

Metropolis and SOA Governance

Part 1: Towards the Agile Metropolis

By Richard Veryard and Philip Boxer

Introduction

From Time to Space

The defining technology of the past thousand years was the clock; and the prevailing technological imperative has been to save time, to make things go faster or better than before. See *Box 1*.



"I was your slave, now you are mine: I am Time."
(Incredible String Band)

Modern mechanical clocks with falling weights were invented around a thousand years ago. The invention is often attributed to Pope Sylvester II, who was an accomplished mathematician and scientist before becoming pope. Under the Rule of St Benedict, mediaeval monasteries used the clock to regulate labour and prayer. Lewis Mumford traces the origins of the Industrial Revolution to the Rule of St Benedict, and to the domination of the clock. Charlie Chaplin's film *Modern Times* shows (in exaggerated form) the relentless power of the clock over the production line worker. Business process engineering in the late twentieth century was focused on reducing cycle time, eliminating waiting. The key slogan: *Just-in-time*.

Box 1. Chronarchy – The Power of the Clock.

Of course, it is too early to say what the defining technology of the next thousand years will turn out to be. But there are already some signs of a shift from an emphasis on time towards an emphasis on difference. The Internet, for example, creates new kinds of difference in the relationships between people and organizations; business organizations operate as differentiated nodes or clusters within complex networked ecosystems; social and institutional cohesion is mapped against the coordinates of complex abstract dimensions of difference.

Our understanding of complexity itself rests upon the recognition that once we go beyond a certain threshold of difference in the behaviours of systems, it becomes impossible to predict their composite behaviour. Thus in the service economy,

we expect service-oriented systems to emerge that are increasingly large and increasingly complex, but that are also capable of behaviours that are increasingly differentiated. As we shall see, this is one of the key challenges of Service Oriented Architecture (SOA).

The city or metropolis is another large complex system, which many of us encounter in our everyday lives, and which we typically experience as differentiated in its behaviour. Our familiarity with cities makes the metropolis a good starting place for working out the best approach to building and governing such large complex systems. Furthermore, many of the key issues for the design and governance of large complex service-oriented systems arise also in the field of urban design, where they have been debated for decades (although without reaching consensus).

As it happens, the latest heavyweight contribution to the debate on physical architecture and urban design comes from Christopher Alexander, whose long-awaited work on the *Nature of Order* has finally been published. Alexander has had a profound influence on software engineering for many years – his early work on the *Synthesis of Form* influenced the structured methods of Yourdon and deMarco, while his middle work on *Patterns* was taken up enthusiastically by large numbers of software engineers, especially in the OO world.

From Cities to SOA

In the twentieth century, two of the best writers on the nature of cities were Lewis Mumford and Jane

Jacobs. Mumford thought a well-ordered city needed central planning and infrastructure, while Jacobs took a more anarchist position.

Here are some of the issues raised in the city debate.¹

Adaptability In nineteenth century England, Manchester was highly adapted to the cotton industry, but failed to adapt to later waves of industrialization. Meanwhile, Birmingham was far more adaptable, and this enabled it to accommodate a series of industrial innovations.

Complexity A city contains a vast quantity of social and commercial interaction. A living city allows for many different levels of such interaction, and for meaningful clusters and subclusters to emerge, which form an abstract hierarchy.

Governance City planning requires orchestration of developments large and small, balancing local initiative and autonomy against global coherence.

There are some strong parallels between town planning and service-oriented architecture, which make it reasonable to translate ideas and experience from urban design into the SOA domain.

The distribution of design. Detailed design decisions are taken within different organizations, each following its own agenda (e.g. commercial or political goals).

Box 2. SOA Governance Questions from Pat Helland

<i>[In contrast to metropolitan city governance ...] IT governance is not so mature.</i>	<i>Enterprises might learn a lot by looking at how cities manage the difficult process of resource allocation.</i>
Who makes the tough choices in IT? Is it the CEO, the CIO, the business unit leaders, techies, or perhaps committees?	What proposals are projected to pay for themselves?
Are priorities established based on cost, flexibility or asset utilization?	What is the timeframe and risk analysis around these projections?
What is success and how is it measured?	What in your organization is sacred?
Are we seeking cost reductions, business process transparency, or competitive advantage?	What resources remain after funding these efforts?
	What balance of short-term, long-term, and speculative investments are right within the specific corporate culture?
<i>These problems are common for metropolitan and IT environments.</i>	

¹ For a useful summary, see Coward & Salingaros.

The constancy of change. Elements of the whole are constantly being redesigned and reconfigured, and new elements are constantly being added. Structures must evolve in robust ways.

The need for progressive improvement. Each design increment should not only make local improvements, but should have a positive effect on the whole.

The recursive nature of the architecture. Similar design tasks must be carried out at different scales (levels of granularity).

Helland on Metropolis

In a recent article in the Microsoft Architects Journal², Pat Helland offers an extended analysis comparing the planning and management of IT with the planning and management of cities. He argues that IT governance has a lot to learn from city governance (*Box 2*), and raises some interesting parallels between urban design and SOA (*Box 3*).

Cities	IT shops
Factories or Buildings	Applications
Transportation	Communication
Manufactured Goods	Structured Data
Manufactured Assemblies	Virtual Enterprises
Retail and Distribution	Business Process
Urban Infrastructure	IT Infrastructure
City Government	IT Governance

Box 3. Key Parallels between Cities and IT from Pat Helland

Helland's article makes the following argument.

1. Progress requires standardization. (According to Helland, people didn't even wash properly until they had standard clothing.)
2. Standardization is associated with commoditization.
3. Standardization requires concentration of power (and if this involves pathological distortions of socioeconomic relations, so be it).
4. Infrastructure requires central investment. (Since we may regard infrastructure as an act of local standardization, it follows that it must involve concentration of power.)
5. Central investment preserves the "sacred".

² Pat Helland, Metropolis. Microsoft Architects Journal2, April 2004. Available at <http://msdn.microsoft.com/architecture/journal>

Let's look at the steps of Helland's argument in detail.

Standardization

Progress should involve enriching people's lives, while many acts of standardization impoverish them instead. Not everyone is willing to regard Wal-Mart as the epitome of progress. Living systems are heterogeneous.

Helland's Utopia appears to be a relatively homogeneous one. Citizens even all smell the same. The exclusion of antisocial odours is engineered though standard interchangeable clothing (although doubtless soap and clean water contribute something as well).

The homogeneity of personal hygiene can of course be disguised by the fact that we can all choose from thousands of different scents. Poor people have plain soap, and rich people have fancy perfume. But this consumerist illusion of choice (enthusiastically promoted by Wal-Mart) does not represent genuine diversity, let alone heterogeneity.

Modern production methods allow for mass customization. This involves a separation of production into two layers – a homogeneous layer of mass production and a heterogeneous layer of customization.

In urban design, the standardized stuff is what goes underneath the pavement – standard utilities such as water and power, for example. The human stuff goes above the pavement. It is a question for city governance to decide what should/may go under the pavement.

The articulation of a complex system into two layers (one homogeneous, one hetero) is an architectural question. Obviously there are companies with a significant commercial stake in this question. It is therefore worth being suspicious of an articulation that is presented as given by some historical tradition or technical imperative³. Apparently pure technocratic architectural judgements often conceal a commercial agenda. One of the functions of governance is to maintain a "level playing field" between different commercial agendas. For this reason, regulators often aspire to intervene at the architectural level.

It is conceivable that something could look homogeneous from the supply side and heterogeneous from the demand side, or vice versa, so the boundary is itself relative to some supply/demand formulation. Technology is constantly creating homogeneous/ heterogeneous splits – for example, Voice over IP technology

³ Practical example – phone companies would like to regard the location of mobile phone masts as mere infrastructure, to be decided on technical grounds and requiring no public consultation. However, people are becoming concerned about the radiation from these masts, especially near homes and schools, and this politicizes the location of masts.

creates a split between devices that care whether a bitstream represent voice or data (and therefore regard the traffic as heterogeneous), and devices that don't care (and therefore regard the traffic as homogeneous).

One plausible basis for articulation of layers is the differential rate of change. It may appear to make sense to standardize and regulate the slow-moving layer, and allow greater diversity in the fast-moving layer. But an ecological perspective informs us that the slow-moving dominates the fast-moving.

This entails a new role for architecture:

- To maintain appropriate stratification of layers and coupling between elements within and across layers, and
- To operate at a higher level of abstraction, implementing evidence-based design policies.

Existing approaches to defining architectures may not work very well even for the homogenous bits. They certainly don't work for the heterogeneous bits, and they also don't help with defining the boundary between the homogeneous layer(s) and the heterogeneous layer(s). A consequence of our argument is that the boundary between homogeneous and heterogeneous (as illustrated above) is a proper focus of attention for architecture. We shall come back to what this means in practice.

Commoditization

We cannot avoid the commoditization of our lives – but we should be wary of its dangers.⁴

Fortunately, we no longer have to put up with one-size-fits-all software. Situated software – software designed in and for a particular social situation or context – resists the traditional software engineering pressure towards generalization, and apparently disregards the economics of scale/scope. Instead it works solely within a collaborative socio-technical system (the "community"); the conditions for the success of the software (including meaning and trust) being co-created by the members of the community.

One of the earliest forms of situated software was the spreadsheet. Power users built themselves complicated structures using Visicalc or Lotus 123 or Excel. These were essentially non-transferable artifacts with many hidden assumptions, but they served a useful purpose within a given context. This illustrates the fact that situated software is assisted by the existence of tools and platforms that provide generalized support for situated software.

⁴ A possible reconciliation of commoditization with human values is provided by Albert Borgmann.

The overwhelming success of the spreadsheet was due to the fact that it performed a useful function, while leaving the user free to create context-specific meanings. But the spreadsheet was also limited by the fact that these meanings were private and undocumented, and attempts to turn spreadsheets into shared artifacts typically failed.

This is where the software factory comes in. There is a great opportunity here to produce software tools capable of supporting a rich diversity of end-user demand. A Domain Specific Language (DSL) can be a way of bridging and holding open the gap between the general/public and the context-specific/private, and maintain the dynamic interaction between supply and demand. This dynamic needs to be driven by the way the end-user defines the relationship between domains and their business as a whole.

The service economy is a complex ecosystem. Service-oriented solutions are essentially systems of systems, and their composition should be mindful of complex systems theory. To maintain requisite variety (and therefore survival of the fittest) in such an ecosystem, we need diversity at all levels of abstraction.

Concentration of Power

The Wal-Mart economic system is unsustainable. It destroys the fabric of small shops on which a rich urban life depends.

Cities are somewhat paradoxical. On the one hand, a city is itself already a concentration of human activity. But the concentration processes are unstable, and can result in highly dysfunctional urban forms.

Historically, cities have had walls, to keep out unwanted visitors. Elsewhere, Helland has advocated the fortress model of computing. But here, he seems to be envisioning a continuous metropolitan fabric, where one city fades into the next (as Manchester merges into Salford).

Central Investment

Helland's plea for central investment (the Lewis Mumford position) provides a justification for corporate central planning and investment in IT. Many large and small organizations attempt to impose central IT planning. However, in many organizations this is a losing battle. The actual situation of the IT industry emerges from millions of small procurement decisions, and is closer to the idea of anarchic procurement (the Jane Jacobs position).

⁵ For a version of this cycle as it applies to health care see "Triply Articulated Modelling of the Anticipatory Enterprise" by P Boxer and Prof B. Cohen, Proceedings of the International Conference on Complex Systems, Boston 2004.

Salingaros draws on Alexander to describe how the Mumford and Jacobs dispute can be resolved – but by adopting an approach that goes beyond simply attempting to reconcile the top-down with the bottom-up.

Preservation of the "Sacred"

A modern version of the sacred may be found in Borgmann's notion of focal things and practices. The role of urban/system governance would be to create/preserve a space in which these focal things and practices can be developed and honoured.

In an agile demand economy, the source of the sacred is demand. This contrasts with a supply-side logic based on a presumption of *symmetric* demand, in which markets are defined to reflect the supplier, so that formations of demand are symmetrical to the formations of supply.

In talking about Virtual Enterprises, Helland writes: "you have to consider the context in which the part will be used. Is weight or ruggedness the principle concern?" Helland argues that standards are the key to enabling component providers to leverage the cost of optimization across a broader market, and this can be understood as supply-side logic. However, this issue goes beyond standardization (here extended to business component models, enabling them to be encapsulated as component capabilities and orchestrated as a part of larger assemblies of processes) and opens up the *granularity* of component capabilities with respect to each other i.e. what is the repertoire of alternative component capabilities available. This needs to be understood from the demand-side as well.

In considering Business Process, this argument is extended by analogy to the need for standardization and interchangeability in data and operations: "people cheerfully accept standard stuff and customization is rare and expensive. But business process is still largely hand-crafted. There are poor standards..." And so the argument goes for standardization providing a basis of extending a supply-side logic deeper into the provision of services, with business process becoming the driving force that dictates the shape and form of applications "as surely as Wal-Mart drives the standards for many, many manufactured goods."

This doesn't meet the challenge of the above retail cycle⁵. This cycle describes the emergence of a new form of supply-demand relationship (destination), which expands to become a new form of offering alongside others (comparison), before it becomes commoditized (cost), and ultimately embedded into the user's context (custom). From here, the ground

is prepared for a new cycle (the transitional cusp), and so on. This cycle is a dynamic process, in which the supply-side is constantly learning new forms of supply in response to a demand which is always evolving – and never fully satisfied. *Asymmetric* demand describes the demand in its particular context-of-use, and this something-always-left-to-be-desired is a structural deficit which is always there driving the development of markets. Commoditization is only the supply-side part of the story, the real issue being the way the dynamics of the formation of demand itself has to be supported.

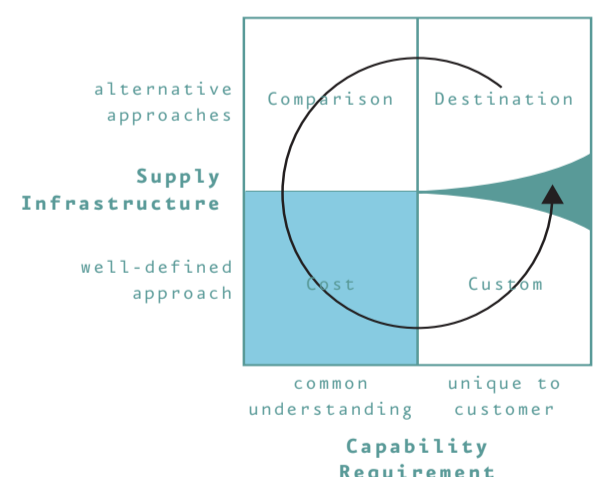
Nature of Order

For several decades, Christopher Alexander has been exploring alternatives to conventional architectural practice. His latest work, the Nature of Order, was published last year.

The Nature of Order (in its complete four volumes) has implications for the involvement of people in the design of buildings and in the detailed ways in which this involvement is likely to be successful or unsuccessful. It has implications for the flow of money. It has implications for the handling of architectural detail and for the successful integration of structural engineering into the framework of design. ... It affects virtually every part of the profession we now know as architecture, and it indicates necessity for change, in almost all of them. There is no question, that under the impact of this theory, architecture will be deeply changed and it will be changed for the better.

Box 4. The Nature of Order. Christopher Alexander's Manifesto

Figure 1. Capability Requirement



According to Alexander, large complex systems cannot be produced by a conventional design process – either top-down or bottom-up. Instead, they emerge from an extended and collaborative (evolutionary) process. Order and coherence comes from the rules that govern this evolutionary process.

This evolutionary process can be broken down into discrete steps, which may either preserve structure and wholeness, or destroy it. Structure and wholeness is articulated as a recursive system of centers.

Services can easily be considered as centers of value⁶. Services can be composed into composite services, with service orchestration (hopefully) yielding coherence between recursive levels.

A service-oriented enterprise can then be understood as a continuous corporate web of services. The architectural properties of this enterprise depend on the numerous collaborating processes that bring about its composition. If these are appropriately structure-preserving, then the enterprise can become both increasingly differentiated and increasingly integrated, without loss of coherence.

Alexander is highly critical of the conventional governance over city planning and urban design, and highly critical of the inflexible and inhuman results of directed composition. His ideas on design and order are, we believe, consistent with the needs of collaborative composition, as outlined in this paper.

Structural Implications

Business Geometry

A business or value chain is composed in a geometric structure. In SOA, we design a business or value chain as a network of services. This is a powerful geometrical pattern. But there may be many possible network geometries capable of satisfying a given business requirement, all of which count as satisfying the principles of SOA. For example: hub/spoke or peer/peer.

A key characteristic of SOA is **stratification**. A business process is composed of services from a set of lower-level services, presented as a platform. A good example of a business platform is the set of retail services offered by Amazon and eBay. Other service providers have built further retail/logistical services on top of the Amazon/eBay platforms.

Each platform is in turn built upon even lower services. At the lower levels, there may be collections of IT-based services, known as

Enterprise Service Bus (ESB). There may also be socio-technical service platforms, such as call centres. Some of these lower layers of the stack may appear to be purely technical services; however a more complete picture should reveal the existence of an IT organization maintaining the platform, in other words it too is a socio-technical platform which includes its administrators and programmers.

Thus we have a stratified geometry, in which a person tackling a problem at a given level is presented with a collection of available services, formed into a virtual platform. This can be thought of as a **business stack**, with one platform stacked on top of another. And while the SOA principles may provide some geometrical guidance, and mandate certain geometrical patterns, there is still a design job to determine the form of geometry most appropriate to supporting the service demanded. This design job may be easy when the requirement is trivial, but gets harder as the complexity increases.

In many situations, the demand side has more variation than a human designer (or design team) can accommodate. (We characterize this as an asymmetry of demand, which calls for a process of asymmetric design.) Under these circumstances, we need to go still further and start thinking about variable geometry solutions, where the geometry itself can be adapted on-demand.

For example, in the past we have assumed that granularity has to be fixed at design time. But we can conceive of a web service platform that detects patterns of demand-side composition, defines new component services automatically, describing and publishing these new services in real time, and notifies likely users of the new service, complete with an incentive to switch to new ways of orchestrating them in support of demand-side composition. We can conceive of a such a web service platform analysing the message content of a certain service, and producing a substitute service with a smaller footprint that would satisfy most of the uses in a more elegant way.

We use the term **value landscape** to refer to the distribution of cost, benefit and know-how across a complex market ecosystem, such as the insurance industry, for a given level of risk. Technology (including SOA) influences business geometry, because it not only affects transaction costs, but also the way know-how can be leveraged in relation to demand. The shape of the value landscape changes (has already started to change) as the result of B2B, B2C, P2P and BPO. Companies that once occupied safe market positions may find their

commercial advantage slipping away, or they may find themselves cut off from their former customers or supply chains.

Let's suppose an insurance company has the following strategic aims:

1. Profitability, short-term viability. To deliver the maximum service value as cost-effectively as possible, using available input services and technologies as efficiently as possible, with minimum costs/risks of change.
2. Adaptability, medium-term viability. To understand and respond to changing demand for insurance services, and to trends in cost and risk, both internally and across the industry. To develop and deploy new services to exploit new business opportunities and avoid emerging business threats.
3. Survival, long-term viability. Making sure the core business proposition remains valid, and doesn't get eroded by more agile players. Taking strategic action in relation to structural changes in the insurance industry.

If we are doing business geometry for an insurance company, we need to think about the insurance industry as an evolving ecosystem. We need an AS-IS model of the present ecosystem (largely based on pre-SOA technologies) and a TO-BE model of an emerging ecosystem (based on the effects of SOA). We can expect the pre-SOA ecosystem to evolve into some form of post-SOA ecosystem, although we may not have much idea which of the possible changes is going to happen first. To satisfy all three strategic aims, an insurance company needs to exploit the pre-SOA ecosystem, and prepare for the post-SOA ecosystem.

Note that this situation may force the insurance company to implement a variable geometry across its business stack, both in the organizational platform and in the underlying IT platform. Otherwise, it will either have to operate sub-optimally for an extended period, or incur significant organization costs and IT costs every time the industry takes another step towards SOA. Variable geometry involves a dynamic collaboration (**collaborative composition**) between an efficient supply side and a variable (asymmetric) demand side, as shown in *Figure 2*.

Asymmetry of Demand

Asymmetry means that the forms of demand are increasingly specific to the context in which they arise.

The first asymmetry involves separating out technology from the supply of specific products.

⁶ See article by Murphy.

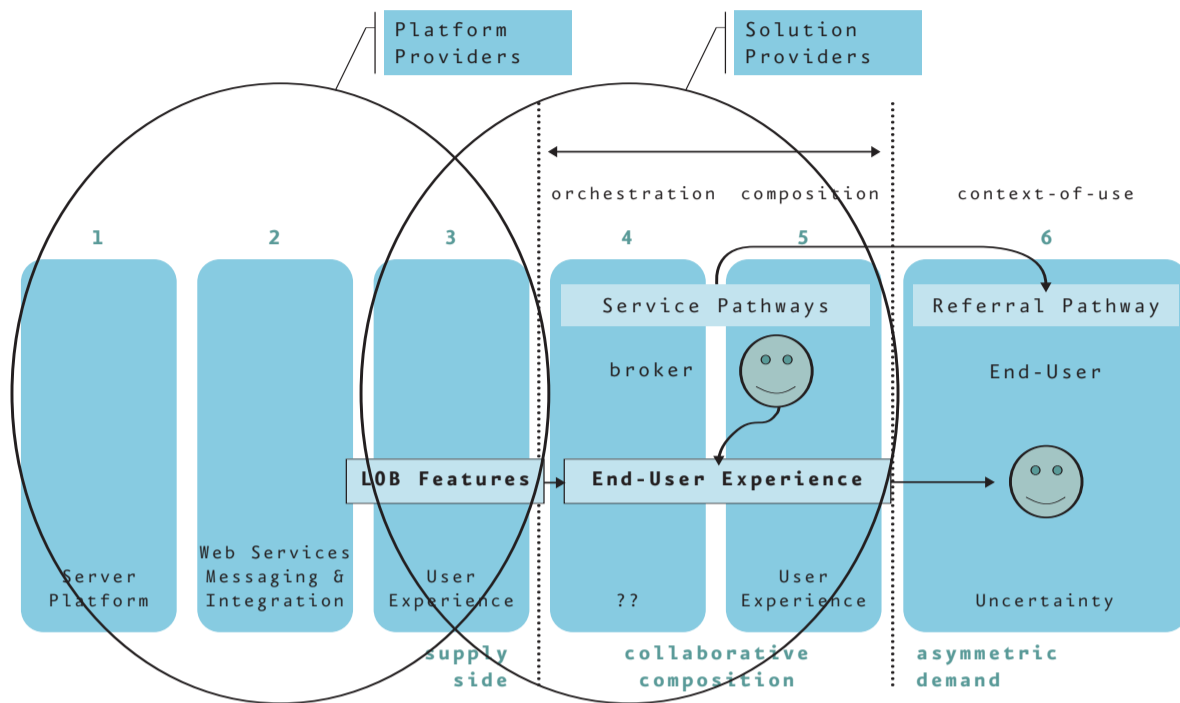


Figure 2. Collaborative Composition in the Insurance Value Chain (Based on Microsoft IVC)⁷

This requires modelling of possible behaviours that can be supported (so Microsoft or car manufacturing has to modularize itself in support of families of technology use).

The second asymmetry requires separating out business models that can organize supply from the solutions that are on offer. This requires modelling of the possible forms of business geometry (so rail maintenance or retail services have to use a franchise model to allow the variation in business organization to accommodate the variety of ways in which the service needs to be implemented).

And the third asymmetry requires separating out the different contexts of use. This requires modelling of the possible forms of demand (so that financial or care services are having to take up the way the through-time wealth/conditions are managed in a way that responds to different forms of context-of-use).

These asymmetries are summarized in Figure 3, and it is worth considering what happens if they are ignored. In the first asymmetry, this means defining the product by the technology. This is typical of the early stages in the emergence of

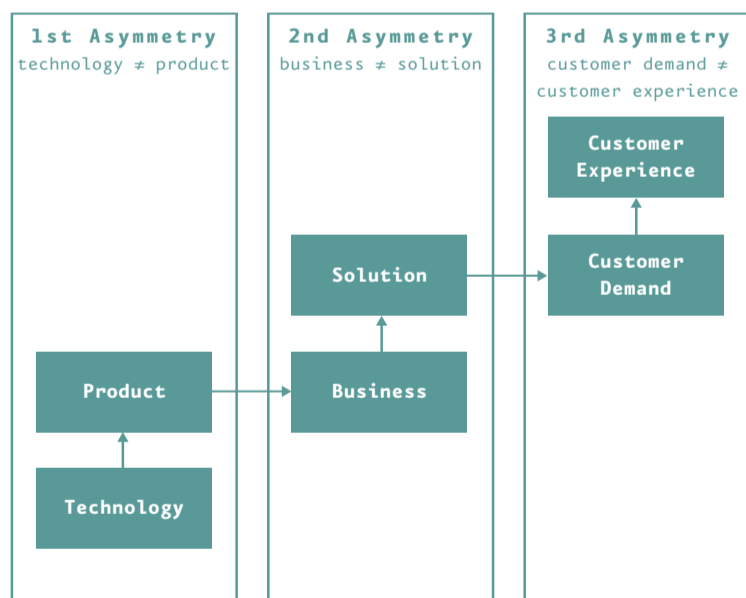


Figure 3. Three Asymmetries of Demand

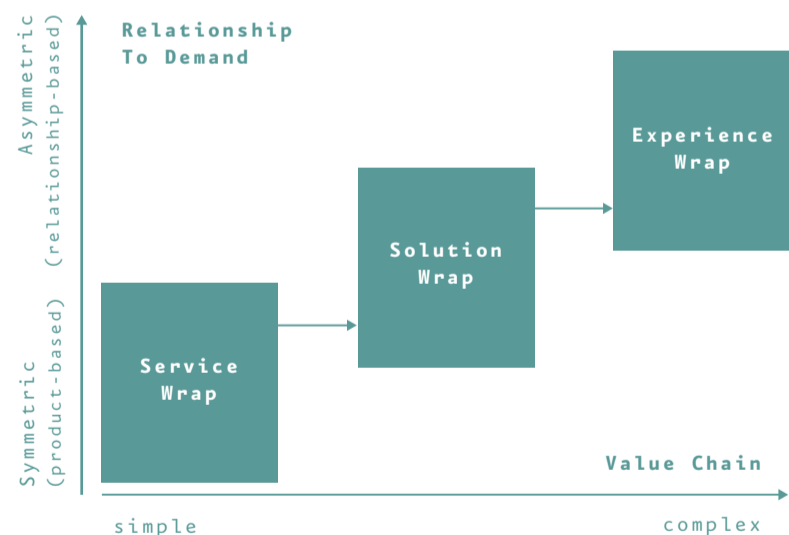


Figure 4. Supply-Demand Roadmap

new technologies. (Do you remember how we used to have to use mobile phones?). In the second asymmetry, this means defining the solution for the customer by the way the business is organized. (Do you remember how large businesses used to relate to their customers before CRM?). And in the third asymmetry, the solution to the problem presented by the customer is assumed to be what the customer actually needs. (Have you ever received a prescription from the doctor that turns out only to treat the symptom?). We see the major competitive impact of SOA being that it changes what the supplier can afford to ignore from the customer's perspective.

To become better at capturing asymmetric forms of demand, an organization needs leadership that will enable it to do two things.

1. **Take power to the edge of the organization** – the people at the edge of the organization with the relationship to the asymmetric demand must be able to organize the business model they need to capture that demand.
2. **Develop an agile infrastructure** – providing business services that can be orchestrated and composed at the edge in response to the particular forms of demand they are targeting. This then allows the supply-side of a business to extract economies of scale or scope when providing support across multiple business models.

Box 5 shows what this means in practice⁸.

⁷ Picture taken from report on Service-Based Business in Insurance, CBDI Forum November 2004.

⁸ This example is derived from work done for PASA. http://www.pasa.nhs.uk/orthotics/Orthotic_Pathfinder_Report_July_2004.doc

Healthcare example.

Asymmetry in the variety of conditions that people have, and the ways in which they reveal themselves in people's lives over time is ever increasing. Meanwhile hospitals and clinics are having to become ever more efficient in how they administer particular care pathways.

We see this challenge particularly in relation to conditions that are chronic. Thus patients don't produce conditions that fit the treatments that are available, while organizing acute care systems around the treatment of chronic conditions becomes exorbitantly expensive. Patients' conditions are relentlessly asymmetric from the point of view of the medical specialisms trying to care for them. For example, acute surgery can often be traced to an earlier failure to provide timely prophylactic treatment.

Strategic Response

Getting power to the edge in healthcare means the funding following the patient, and the doctors having the ability to craft a treatment plan that is particular to the patient's condition.

This creates a double challenge for healthcare provision. Not only must it increase the flexibility with which its component services may be made available to patients, but also the doctors have to have a much greater involvement in formulating whole-life strategies of healthcare provision that can be tailor-made to a patient's condition over time, and for the delivery of which they can be held accountable. Where this can be done, the total cost of care is reduced.

A healthcare purchasing and supplies agency (PASA) was responsible for the supplies of equipment to a clinical service. They were concerned about the side-effects of minimising the costs of these supplies – reduced investment in the industry and a vicious circle of decline in the quality of the clinical service itself. They decided that the conventional symmetric design wasn't working for them; in an attempt to improve the quality of the clinical service itself, they decided to consider developing an approach that addressed the asymmetric nature of the clinic's demands. They conducted an initial pilot study to establish the feasibility of an asymmetric design process.

What PASA accepted was that they needed to address the demand-side of the clinic, and establish how best to satisfy its needs. Within this context-of-use, they could then address the question of the costs of supply.

A national process was set up by the modernisation agency. This process ran six 'pathfinder' projects, each of which aimed to establish how to effect change in each context. It was the national project's task to secure long-term support and funding for the change process in the light of the learning and results of the pathfinders. This ultimately involved modeling the regional and national impact of the changes on NHS and Social Services budgets.

From the point of view of the supply to the clinics, the demand-side modelling was of the referral pathways and the services offered by each clinic in response to the demands arising from those referral pathways. The supply-side modelling was of the organization of the clinic itself, together with its use of suppliers, in order to establish how the one was aligned to the other. Where this "cut" came between the supply-side and demand-side was a function of who the client was, and what was their change agenda. Nevertheless, in examining the referral pathways and the particular ways in which they themselves had been "colonized" by suppliers, further questions were raised about the organization of primary care itself. These questions were left, however, to be taken up by a different client system at a later time – and which would have to address the interests of the Strategic Health Authorities.

The key challenge was to give the clinicians "design control" over how the clinics operated (i.e. power at the edge). Fundamental to this was to grasp on the supply-side the chronic multi-episode nature of the conditions being treated by the clinic, and on the demand-side the processes of delegation and/or teaming of clinical responsibility for patients' conditions. The clinicians lacked the means of defining the different characteristics of the former, and managing the complexities of the latter. Furthermore, without the means of doing these things, there was no practical way of holding the clinicians accountable for the clinical performance of their clinic. The solution was to build a reporting platform that could support the doing of these things.

This modeling involved defining the referral pathways and their actual characteristics. From this emerged the requirement to change the way the clinic was relating to demand.

This modeling involved defining the service propositions and clinical business models needed. In the former case, this meant establishing episode protocols for different conditions, and in the latter, changing the workflow processes between the clinic and front and back office processes in support.

No realignment of the supply infrastructures was attempted, considerable gains being available simply through the way the alignment of suppliers to the demand was managed. The reporting platform built provided the means of achieving this. The platform allowed the clinic to define its own treatment protocols in relation to its own definitions of referred conditions. The back end of the platform was able to lift data out of the NHS environment on patients, appointments etc.

Box 5. Healthcare Example

Supply-Demand Roadmap

Figure 4 shows a typical roadmap for the service-based business as it responds to the competitive impact of SOA. The business tackles each form of asymmetry in turn:

1. It uses a “service wrap” to decouple the product from the technology. This includes defining a different object model for the demand side, separating “raw” data from what we might call “cooked” data.
2. It uses a “solution wrap” to decouple the solution from the business. This includes defining different rules for the demand side, separating the business logic from the orchestration of different solutions.
3. It uses an “experience wrap” to decouple the ongoing customer experience from the particular solutions bought by the customer at any one time. This includes new forms of process modeling to understand the customer experience of solutions within their particular contexts of use.

The point about this progression is that it confronts the supplying business with the need to manage increasing complexity (and concurrency) in the value chain relates to the customer. Hence the competitive significance of SOA.

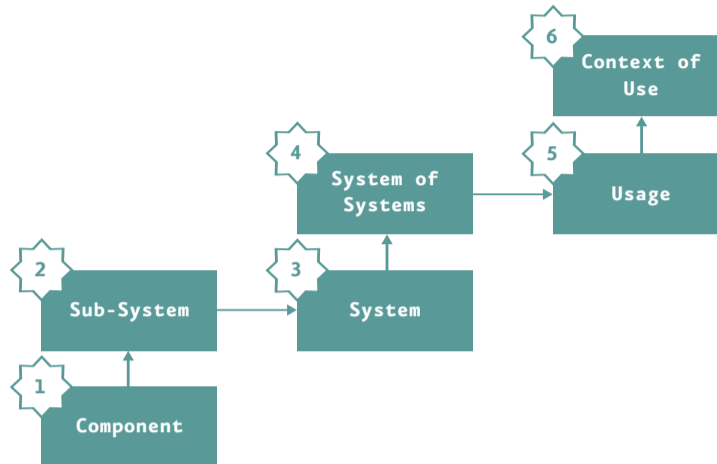


Figure 5. The Stratification of Models

Breaking the Three Symmetries

With the first symmetry, the business only needs one model, because demand can be inferred from supply (or so it seems at the time!). When this symmetry is broken, two models are needed: one to manage the technology, and the other to manage the business. How many times has an investor of venture capital had to teach this to a start-up business?

If the full potential of a business proposition is to be realised, the second symmetry must also be broken (this is something that every marketing MBA learns!). Now there are three models: one

to manage the technology, one to manage the business, and one to manage the market solution. Thus in the railway example in Box 6, you need to separate out how you deliver the service because you cannot derive the output-side service levels for the train operating companies purely from the input-side service requirements of the rail infrastructure (unless, that is, you are prepared to service to the highest level of quality under all circumstances).

But what effect does breaking the third symmetry have? What CRM tells us is that the ‘market solution’ is also a customer relationship. But in breaking the third symmetry we have to learn that each customer incorporates the solution into a context-of-use that is different. Thus in healthcare, it is no longer just a question of delivering particular treatments cost-efficiently, but also of delivering the right treatment at the right time within the context of an evolving patient condition – like old age for example!

Thus in the pharmaceutical case, you cannot derive the output-side knowledge (what the GPs and pharmacists need to know about the patient’s condition) from the input-side knowledge. You need at least two models because you cannot infer the effects of the treatments purely from your knowledge of interactions between the prescription and the situation as it presents itself when you prescribe – it requires a model of the dynamic effects of the prescription within its context-of-use.

Overall, then, the breaking of symmetries requires six different levels of model to separated out (see Figure 5). Even though each of these strata may contain many different levels within it, the agility of the infrastructure will depend on how much variability they will allow in the geometry of how they can be composed.

Box 6. Implications of Asymmetry – the Three Dilemmas

Symmetry	Assumptions	Implications of Asymmetry
Technology = Product	The first symmetrical assumption is that the first three challenges are all lined up. Thus these three challenges collapse into a single dimension defined by the technology.	With SOA, we are increasingly confronted with businesses that are nothing more than a technology platform for other businesses. (From Microsoft itself downwards.) So the simple alignment doesn’t work. Instead, they are pushed into some form of stratification.
Business = Solution	The business rules and procedures espoused by the supplier match the ways the services are to be used.	Rail maintenance is supposed to provide reliable and safe railways. In the UK, Network Rail (formerly Railtrack) take input services from engineering companies and turns them into output services for train operating companies. It has proved to be extremely difficult to align the input requirements against the output requirements.
Customer Demand = Customer Experience	The banking fantasy of straight-through-processing is based on symmetry and shared values all along the value chain.	Contrast this with the situation in the pharmaceutical industry, where a drug company’s set of relationships with GPs and pharmacists are of quite a different nature to those of the GPs and pharmacists with their patients, despite the tendency of the drug companies to assume that it should be otherwise.

Role of the IT Architect in SOA

What, then, does all this require of the “IT architect” in support of these new forms of agility? Our main argument is that there is no single right level of granularity. There is a supply-side vs demand-side dilemma which needs to be addressed by an organic method (c.f. Alexander) in which the level of stratification at which a change is made must always also address its impact on the levels above and below. As a result, a trade-off has to be made between two types of pressure on any given layer of a stratification, one standardizing and the other requiring greater diversity. This places a different kind of demand on the role for architecture.

We may think of the role of architecture as the making of certain kinds of (structural) judgement⁹. We see architecture as something that emerges from a collaborative process which may (but doesn't have to) involve people with the word “architect” in their job title. We may refer to these people as professional architects, although in systems and software engineering there is no professional body granting any official status to this job title. These judgements belong in an asymmetric design regimen.

We can then separate the products and processes of architectural activity from the professional responsibilities of suitably qualified Architects. The structural properties of a city depend on a complex collaborative process, in which the professional judgements of architects are pitted against a broad range of commercial and political interests. Professional architects rarely dominate the design of cities, although they may sometimes exert some influence.

And we may note a similar situation within IT. Although there are numerous people in the IT industry calling themselves “architect”, the structural properties of large IT systems often depend largely on decisions taken elsewhere. For example, the degree of coupling between two modules may have a significant impact on the structural cohesion and flexibility of a large system, but this may be determined by fairly low-level development choices that are not even visible to the software architects.

In many organizations, software architects can be outmaneuvered and marginalized by coalitions of developers and users. And as outsourcing increases, the position of the software architect may become yet weaker – especially in those instances where contractual specifications focus on the functional requirements and under specify the structural properties; and where there are

inadequate mechanisms for the architects to verify the structural properties of the delivered software. (For example, hidden coupling that compromises the intended flexibility of a software artifact.)

So the challenge for software architects is to remain relevant to an SOA world, to a world of distributed production of distributed services, by paying attention to the real structural issues emerging in this “on-demand” world. Otherwise, they will be unable to contribute anything of value to the design and management of on-demand systems of systems. This leads to a need for forms of analysis that support an asymmetric design regimen, and can enable explicit consideration to be given to implicit choices being made concerning geometry.

Conclusions and Next Steps

We have seen some vendors recognising the supply-side issues of reconciling multiple web services (viz IBM Rational), while other platform vendors are creating the conditions for an explosion in the numbers of (behavioural) domains needing to be brought into relation with each other (viz Microsoft). In both cases, the service-oriented business is configured as a continuous fabric of services – “the corporate web”. But this can never be achieved in one large ambitious project. It has to be achieved progressively through a continuous stream of small and medium projects.

In the collaborative planning approach, order and coherence emerges from distributed activity, with no central design authority. Each unit of procurement, development or maintenance activity has to be regarded as a project, with project outputs being constituted as services. Thus each project contributes something positive to the emerging corporate web of services. So what form of governance is needed to maintain architectural order?

SOA Governance is required to ensure that each project satisfies the global demands of the corporate web, and to ensure that there is a well-balanced mix of projects – different types as well as different scales (large, medium and small).

What our discussion in Part I has outlined is the limitations of a supply-side approach to SOA governance – directed composition is limited in its capacity to respond to the full heterogeneity of demand. It leaves too great a value deficit in relation to demand which is increasingly heterogeneous, asymmetric, and spatially as well as temporally differentiated. So we need to take governance to the edge of the organization, acknowledging that we are engaged in processes of asymmetric design, and

prepare to meet the twenty-first century on its own terms. In Part 2 of this article, we shall open up the question of what taking ‘governance to the edge’ means for the design of SOA infrastructures as well as for the relation to demand.

References

- Christopher Alexander, *The Nature of Order*. 4 Volumes. The Center for Environmental Structure, 2002-2004.
- Albert Borgmann, *Technology and the Character of Contemporary Life*. Chicago, 1984.
- Philip Boxer & Bernie Cohen, *Triple Articulation*. BRL Working Paper, revised 2004.
- Andrew Coward & Nikos Salingaros, *The Information Architecture of Cities*, *Journal of Information Science*, Volume 30 No. 2 (2004), pp. 107-118. Reprinted in Salingaros 2005.
- Jack Greenfield & Keith Short, *Software Factories* (Wiley, 2004).
- Pat Helland, *Metropolis*. *Microsoft Architects Journal* 2, April 2004. Available at <http://msdn.microsoft.com/architecture/journal>
- Jane Jacobs, *The Death and Life of Great American Cities* (Vintage Books, New York, 1961).
- Lewis Mumford, *The Culture of Cities*. (Secker & Warburg, 1938).
- Richard C. Murphy, *Centers: The Architecture of Services and the Phenomenon of Life*. FTP Online March 2004. <http://www.ftponline.com/special/soa/murphy/>
- Nikos Salingaros, *Principles of Urban Structure*. Delft University Press, Delft, Holland, 2005, in press. <http://www.math.utsa.edu/~salingar/urbanstructure.html>
- Richard Veryard, *Component-Based Business* (Springer 2001).
- Richard Veryard & David Sprott, *The Service Based Business* (CBDI Journal, 2003). *SOA Governance and Business Driven SOA* (CBDI Journal, 2004).

⁹ We use judgement here as defined by Vickers, to include appreciation and value judgement, as well as action judgement.

Thanks to Bernie Cohen, Pat Helland and Nikos Salingaros for comments on earlier drafts.