Service Grids: The Missing Link in Web Services

By John Hagel, III and John Seely Brown

Service grids are a critical architectural component required to realize the business potential of Web services. Without the emergence and evolution of robust service grids, Web services technologies will remain relatively marginal, especially in terms of supporting inter-enterprise connections. As our other working paper "Break on Through to the Other Side" suggests, much of the earliest potential impact of Web services is concentrated at the edge of the enterprise. Without service grids, this potential will largely be unrealized. In spite of the importance of service grids, all the attention of technology vendors and pundits remains focused on two other layers of the Web services architecture – the foundation standards and protocols and the application services ultimately enabled by these standards and protocols. Service grids truly have become a critical missing link in Web services.

WHAT ARE SERVICE GRIDS?

Service grids need to be understood from two perspectives: technology architecture and business ecology. Their critical and distinctive role in a distributed service technology architecture creates an opportunity for a diverse array of businesses to form and create significant economic value.

The concept of service grids in technology architecture

Let's start with the architectural view. When most technologists talk about Web services, they start appropriately by highlighting the key role of a new set of standards in defining this new generation of technology. Riding upon the existing Internet standards of TCP/IP and http, Web services technology begins with XML as a foundation standard establishing a format for exchanging data and application functionality. This foundation standard in turn has spawned a whole series of derivative standards and protocols including, most importantly, SOAP (establishing a common format for addressing messages), WSDL (creating a standard way of describing the functionality of a Web service and instructions on how to access it) and UDDI (providing a uniform way of registering Web services so they can be searched for and discovered by others). These standards and protocols represent essential building blocks for Web services technology. The fact that they have been so quickly and universally adopted by technology vendors creates a basis for optimism.

Once technologists have finished describing the core standards and protocols underlying Web services technology, they generally move quickly to the application level. They paint futuristic visions of enterprises using Web services standards to dynamically compose new applications to address the specific needs of the business at any point in time. Generally, these visions suggest that many micro-services will be knit together to create broader application services tailored to specific enterprises or even specific transactions.

Occasionally one might hear a reference to a broker helping to compose these applications, but most of the discussion remains tightly focused on providers of application services and users of application services. In part, the close relationship in the eyes of many technologists between Web services technology and peer-to-peer architectures explains this bias.

Web services and peer-to-peer architectures. Peer-to-peer architectures are essentially network-based architectures that enable smart devices (PC's, servers, PDA's, etc.) at the end-points to communicate with each other as equals without requiring servers to facilitate the connection. In many respects, this was one of the early insights of Arpanet, the predecessor to the Internet that sought to connect computers with each other as peers. Peer to peer architectures stand in contrast to client-server architectures where there is a clear and unchanging hierarchy – clients (such as PC's and smart terminals) must communicate with each other by going through a server. Client devices always play the role of client and server devices always play the role of servers. In the peer-to-peer world, a client in one transaction may just as readily play the role of server in the next transaction. Roles are redefined to meet the need of the moment.

This vision is seductive. In a world where all forms of hierarchy are suspect, the peer-to-peer vision is profoundly liberating. Eliminate the middleman and free all devices to connect with each other at will. Certainly the great success of the Internet as a peer-to-peer network lent credibility to the champions of peer-to-peer architectures. The enormous popularity of Napster and ICQ, often touted as peer-to-peer applications, fueled the enthusiasm of peer-to-peer champions. Many of these same champions have also become advocates of Web services technology. It is therefore not surprising that the technology discussions remain so silent on anyone standing in the middle between the provider of a Web service and the user of a Web service.

Yet, even if we look closely at peer-to-peer architectures, we find that they are not quite so pure as they at first might seem. Napster for example uses a central server as a registry for the music resources. Users must go through this server to establish a connection with the music resource provider. Groove, another popular example of a peer-to-peer application, uses a message queuing server independent of either node to ensure reliability of message delivery. The purity of peer to peer rapidly dissolves in the face of the pragmatic need to offer more functionality to support communication among the nodes. Yes, the Internet is a great example of a peer-to-peer architecture, but have you ever tried to consistently establish reliable connections on the Internet? Pure peer to peer comes at a significant price – either one must sacrifice functionality or accept much higher complexity at the nodes to provide functionality that centralized servers might otherwise provide.

The same is true of Web services technology. It is no accident that many of the most fervent champions of Web services technology keep falling back on relatively simple application examples like the delivery of a currency converter or stock quotes to a cell phone. In these examples, robustness of connection is not a big concern. Peer to peer connections with their limited functionality work just fine. So what if a connection can't be established or maintained or someone hacks into the connection to access the data being delivered? Not a big deal. Now, think about a lean supply chain where factories start to close down if a connection goes down for more than a few hours. How about a payment processing application where millions of dollars are being transferred in individual transactions from one account to another? Robustness of connections now becomes a major concern. Pure peer-to-peer just won't do.

We are starting to see a broader movement away from purist peer-to-peer paradigms, driven in part by the inexorable improvement in performance of key technology components – computing power, bandwidth and storage. These improvements in technology performance make it possible, for example, to use message queues cost-effectively to implement asynchronous transactions. Queues in turn help to solve the monitoring problem, making connections accountable for performance and diagnosable so that shortfalls in performance can be quickly addressed. Message queue managers sit in the middle, facilitating connections among end points. From a purist peer-to-peer perspective, this is anathema. Perhaps we need to start be more forgiving of such heresies and embrace hybrid peer-to-peer models where devices at the end-points retain the ability to connect with each other as peers, but aided and abetted by specialized (and optional) services helping to make connections more robust. Service grids play this latter role in a distributed services architectures.

The importance of managed services. A distributed services architecture will be required before Web services technology can be broadly deployed to support mission critical applications within and across enterprises. Service grids constitute a key component of this distributed services architecture as the scope of the architecture expands beyond the boundaries of the firm to encompass a broad range of business partners. These service grids are analogous to the electrical power grid in that they provide a set of enabling utilities and services to support more robust connections between providers and users of Web services.

This enabling functionality is distinct from application functionality that is directly useful to end-users. Rather than applications like inventory management systems or order entry systems, we are focusing on supporting, or enabling, these applications with functionality like security, routing of messages across applications or data transformation so that one application can access data from another application. In many respects, you might visualize the enabling functionality provided by service grids as the equivalent of middleware for enterprise applications, only in this case, delivered as a set of managed services, rather than installed in the computers communicating at either end of the connection.

This notion of managed services is critical to the concept of a service grid. Despite their importance, standards alone cannot do the job. These standards need to be harnessed in the form of a managed service in order for their full value to be realized. Take the example of WSDL, the standard for the representation of what a Web service can do. The provider of a Web service can make all kinds of representations regarding what that Web service can do. But who will in fact verify that those representations are accurate? Who will provide a third party auditing function to monitor the performance of Web service in action and confirm that the serve performs as represented? For WSDL and UDDI to provide trusted services, some managed service will be required to administer reputation systems. These systems would operate on an eBay type model where users can indicate whether the Web services performed as represented.

Or, take the case of various security standards. Standards are important, but what matters even more is the ability to compare and align security policies among the participants in a business process. Some kind of managed service would be required to compare and align security policies, as well as deal with the inevitable exceptions that the automatic rules can't handle.

Within the enterprise, these managed enabling services may be developed and administered by the IT department. But, what happens when Web services are provided and accessed across multiple enterprises? In some cases, it may make sense for a dominant, larger enterprise to take on the role of managed enabling service provider. In other cases, robust connections simply won't be established in the absence of a diverse set of third party managed enabling service providers. Service grids perform many of the same functions as a CIO performs within the enterprise – e.g., defining and enforcing policies, resolving exceptions when they arise and refining policies accordingly, delivering services to support applications, recruiting and developing appropriate technical expertise. Service grids thus become particularly useful in coordination of Web services deployment across enterprises where there is no single CIO to deploy expertise and to handle inconsistencies and inadequacies in implementations.

Service grids in many respects are devices to focus human intervention. They automate the more mundane administrative activities while freeing up skilled personnel to accelerate learning by handling a broad range of exception conditions. Service grids also become gathering points for the formation of communities of participants concentrating on addressing current business needs and anticipating the next wave of business needs. In the end, service grids provide a way to blend technology with human judgment.

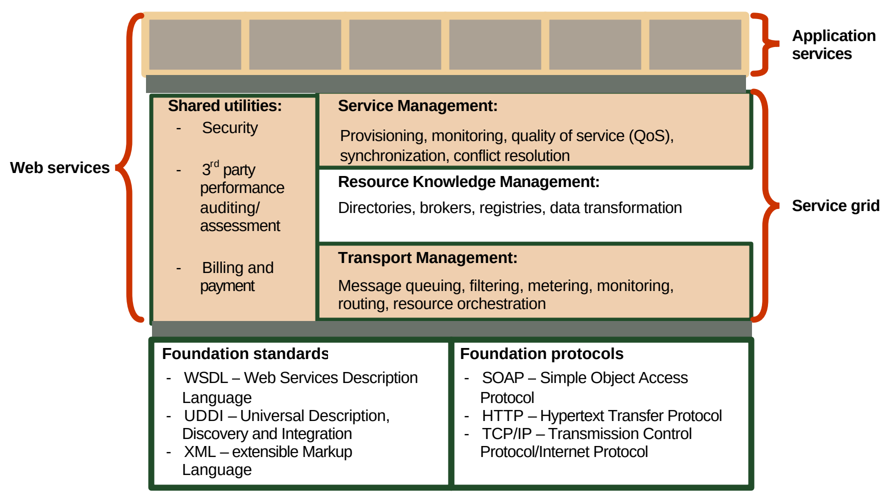

Figure 1 – Web Services Architecture

Utilities within service grids. As Figure 1 indicates, service grids provide four broad categories of managed services:

1. **Shared utilities** provide services that support not only the application services but also the other utilities within the service grid. There are three types of shared utilities. Security utilities provide services like authentication, authorization, and accounting. Performance auditing and assessment utilities provide assurance to users of Web services that they will obtain agreed-upon levels of performance and will be compensated for damages if performance falls below these levels. Billing and payment utilities aggregate charges for the use of Web services and ensure prompt and full payment.

2. **Transport management utilities** include messaging services to facilitate reliable and flexible communication among application services as well as

orchestration utilities that help companies assemble sets of application services from different providers.

3. **Resource knowledge management utilities** include Web services directories, brokers, and common registries that describe available application services and determine correct ways of interacting with them. They also include specialized services for converting data from one format to another.

4. **Service management utilities** ensure reliable provisioning of Web services. They also manage sessions and monitor performance to ascertain conformance to service quality specifications and service-level agreements.

Service grids as optional network overlays. Service grids and their component utilities are entirely optional components of a distributed services architecture. They operate as optional overlays on existing networks – providers and users of Web services can choose whether or not they wish to avail themselves of the enabling functionality offered by service grids. In this way, a distributed service architecture preserves both the simplicity and ubiquity of the underlying Internet platform, but offers participants additional enabling functionality if they need it. Service grids do not imply a return to the feature-heavy, "smart" networks of the voice telecom world. They honor the value of simple, "dumb" networks while also honoring the need of participants to access higher levels of enabling of functionality through optional overlays.

Service grids as federations of service utilities. From an architectural perspective, service grids are not hard-wired bundles of utilities. Instead, they are loosely coupled federations of utilities tailored to the needs of specific application environments. Like the domain name system (DNS) that provides a critical foundation for the Internet, these federations of utilities will be layered with abilities to escalate to resolve inconsistencies until, if necessary, a single, ultimate arbiter is reached. Think of this layered federation as equivalent to the system of government in the United States with the U.S. Supreme Court as the ultimate arbiter of disputes.

In some cases, a service grid may not contain all the service utility groups identified above. In other cases, additional, more specialized, utilities may be included to address the unique needs of a particular application environment. In fact, service grids are likely to be federated with each other, in much the same way that electrical power grids are federated to provide access to a broader range of production capacity.

Willingness to federate utilities within a service grid in part will depend upon the performance needs of the applications and processes being supported. For certain kinds of trading settlements millisecond response times and a high degree of predictability may impose performance constraints beyond the capabilities of a federated approach. In these situations, utilities will need to be more tightly integrated to deliver the required performance. Most business environments are not as demanding in terms of performance. As a result, a broad range of environments may support and in fact seek a federated approach to increase flexibility.

Federation of course increases concern about trust-worthiness. The more loosely coupled the service grid becomes, the more critical it is to develop reputation mechanisms like those offered by EBay to determine the trust-worthiness of utilities.

The concept of service grids in a business ecology

Service grids are not only components of a distributed services architecture. They are also business formations representing an array of service businesses or utilities owning and deploying specific enabling services. These service businesses may either be specialized independent businesses or revenue centers within larger enterprises offering their specialized enabling services to other enterprises.

Service integrators and aggregators. Service grid business formations also include a broad range of businesses operating as service integrators or aggregators, knitting together the enabling services into complex bundles of services targeting the needs of specific customers or segments of the market. Service integrators will design and source the services for highly specialized service grids tailored to meet the needs of individual large enterprises. Think of them as the Web services equivalent of systems integrators today. Or, they might be a large enterprise itself with deep internal capabilities to integrate the various utilities required for the Web services platforms used by that enterprise. Service aggregators will have a somewhat broader focus, designing and sourcing the services for service grids targeted to meet the needs of specific market segments like insurance companies or managers of supply chains.

Enabling service utilities. These integrators or aggregators might own some of the specialized enabling service utilities, but in most cases they are likely to be reselling services provided by third party utilities. Once again, these integrators or aggregators may be specialized independent businesses or, as indicated above, the integrators may be larger, more diversified enterprises offering this specialized capability to participants within their own enterprise or within other enterprises. In some cases, industry consortia might own these specialized integrators or aggregators to deliver service grid capability tailored to the unique needs of the consortia participants.

Other business opportunities. No doubt, professional service firms will also focus on helping participants in distributed services architectures to determine their service grid needs and to evaluate different service grid offerings. In addition, there are likely to be specialized hosting businesses that will operate the enabling services developed by specialized utilities or even entire service grids assembled by integrators or aggregators.

WHY ARE SERVICE GRIDS SO IMPORTANT?

Service grids are a pre-requisite for the economic impact of Web services

Mission critical functionality. Unless businesses can access the specialized capabilities offered by service grids, they are unlikely to adopt Web services technology to support mission critical business processes. If the technology is not used to support mission critical business processes, it will be relegated to support marginal activities within the enterprise and generate only modest economic returns. The only way for Web services technology to have significant business impact is to make the technology relevant to mission critical activities.

Why are service grids so important for mission critical business activity? Because connections across application services supporting mission critical business activities must be available at all times, perform reliably, maintain appropriate security and deliver results quickly. Any shortfall in connection performance along these dimensions could result in significant disruptions to the operations of the enterprise. When these connections are established using Internet technology, concerns immediately arise. The Internet is low cost and ubiquitous, but it is also has performance issues in terms of reliability, security and speed. Managed enabling services delivered by service grids are necessary to design and deliver additional functionality around the connections established across Web services.

In many cases, the functionality delivered by enabling service utilities in the service grid will have a strong technological component – e.g., ensuring secure connections, monitoring and addressing performance shortfalls in connections and routing messages based on pre-defined business rules. In other cases, the mission critical functionality delivered by service grids will have little, if any, technological component. For example, specialized utilities in a service grid may provide new forms of risk management capability through performance bonds and warranting services designed to address concerns about relying on Web services. Of course, such utilities would rely on other more technological utilities to monitor and audit performance of Web services.

Enabling revenue generation from Web services. Addressing mission critical performance needs is a pre-requisite for any real economic value to be generated by the technology. Farther down the road, as adoption becomes more widespread, service grids play additional roles in enhancing the economic impact of Web services technology. For example, the technological vision of the broad-based knitting together of application or enabling services from different enterprises tends to overlook one small item: who is paying whom for the use of these application services? Behind that question are a host of others. Who will monitor and measure the use of application services? Who will compare performance of the application services against relevant service level agreements to determine whether threshold performance levels were achieved? Who will offer the billing capability to the providers of application services? Who will collect payments? Once again, service grid utilities can play a key role in harnessing the economic potential of Web services technology by making it feasible for providers of Web services to generate revenue. These specialized capabilities are necessary for any robust ecology of specialized businesses to emerge in the Web services arena as providers of Web services. In their absence, we will undoubtedly see Web services deployed within the enterprise and perhaps within private trading networks of closely related business partners who are not focused on revenue generation from the Web services provided.

Easing the task of finding appropriate Web services. Even farther down the road, as a broader array of Web services and providers emerge, service grids play a third critical role in harnessing the economic potential of the technology. Let's go back to that technological vision of the broad-based knitting together of application or enabling services from different enterprises. Another small item these visions tend to underplay: who will manage the directory services required to discover and evaluate this exploding array of Web services options? As indicated above, UDDI and WSDL are useful standards to support the development of these directory services, but who will actually develop and manage these services? Who will work to refine the descriptions of Web services based on a growing set of user experiences and learning about the kind of information that users of Web services will need to make the appropriate choices for their application needs? Many factors will determine appropriateness, including such features as scalability, reliability and performance within specific technology environments. Without these specialized directory services utilities, deployment of specific Web services will remain confined to very limited business arenas and the ability to generate significant revenue will be substantially throttled back.

Service grids play a vital role in accelerating the adoption of Web services technology

The specialized enabling services offered by service grids help to remove key obstacles to adoption. Some of these obstacles have to do with availability of key technology and limited skills to implement the technology. Other obstacles involve the continuing evolution of standards necessary to support a broader range of technology deployment. Yet other obstacles involve perhaps the biggest challenge of all: establishing shared meaning, especially across enterprises, so that business activities can be effectively coordinated.

Reducing complexity at the edges. By taking much of the functionality that would otherwise need to be installed in the devices connecting with each other and instead delivering it as a set of managed enabling services, service grids make it much simpler for the end-points to connect. By shifting much of the complexity from the edges of a connection to the center of the connection, service grids play a key role in reducing barriers to adoption and increasing incentives for the provision and use of Web services.

This benefit is particularly stark in connections across enterprises, especially when many of the companies involved are small businesses with very modest technology platforms on their premises and even more modest technological skills among their staff. They simply could not afford to participate if it required expensive new technology and deeper skills to establish connections. But these benefits are even powerful within the enterprise, especially one that is geographically dispersed with a broad number of smaller facilities with limited resources.

Service grids take the escalating complexity confronted by any one end-point as it seeks to establish many-to-many connections across a broad range of other end-points and move that complexity into shared utilities that have deeper capability to handle it. Service grids take this "n-squared" growth in complexity and reduce it to 2n – all the end-point need master is how to connect its technology with the technology interface offered by the service grid. The service grid takes care of the rest.

Bootstrapping with the use of existing technologies. If users had to wait for a whole new set of enabling services to be developed from the ground up using Web services technology, the adoption process would lengthen considerably. It might even stall completely in the face of the "chicken and egg" dilemma. Enabling services would be slow to develop because few Web services connections have been established. Few Web services connections would be established because only a limited number of enabling services are available.

Using Web services technology to "expose" technology that is already available and operational in enterprises can help to more quickly assemble service grids.

Citibank offers an early example of this. Citibank recently used Web services technology to "expose" its payment-processing engine that had previously only been accessible within the bank. The capabilities of this payment-processing engine are now available to other companies through a Web services interface. In effect, Citibank has created a key enabling service utility for service grids to access and provide to a broader range of business users.

This example highlights a key point. Service grids may be essential to enhancing the value of Web services technology, but the specialized utilities they encompass do not themselves have to be built exclusively using Web services technology. In fact, in the early stages, many of these specialized utilities are likely to have been built with previous generations of technology. Web services technology can be used at the outset simply to implement the necessary interfaces required to deliver the capabilities of the utilities as an enabling service. Over time, undoubtedly Web services technology will be extended into the core of many enabling services to enhance their flexibility, but this is not a pre-requisite for deployment of service grids. Service grids therefore can be implemented much more quickly, piggybacking on more conventional technology that is already in place.

Mediating among competing standards and policies. The core standards of Web service technology are already in place – XML, SOAP, WSDL and UDDI. An entire additional set of standards are in much earlier stage of development, designed to address such important business issues as security, business process management and management of long lived transactions. At this stage, candidates for these standards are proliferating and there is little near-term prospect of convergence around a single set of standards in any of these categories. Even the core standards themselves are evolving and a broad range of companies are proposing new enhancements to make these core standards more robust. In light of all this effort, there is a growing sense that reconciling these evolving standards will be critical to broad adoption of the technology.

If we had to wait for these various standards to mature and converge, adoption of Web services technology would be significantly delayed. Once again, service grids can help to remove this obstacle and accelerate adoption. Specialized utilities in service grids can help to reduce the near-term complexity created by proliferating standards. Rather than requiring each end-point in a Web services connection to understand and translate the various Web service standards in use at every other end-point, specialized utilities can perform that function and make it easier for end-points to connect, regardless of which variants of the standards are in use at any point in time.

The challenges created by evolving Web services standards are only one part of a broader mediation challenge. Take the example of application frameworks. Technologists realize that, with the advent of application servers, we have seen growing reliance on application frameworks that can be very helpful in terms of providing developers and users access to shared enabling services. These shared enabling services parallel many of the enabling services offered by service grids – e.g., security services, data conversion services and message queuing services. But there's one problem. These services are typically only available to applications developed within that application framework. Microsoft has its own application framework and a variety of vendors in the Java world have developed their own application frameworks. None of these frameworks talk very well to each other. Service grids in many respects can create a meta-framework, helping participants operating within diverse application frameworks to communicate more effectively with each other by mediating the differences.

Conflicting business policies create another kind of mediation problem. For example, we talk a lot about security technologies, but we spend a lot less time on security policies. Usually (but not always) these security policies have been standardized within an individual enterprise. But watch out as soon as you try to establish automated connections across enterprises. Each enterprise typically has evolved its own security policies and, especially in light of recent events, these security policies are continuing to evolve in significant ways. Specialized utilities within the service grid can be very helpful in registering, comparing and mediating differing security policies to ensure that automated connections are tailored to the diverse security policies of the participants.

Helping to develop shared meaning. As most executives are quickly coming to understand, XML is not a panacea for enabling communication across diverse enterprises. It plays a critical role by establishing a common format, or grammar, to support communication. But grammar alone is not a foundation for communication. Nouns and verbs must have meaning and that meaning must be shared among all participants. For example, an XML tag may be helpful in designating where product size information is presented, but do all parties agree on the meaning of product size? Is size represented in inches or centimeters? Does it list height before length and width or vice versa? Without agreement on specific terms, serious misunderstandings can occur.

This shared meaning is being slowly generated, in large part through the collaborative efforts of small groups of business partners who see the value of shared meaning in helping to automate connections across applications and enterprises. These efforts are typically isolated, largely unaware of other, similar efforts and relying upon ad hoc methods and tools to fashion this shared meaning.

Such efforts could be significantly enhanced and accelerated through the availability of specialized enabling services focused on helping business partners to create, disseminate and refine shared meaning. These specialized utilities could develop sophisticated methodologies and tools by surveying and understanding experiences to date in building shared meaning through the use of XML formats. These utilities would realize that shared meaning cannot be specified fully at the outset – it evolves as misunderstandings and exceptions arise, indicating where shared meaning does not yet exist. This suggests a continuing role for these specialized utilities, monitoring exceptions and working with participants to refine meaning over time.

Service grids offer compelling economics in terms of amortizing development cost

Enabling services cost money to develop, maintain and enhance. If every company had to create its own enabling services to support connections across application services, it would incur significant costs. Much of the same capability would be replicated across many enterprises. Companies that could not afford to develop their own enabling services would become increasingly isolated and disadvantaged or they would become increasingly dependent on the good will of larger business partners to access the necessary functionality.

Service grids facilitate a higher degree of specialization and help companies to more effectively amortize the cost of developing, maintaining and enhancing enabling services. Rather than have every company create the same enabling service, service grids encourage the growth of specialized utilities that can offer their enabling services across a broad range of other companies. Since development costs can be more broadly and rapidly amortized, strong incentives exist for the formation of new specialized utilities.

Service grids enhance access to world-class capability

By encouraging specialization and helping specialized utilities reach a broader set of users, service grids also accelerate learning and performance improvement. Each specialized utility is able to focus on providing only the enabling services in its distinctive area of competence without diverting its resources into other business areas. Since the enabling services are its only business, it is typically able to attract and retain the best talent in that area. In contrast, when more diversified companies seek to offer these enabling services as a secondary business, the companies often treat the talent required to develop and improve these services as second-class citizens.

Specialization and reach provides an even more powerful engine for learning and performance improvement. These enabling service utilities typically operate in a much broader range of environments and see a much broader range of events than “captive” service utilities that support the operations of one company or segment of the market. As a result, they gain more experience more quickly and can learn from that experience to refine their own service offerings to become even more effective. This learning effect can be particularly critical in areas such as intrusion detection in security services or shared meaning and data conversion services where the ability to see many situations or context provides a distinctive advantage.

IS ANYONE ADDRESSING THE CHALLENGE TODAY?

Service grids are still at a very early stage of development. Nevertheless, we are already seeing significant initiatives leading to the emergence of early generations of service grids and specialized service grid utilities. This is one area in the Web services space that represents significant opportunity both for existing enterprises and entrepreneurial start-ups.

Enabling service aggregators and integrators. Perhaps the most significant early activity is in the area of enabling service aggregators and integrators. The early candidates for these roles are typically companies that began with a different focus, but had developed many of the enabling service capabilities required in a service grid and saw an opportunity to embrace Web services technology to enhance the value to their customers.

Much of the early activity for example is concentrated in the first generation of electronic marketplace and business-to-business startups. For example, a consortium of companies in the high tech industry led by IBM formed E2open to develop and operate a platform for collaboration with business partners in areas like product development, supply chain management and procurement. E2open has developed its Global Collaboration Network as a technology platform providing a robust set of enabling services like directory services, security services, data conversion services and transaction monitoring services. E2open supports the exchange of XML documents using the Web services based RosettaNet Implementation Framework and it is one of the first companies to introduce a commercial implementation of UDDI directory services. In general, it is aggressively implementing Web services technology interfaces to facilitate access to its enabling services by customers.

Comergent is another early contender for a service grid aggregator role. The company targets the complex coordination challenges facing companies seeking to collaborate with sales channel partners. Creating a platform independent service, Comergent offers a variety of enabling services relative to XML documents including transformation and messaging in addition to its broader suite of partner relationship management applications.

Both E2open and Comergent initially developed these enabling services as part of a broader application service. Over time, they may decouple the enabling services into a distinct service grid offering to support a broader range of customers who may not wish to participate in the application services. They may also find that they can enhance the capabilities of their early enabling services offerings by plugging in more specialized enabling services developed by third parties. The fact that they already offer broader service grid capability increases the opportunity for more specialized enabling service utilities to emerge and address existing gaps in the service grid.

Other candidates for enabling service aggregators include existing industry collaboration hubs that began operation with earlier generations of technology. Think of Sabre in the travel industry and SWIFT in the financial services arena. Both of these companies are implementing Web services interfaces to their earlier service offerings to reduce integration costs for their customers. They too may find it attractive to enhance their current service offerings by aggregating more specialized enabling services developed by third parties.

Another fertile ground for the emergence of enabling service aggregators appears to be the grid computing space. This can get confusing. Grid computing is different from the service grid concept. Grid computing emerged as a way to help users more effectively share computing resource capacity. This is particularly valuable in computationally intensive arenas like biogenetics, astrophysics and weather forecasting. Over time, it has expanded its focus to include approaches to help users share a variety of technology-related resources, including data and applications. To address these opportunities, early grid computing initiatives have adopted distributed services architectures and are leveraging Web services technology to help design and operate these architectures. Companies like Avaki Corporation and initiatives like the Legion project at the University of Virginia are starting to offer early forms of service grids aggregating a variety of specialized enabling services like fine-grained security services, policy-based configuration and administration of resources and automatic failure detection and recovery.

Yet other candidates for the role of enabling service aggregators include the few survivors of the first generation of application service providers (ASP's) and broader providers of network services. One of the challenges faced by the first generation of ASP's was that they had to develop from scratch an entire set of enabling services to support their application offerings. These application offerings frequently had substantial limitations (discussed in more detail in John Hagel's forthcoming book Out of the Box), but the enabling services represent an important foundation for a potential service grid offering.

Providers of network services like AT&T and Verizon could also play the role of service grid aggregator by leveraging their customer relationships and expertise in operating scalable and reliable network services. These companies typically have fewer enabling services of their own to offer, but could become attractive packagers and resellers of more specialized enabling services, especially from newer companies without a lot of brand recognition or operating track record.

Let's also not forget the role of the large enterprise as a spawning ground of enabling service aggregators. Many larger companies like General Motors, Dell and Merrill Lynch are rapidly deploying Web services technology within their own enterprises and with business partners. In many cases, they are creating enabling service capability internally and drawing upon more specialized third party utilities to knit together their own service grids. Over time, they may see an opportunity to offer other companies access to their service grids. This will likely happen first for their close business partners, but they may eventually seek to generate additional revenue by spinning out these service grid aggregation platforms as distinct businesses available to a broader range of customers.

So far, we have focused on service grid aggregation – creating service grids designed to meet the needs of broad segments of the market. Early initiatives are also beginning to be launched to provide service grid integration – knitting together enabling services to meet the needs of an individual customer. Of all the traditional systems integrators, IBM Global Services has shown the most interest in playing the role of service grid integrator. Other firms like Accenture are also candidates for this role, although they remain much more focused on conventional systems integration businesses to date.

This leaves an opening for smaller, more focused integration firms to enter the fray. NerveWire is a particularly promising candidate in this category. A professional services firm based in Boston, NerveWire focuses on helping clients develop more effective collaboration platforms with their business partners. It combines business strategy development with deep expertise in technology implementation. Since it is focused on establishing technology connections across enterprises, NerveWire will need to help clients integrate appropriate enabling services to support deployments of Web services technology across enterprises.

Service grid utilities. As enabling service aggregators and enabling service integrators emerge and become more prominent, specialized service grid utilities will also proliferate. Although most observers tend to focus on the role of entrepreneurial start-ups in this category, it is important to recognize the opportunity for larger, more established enterprises as well. We have already discussed the example of Citibank and its payment-processing engine to highlight the opportunity for companies with well-recognized brand names and scalable operational expertise to participate in this business arena. Recall as well that the technology underlying the enabling service need not be Web service technology. Earlier generations of technology can play an important role in service grids as long as their capabilities are exposed and made accessible to others with Web services interfaces.

Other, more focused technology companies can play a growing role in providing specialized enabling services. Take the example of Verisign in issuing SSL certificates, delivering Public Key Infrastructure (PKI) services and offering access management services to enhance the security of Web services connections. Verisign developed this business before the emergence of Web services technology, but this capability is very valuable in a distributed services architecture and can become a key component of service grid offerings. Verisign has a well-established reputation in this field and has a large installed base it can draw upon to serve this new market.

Of course there will also be room for entrepreneurial startups in this field as well. We were both associated with the formation of Grand Central, a company focused on providing a Web services network to support inter-enterprise connections with security, messaging, orchestration and monitoring, delivered as a subscription service. Grand Central illustrates the opportunity to offer a bundle of related enabling services, rather than simply one isolated service. It does not yet resemble a full-blown service grid in terms of the range of services provided, but it has the potential to evolve in that direction over time.

The early generation of Web services technology start-ups focused much more on providing development tools and enabling software to be installed on servers within the enterprise rather than providing enabling services on a subscription basis. Most of these development tool companies and enabling software companies will likely end up being acquired by application server providers and integrated into their technology platforms. A few, wishing to carve out an independent existence, will potentially evolve into enabling service businesses, leveraging their expertise in particular domains like directories, data conversion and monitoring of Web services performance to deliver managed services that are shared across multiple enterprises.

WHAT FORMS WILL SERVICE GRID BUSINESSES TAKE OVER TIME?

We are still at the earliest stage in the emergence and evolution of service grids and their constituent enabling service utilities. It would be presumptuous to pretend to know at this early stage how these businesses will look five or more years from know. We may not know, but we can speculate. Some, but not all, of the hypothetical evolutionary paths presented below will likely be explored and, of these, some, but not all, will possibly prove successful. We only want to present a map of possibilities to stimulate thinking. We are certainly not predicting the future. We are inviting readers to imagine it and then to invent it.

Forces shaping the outcomes

We can speculate by looking at the forces that will help to shape the outcome. These forces will play out on three levels – customer behavior, technology capability and economics. Of course, what makes it interesting is that these three levels are not independent but instead interact with each other in subtle and unpredictable ways.

Customer behavior. All customers are not created the same. They segment based on different needs and preferences that shape their behavior in sourcing technology. Broadly, business customers tend to fall into one of two camps: sophisticated piece-parters vs. convenience driven one-stop-shoppers.

Sophisticated piece-parters invest to create deep internal expertise in the integration of technology platforms because they don't want to compromise in terms of sourcing best in class technology products. Typically, these companies reside in technology intensive industries like financial services and they are generally the companies that compete by continuously innovating in terms of technology-enabled business design.

A much broader array of companies tend to be convenience driven one-stop-shoppers. This segment has driven much of the growth of the enterprise application market over the past decade – they would rather source a full suite of capabilities from one vendor, even if it means sacrificing in terms of ability to access best in class products. This segment has tended to grow over the past decade precisely because of the enormous challenge and expense required to integrate applications from different vendors.

Service grids represent an interesting opportunity to address the needs of both segments. On the one hand, sophisticated piece-parters may be more willing to do business with enabling service aggregators or, more probably, enabling service integrators because they would not require any compromise on accessing the best in class service grid utilities. On the other hand, convenience driven one-stop-shoppers may also rely on enabling service aggregators because they can still enjoy the convenience they seek in terms of one-stop-shopping while at the same time gaining access to a broader range of best in class service grid utilities.

Technology capability. Service grids can ultimately encompass an extraordinary array of specialized enabling services. Some of these enabling services are closely related to each other, both in terms of functionality and skill. For example, monitoring of Web services is likely to be closely related to the metering of usage (measuring frequency and/or duration of use by specific users) required for billing. In this case, it is likely that the same business that offers monitoring capability will eventually offer metering services as well. On the other hand, the billing for service requires such a different set of functionality and skill from either monitoring or metering that it is likely to be provided by a separate business.

It is this kind of functional bundling logic that led us to differentiate four different categories of enabling services in the service grid depicted in Figure 1. Broadly speaking, the services in each of these categories share enough similarities with each other that it may make sense for bundled service providers to emerge within the category. The shared utilities category is the one exception to this – these are set apart because they are so different from the other enabling services and yet they do not share much similarity in terms of functionality or skills with each other.

The similarity in functionality and skill may also help to drive many large enterprises to outsource enabling services over time, even though they may initially emerge within the enterprise. For example, many large enterprises today operate their own message queuing platforms and related transport management utilities. As more specialized third parties offer these services, larger enterprises may decide to outsource this capability to these third parties to enhance the ability to access world-class functionality and skill.

Economics. Enabling services may benefit from shared economics in terms of technology development, services operation or customer acquisition and management. These shared economics may in turn drive decisions to merge with related utilities or to outsource certain activities to even more specialized providers.

For example, the operation of the facilities required to deliver enabling services as part of a service grid may be much more efficient if the facilities are shared across a range of enabling services rather than dedicated to a single enabling service provider. In this case, facilities management companies may become an attractive outsourcing option for service grid utilities.

In another example, the various enabling services grouped in the service management utility category are likely to be selling to the same decision-maker. There may be an economic advantage for all of these enabling services to be marketed and sold through a common sales force. This may drive a tendency towards consolidation within this category of enabling services.

Likely trajectory

We are likely to see service grid business structures evolve through three broad waves of development, starting with bundling, then going through a fragmentation phase and ultimately culminating in a different form of bundling.

Initial bundling. In the early stages of the development of service grids, we are more likely to see bundles of enabling services created by early aggregators or within large enterprises. Earlier, we discussed the role of the first generation of electronic marketplaces and industry collaboration hubs in developing many of the enabling services required for their own operation and offering these enabling services in a bundled offering to their customers. Similarly, large enterprises will develop much of the enabling service capability to support early connection initiatives with business partners and within their own operations. This enabling service capability is likely to be offered to business partners as a bundle in encourage them to establish more efficient connections.

In general, this trend towards bundling will likely respond as well to customer risk averseness shaped by limited skill sets within most customers and the limited track record of more specialized service grid utilities. Customers are more likely to want to gain access to enabling service capability from a larger, better-established aggregator rather than venture out on their own and try to knit together a variety of specialized enabling services.

Fragmentation and specialization. In the second wave of evolution, we are likely to see a process of unbundling and fragmentation. The rapid development of even more specialized enabling services will drive at least some of this fragmentation. Both existing companies and start-ups will begin to offer more specialized enabling service capability to fill in the gaps in the initial bundles of services offered by enabling service aggregators. These more specialized services may be marketed to customers by enabling service aggregators, but they will remain independent businesses.

Customers will also become more sophisticated and confident in their ability to locate and evaluate the capabilities of service grid utilities. In particular, the piece-parter segment will likely become more active in seeking out best in class service grid utilities and creating their own internal integration capability. More specialized enabling service integrators will also become more prominent and help the one-stop-shoppers to access best in class capability, creating more opportunity for independent service grid utilities to reach and serve customers directly.

The first generation of enabling service aggregators will also start to spin out some of the more specialized enabling service capability they had developed internally as independent businesses. The economics of these aggregators will force them to focus heavily on investing in marketing, sales and customer support capabilities, rather than continuing to innovate in the technology required to support enhancement of enabling services. In part, this spin out of enabling service capability will be driven by competitive pressures from more focused aggregators who come into the market without any prior ownership of enabling services. Aggregators will also want the flexibility to access capabilities of other independent service grid utilities. These providers may be more reluctant to work with aggregators if the aggregators own enabling services themselves.

Thus, we are likely to see more specialized aggregators evolve that are less and less backward integrated into enabling services. In parallel, we will see more independent service grid utilities, each highly specialized in a specific type of enabling service.

Eventual consolidation. Over time, we will start to see consolidation occur within the aggregator business and within certain segments of enabling services. Enabling service aggregators specializing in similar customer or market segments will begin to merge with other smaller players in the same segments, driven by economies of scale and scope.

Enabling services providers will begin to bundle their services together through mergers and acquisitions of related services providers. This consolidation will in particular occur within the categories other than shared utilities identified in Figure 1. It will in part be driven by similarities in functionality and skill. Compelling efficiency benefits will also be a factor, created by the potential to share in the economics of either development, operation or marketing of enabling service capability.

Likely outcome

Over time, we are likely to see an industry structure designed to serve two needs. On the one hand, it will address customer desire for convenience and access to world-class capability. On the other hand, it will disperse opportunities for innovation to accelerate performance improvement.

Highly specialized utility bundles. Enabling services will tend to bundle together into natural clusters, shaped by common skills and economics. Within those clusters we are likely to see two or three very large providers, surrounded by a set of more highly specialized providers operating in narrow niches. Large enterprises may continue to operate their own highly specialized enabling services in certain areas where the functionality requirements are unique to one firm. In these cases, the enterprise will federate these internally provided enabling services with service bundles delivered by third parties.

Broad based services aggregators. These aggregators will tend to focus on specific target markets or customer segments. Within these segments, they will tend to become highly concentrated, evolving as collaboration hubs to orchestrate activities across a broad range of enterprises operating in the segment. These services aggregators will tend to own and operate few of the enabling services. Instead, they will create federations of service capability, drawing in part on the third party utility bundles described above and in part upon the enabling services that are still owned and operated by their customers.

Two tiers of services integrators. A few large, sophisticated enterprises will continue to maintain their own do it yourself integration capability, but most of the large services integration needs will be met by broad-based third party integration firms. These broad-based integration firms will tend to be large and concentrated businesses, specializing in recruiting and developing deep services integration talent. A more fragmented tier of integration specialists will continue to operate as sub-contractors to the broader integration firms. Over time, this tier of integration specialists will also tend to concentrate within their target niches.

Concentrated facilities management businesses. Managing the operations of service grid utilities will become a highly focused and concentrated business. Ensuring appropriate performance levels and owning and operating large, scale-intensive facilities will be the distinctive expertise of these facilities management businesses. Depending on the situation, these facilities management businesses may either serve enabling service utilities, services aggregators or larger enterprises continuing to own internal enabling services capability.

"Sweet spots" within the service grid industry

As service grids evolve, participants will naturally be seeking to focus on potential "sweet spots" that have the potential to generate above average returns. Identifying these sweet spots today would be highly speculative and depends upon layer upon layer of assumptions. Since we have already articulated an elaborate set of assumptions about the role and likely evolution of service grids earlier in this paper, we might as well go even further out on the limb and offer some early hypotheses about potential sweet spots. We see at least three interesting candidates.

Third party auditing/reputation engines. Choosing Web services and Web service providers (especially if they are not close business partner partners) will hinge upon assurance that these services are not only reliable but also trustworthy. This assurance is likely to be provided through third party auditing services that can independently verify performance of Web services in a broad range of application environments. These auditing services are likely to become highly concentrated and profitable, in much the same way that credit rating and bond rating services are. Reputation engines are also likely to emerge, providing a platform for users to rate the performance of Web services as well as their providers. Operating along the lines of EBay's reputation service, these reputation engines tend to have strong increasing returns economics – once a critical mass of users begin to rely on a particular reputation engine, it will tend to crowd out smaller competitors.

Business policy repositories and mediators. Everyone has tended to focus on the need to develop shared meaning within a Web services framework. Far fewer people have zeroed in on the need for, and challenge in providing, mediation of business policies across enterprises. The ability to provide this kind of mediation first depends on the ability to create repositories that capture and represent the business policies in standard formats. But that is only the beginning. The real complexity begins when enterprises seek to interact with each other even though they have implemented different business policies. Specialized services that can mediate across the different business policies and facilitate the agreement upon a common set of policies will be particularly valued. These businesses will become highly concentrated and profitable because of the proprietary expertise that can evolve from mediating a broader set of relationships than smaller providers of the same services.

Collaboration hubs. As we have indicated above, collaboration hubs are likely to emerge within particular industry or market segments, offering deep vertical or business process expertise to help companies operating in the same arena to more effectively collaborate together. Once again, the specialized vertical or business process expertise will drive the tendency towards profitability and concentration. This expertise will become more refined as early providers of this capability gain a critical mass of customers within the target segment, making it harder and harder for smaller and newer entrants to compete effectively.

WHAT ARE THE IMPLICATIONS FOR VENDORS AND CUSTOMERS?

So what does all this mean for vendors and customers? Let's start with technology vendors first and then move on to looking at customer implications.

Implications for technology vendors

Technology vendors have generally moved aggressively to support the standards and protocols that provide the foundation for Web services technology. However, they have tended to view Web services technology through an enterprise-centric lens. As a result, they have tended to overlook the role of the service grid as a key gating item in the adoption of Web services technology. As noted above, the service grid is particularly important in supporting inter-enterprise connections. It is precisely in this area that Web services technology has its greatest advantage and distinctiveness relative to traditional technologies. Adoption of Web services could be significantly accelerated if service grid capabilities were made available quickly.

Technology vendors need to decide what role to play in the deployment of service grid capabilities. At a minimum, they can play the role of evangelist and facilitator for early service grid utilities and service aggregators. They can educate their customers regarding the role and importance of the service grid. Vendors can also invest in certifying the capability of the early service grid providers and help to introduce them to their customers. In some cases, vendors may be able to play the role of aggregators, identifying essential service grid utilities and knitting their offerings together into a service grid offer targeted to their customer base. This is particularly an option for technology vendors like IBM and EDS that have expertise in operating technology based service businesses. In other cases, technology vendors may be able to support the development of a service grid by offering specialized enabling services. Earlier in this paper, we highlighted the role that Verisign will undoubtedly play in the Web services security arena.

Whatever choices the technology vendors make, they should explicitly recognize the implications of these choices in terms of competitive implications. Service grids will undoubtedly become the seedbeds for some very large new technology service providers as well as allowing existing technology vendors to diversify into related high growth markets. More broadly, it is possible that service grids over time will "hollow out" the functionality of existing technology frameworks and platforms like application servers and EAI middleware platforms. Much of the functionality currently offered by these frameworks and platforms may be more efficiently provided as shared services from specialized providers. This evolution would take considerable time to play out, but existing providers of these platforms may put themselves at risk by ceding this arena to other, more aggressive companies.

At a minimum, emerging service grid capabilities are likely to lead to a commoditization of enterprise-centric technology platforms, ranging from computing hardware to massive enterprise applications. To date, the difficulty in connecting these platforms together has helped to protect the vendors of these platforms once they have built up a substantial installed base. This installed base creates a compelling case for the customer to continue to bulk up on that vendor's products for incremental IT needs. Why look at a competitor's offer when the customer would then have the headache of trying to integrate it with the technology already in place? This logic has helped to support the growth of many technology vendors who have diversified into related hardware and or software offerings and persuaded customers that integration would be far easier if they sourced a broader range of their IT needs from a single vendor.

Service grids undermine that logic. With the ability to establish robust connections across diverse technology platforms, the logic for sourcing from one vendor erodes significantly. As this capability is deployed, customers can identify and source the best technology platform for a particular need, secure in the knowledge that the robust connections required to derive more value from that platform can be quickly implemented at a relatively very modest cost.

Implications for customers

Customers need to closely monitor the deployment of service grid capabilities. The availability of these capabilities will determine the range of environments and applications that Web services technology can effectively support. Customers can appropriately stage their Web services initiatives to exploit capabilities that are available from Web services technology while at the same time reducing risk due to limitations in functionality when other capabilities are not available.

Customers can do more than monitor and adapt. They can also influence and shape the deployment of service grid capabilities. CIO's need to become active evangelists for service grids, increasing awareness of the urgent need for the deployment of these capabilities. These executives can lobby with their primary Web services technology vendors to increase the sense of urgency among these vendors regarding service grid development. They can also help to spawn the development of new service grid utilities, aggregators and integrators by identifying businesses moving early to target these opportunities and then working closely to support these businesses. Their greatest impact will be to serve as early adopters of these services and to offer helpful feedback to the providers regarding enhancements required to make the services even more robust. In this way, early adopters will more quickly develop referenceable accounts and focus their own technology investments to have even greater appeal to customers.

Finally, large enterprises can also play a role in exposing existing capabilities and making them available as enabling services. We highlighted earlier the role that Citibank played in offering its payment processing engine to others through Web services interfaces. Moving beyond existing capabilities, large enterprises can invest to create new enabling services and begin by offering them to their business partners as part of a more far-reaching program to increase the efficiency and flexibility of connections in business processes encompassing these business partners. Over time, large enterprises might be able to offer these same enabling services to a broader range of companies, perhaps through service aggregators or integrators. In this way, they would help to increase the diversity of enabling services that can be mobilized through a service grid while at the same developing new revenue growth platforms.

***

Service grids will have far reaching implications for both Web services vendors and customers. The adoption of Web services technology will be significantly affected by the pace and scope of service grid deployments. In the absence of robust service grids, both vendors and customers will be deprived of significant opportunities for economic value creation. Supported by robust service grid capability, both vendors and customers will ultimately be able to realize the full potential of a powerful new generation of technology.

John Hagel III is an independent management consultant who work focuses on the intersection of business strategy and technology. His most recent book, Out of the Box: Strategies for Achieving Profits Today and Growth Tomorrow, was published by Harvard Business School Press in October 2002. He can be reached through his website www.johnhagel.com or by e-mail at john@johnhagel.com

John Seely Brown was the director of Xerox PARC until 2000. He continues his personal research into digital culture, learning and Web services. His most recent book (co-authored with Paul Duguid) is The Social Life of Information. He can be reached by e-mail at jsb@parc.com.