.uio.no/~oleha/Publications/bok.pdf>.
- Hanseth, Ole, Eric Monteiro, and Morten Hatling. 1996. "Developing information infrastructure: the tension between standardisation and flexibility." *Science, Technology and Human Values* 11: 407-426.
- Jakobs, Kai. 2000. *User Participation in Standardization Processes - Impact, Problems and Benefits*: Vieweg Publishers.
- Klein, Heinz K., and Michael D. Myers. 1999. "A set of principles for conducting and evaluating interpretive field studies in information systems." *Management Information Systems Quarterly* 23: 67-94.
- Lessig, Lawrence. 2001. *The future of ideas: the fate of the commons in a connected world*. New York: Random House.
- Orlikowski, Wanda J., and Jack J. Baroudi. 1991. "Studying Information Technology in Organizations: Research Approaches and Assumptions." *Information Systems Research* 2: 1-28.
- Saltzer, Jerome. H., David. P. Reed, and David. D. Clark. 1984. "End-to-End Arguments in Systems Design." *ACM Transactions on Computer Systems* 2: 277-288.
- Schmidt, Susanne K., and Raymund Werle. 1998. *Coordinating Technology: Studies in the International Standardization of Telecommunications*. Cambridge, Massachusetts: The MIT Press.
- Star, Susan Leigh. 1999. "The Ethnography of Infrastructure." *American behavioral scientist* 43: 377-391.
- Tassej, Gregory. 1995. "The roles of standards as technology infrastructure." in *Standards, Innovation and Competitiveness*, edited by R. Hawkins, R. Mansell and J. Skea. Aldershot, UK: Edward Elgar.
- . 2000. "Standardization in technology-based markets." *Research Policy* 29: 587-602.
- Vogt, W. Paul. 1999. *Dictionary of Statistics and Methodology*. California: SAGE Publications.
- Vries, Henk J. de. 1999. *Standardization: A Business Approach to the Role of National Standardization Organizations*. Dordrecht: Kluwer Academic Publishers.
- Walsham, Geoff. 1993. *Interpreting information systems in organizations*. Chichester: Wiley.
- . 1995. "Interpretive Case Studies in IS Research: Nature and Method." *European journal of Information Systems* 4: 74-81.
- Werle, Raymund. 2001. "Institutional Aspects of Standardization - Jurisdictional Conflicts and the Choice of Standardization Organizations." *Journal of European Public Policy* 8.
- Williams, Robin, Raluca Bunduchi, Ian Graham, and Neil Pollock. 2004. "Understanding the Evolution of Standards: Alignment and Reconfiguration in Standards Development and Implementation Arenas." in *4S & EASST*. Paris.
- Yin, Robert K. 1994. *Case study research: designs and methods*: Sage Publications Inc.
- Zittrain, J. 2006. *The Generative Internet*. *Harvard Law Review*, Vol. 119, p. 1974-2040.