# SERVICE LEVEL AGREEMENTS AND BUSINESS CONTINUITY

*By Andrew Hiles FBCI MBCS*

Whether or not your organization has a Balanced Scorecard approach to mission achievement, the strategic use of service level agreements (SLA) can effectively be used to justify resilience and business continuity spend; to align information and communications technology (ICT) and other support services with business mission achievement; and to justify ICT and facilities infrastructure resilience and redundancy.

The first step is to, define the corporate mission. Take, as an example, a multinational company – call it Klenehost - selling miniature packs of soap, shampoo, hair conditioner and shower gel to the hotel industry. These are packaged in different ways and customized for specific hotel chains.

Klenehost states: "*Our mission is to be the number one vendor, world-wide, of in-room hygiene products to the hotel industry.*"

Fine – but what does that mean? Number one in what way? The biggest (by revenue/turnover)? The most profitable? Having seven of the top ten hotel groups as customers? Having a dominant market share in each of the geographic regions in which Klenehost operates? Having the products most liked by hotel guests?

Following board level discussion and business analysis, critical success factors (CSFs) are developed to reflect the board's definition of mission achievement. High-level key performance indicators (KPIs) are established – these are the numbers and ratios that reflect whether the CSFs have been met. For Klenehost, examples of KPIs could be: return on investment; net profit; turnover; customer satisfaction ratings from hotel guests; return per employee, market share by geographic region; key account penetration; customer churn rates; and employee satisfaction.

Initiatives can then be undertaken to put the necessary products, infrastructure, tools, methods, research etc. in place so that the mission may be achieved. Capacity plans and HR policies can be put in place to support mission delivery. Service specifications can be developed to ensure that services meet customer and business needs.

However, the problem with KPIs is that they are usually lagging indicators: you may only know whether you have hit the numbers when it is too late to take action to correct under-performance. The KPIs therefore have to be broken down into lower level business performance requirements and technical performance measurements: enter service level agreements.

For a graphical description of the above process, see figure one, below.

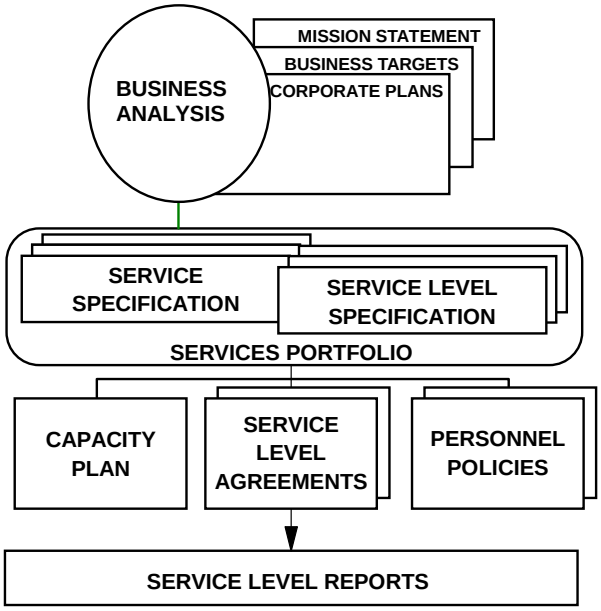



*Figure one: Linking service delivery to mission achievement*

SLAs for ICT used to be written in technical terms, typically reporting to end users in terms of the platforms from which services were provided – so they reported on items such as mainframe, server , WAN or LAN availability and response, typically with minimal business content. SLAs tended to reflect technical measurement over which the end user had no control and in which they had little interest. A parallel is with in-flight information provided to passengers. Because the information is available, the passenger is informed of the altitude and outside temperature. What use is this information to the passenger? What are they supposed to do about it?

Now, SLAs are applied to any internal support service – not just ICT but, for example, facilities, finance, HR, legal services and logistics. And SLAs can be associated with external supplier contracts to ensure that suppliers are aligned with customer mission achievement.

Technical measurement is important – but only to the technicians who can use it to adjust the service to ensure it does meet SLAs and hence support business achievement of KPIs and ultimately of CSFs and the overall mission. Thus the technical measurement is a leading indicator for ICT.

However, technical achievement needs to be put into a business context and the business or support unit needs to have ICT performance reported not on a technical platform level but in terms of overall service quality across all platforms that support the business activity. The CFO may use PCs, LAN, servers, printers, WAN and mainframe: but these are just tools. As far as the CFO is concerned, the deliverable is what matters, not the tool. Are invoices issued on time? Are credit control systems working effectively? Are debtors chased promptly? Is the payroll out on time? Is the Call Center working at optimum effectiveness in handling the maximum number of calls, maximizing sales and minimizing customer churn? The ICT or other support service's technical performance measures need to be translated into business terms, since they then reflect whether or not ICT's customers - the business or other support units – are meeting their service levels and hence their KPIs. Timely production of business performance reports enables ICT's customers to

take any remedial action necessary to ensure each unit is on course to support overall mission achievement.

So far we have de-composed high-level metrics into technical performance measures to establish a direct chain of results running through technical performance; the business performance supported by it; to mission achievement.

But we can do more. We can evaluate all of the business and ICT services, establishing how critical they are to business mission achievement and what the impact on the business would be of failure of each of these services. This is best done with input from the business and support units. Ideally, a high-level business steering committee should be established, with representation from finance and marketing, as well as from key operational and support areas.

Conducting a business impact analysis (BIA), we identify the criticality and the recovery time objective (RTO) for each service (that is, the maximum length of time the organization can afford to be without the service). We can also establish the recovery point objective (RPO) (that is, the point to which data must be recovered – e.g. start of day, end of day, or to a checkpoint). The results of this process will form the basis of the SLA requirements for availability and reliability (the number of incidents of outage) for each service. If a BIA has already been done for business continuity purposes, this needs to be retro-fitted into any existing SLAs so that they are made compatible with the business continuity plan. The same applies to external suppliers: for instance, we may have a requirement for 99.5 percent availability (in a 24/7 operation, this equates to about four hours downtime a year). A maintenance contract for support of this activity which allows four hours to get on site is simply inadequate.

The results can be sorted into tiers. A financial institution might, perhaps, define tiers as follows

| | |
|---|---|
| Tier One: | Continuous availability requirement: 99.999 percent availability, maximum of one outage and four minutes downtime per year. |
| Tier Two: | High availability, maximum of one outage per year, maximum four hours outage per year. |
| Tier Three: | Recovery essential within 24 hours; maximum three outages per year. |
| Tier Four: | Recovery required within 3 days; maximum four outages per year |
| Tier Five: | Delayed recovery – all other services. |

There may be as many tiers as appropriate, with the requirements for each tier adapted to each particular organization's needs.

The next step is to review the quality of infrastructure at each site. Clearly, the quality of infrastructure has to be capable of supporting the availability and reliability requirement for the Tier of service that is to be delivered to that site. Critical component failure analysis can be undertaken to establish the theoretical availability of the equipment, operating systems, applications and network on which the service depends.

Downtime percentages can be (broadly) calculated as follows (figures two and three):

| TIME PERIOD | CALCULATION BASIS | MONTHLY REPORT |
|---|---|---|
| Days in a month | 31 | As per month period |
| Hours in a month | 744 | As per month period |
| Minutes in a month | 44,640 | As per month period |
| Seconds in a month | 2,678,400 | As per month period |

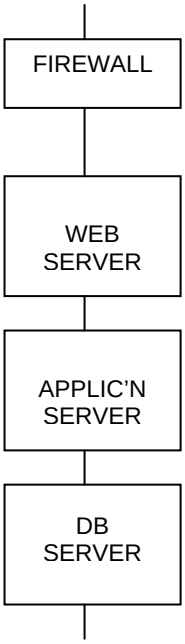
*Figure two: Basis of downtime calculation*

| MONTHLY AVAILABILITY | DOWNTIME MINUTES | MONTHLY AVAILABILITY | DOWNTIME MINUTES | MONTHLY AVAILABILITY | DOWNTIME MINUTES |
|---|---|---|---|---|---|
| 100 | 0 | 99.40 | 268 | 98.2 | 803 |
| 99.99 | 4.46 | 99.30 | 312 | 98.1 | 848 |
| 99.95 | 22.3 | 99.2 | 357 | 98.0 | 893 |
| 99.90 | 44.6 | 99.1 | 401 | 97.9 | 937 |
| 99.85 | 67 | 99.0 | 446 | 97.8 | 982 |
| 99.80 | 89.3 | 98.9 | 491 | 97.7 | 1026 |
| 99.75 | 111.6 | 98.8 | 535 | 97.6 | 1071 |
| 99.70 | 134.0 | 98.7 | 580 | 97.5 | 1116 |
| 99.65 | 156.2 | 98.6 | 625 | 97.0 | 1339 |
| 99.60 | 178.6 | 98.5 | 669 | 95.0 | 2232 |
| 99.55 | 201.0 | 98.4 | 714 | 80 | 8928 |
| 99.50 | 223.0 | 98.3 | 759 | | |

*Figure three: Resultant downtime percent*

The simple diagram at figure four shows a service with one access route, a firewall, a web server, an application server and a database server.

*Figure 4: Calculating Availability*




Including operating systems, middleware and application software there may be, say 15 components involved in the service shown in figure four. In this case, each component is a potential single point of failure. If each component has a 99.98 percent availability, the theoretical availability of the overall service is calculated as:

*99.98%x99.98%x99.98%x99.98%x99.98%x99.98%x99.98%x99.98%x99.98%x99.98%x99.98%x99.98%x99.8%x99.98%x99.98%.*

The overall availability works out at about 99.7%. Clearly this would be unacceptable for a Tier One (continuous availability) service as defined above.

To get continuous availability, the configuration might need duplication or triplication, with each configuration cross-linked by triangulated communications and geographically separated (e.g. one in New York, one in Dallas, one in Paris, France) so that the same physical disaster could not impact all configurations and any one could stand on its own. The system could be affected not just by hardware or software failure, but also by facilities issues such as power failure and also by overload. So capacity on demand and storage on demand could also be considered. Since (effectively) zero downtime is the requirement, data has to be mirrored in real time – there is no time for traditional data recovery from off-site backups. In this case, disaster recovery arrangements are not simply added on later: they are an integral part of the system design and build.

Once the requirement has been agreed by the business for each tier, it is simply a case of applying the rules. Clearly Tier One services are going to require more funding than Tier Three or Four: resilience costs money. However, the budget naturally follows the business decision.

The concept can be adapted to any organization. The construction industry is substantially less demanding than, for instance, banking. The example that follows reflects the development of SLAs

for a construction company where a central ICT function served six hundred different sites owned by six operating companies that operated internationally.

Applications were allocated to tiers in order to ensure that the infrastructure and support was provided to match the criticality of the application to the organization. 24/7 service was not necessary: ICT services were supported for an extended working day. The percentage availabilities shown in table one, below, therefore refer to scheduled availability while the loss of service times refer to elapsed time. Tier ratings were decided by business divisions who had to fund ICT accordingly. For new applications, the category influenced the design and resilience of the application, equipment and infrastructure to be used.

**Table one - application tiers**

| Category | Definition | Availability |
|---|---|---|
| Tier One | A mission critical application or service, where a loss of service of more than 18 hours will result in severe financial loss | 99.5% |
| Tier Two | A critical application or service, where a loss of service of more than 36 hours will result in severe financial loss | 99% |
| Tier Three | An important application or service where a loss of service of over 5 working days will result in significant financial loss | 98% |
| Tier Four | An application or service, where a loss of service of over 10 working days will result in significant customer concern | 95% |

Backup policy for each tier was defined by corporate IT backup policy and was designed to facilitate recovery within the relevant recovery time objective.

We could then specify the resilience, infrastructure and facilities requirements for each tier.

Risk analysis could then be conducted for each of the sites to establish its actual capability. Sites (i.e. the site where the user resides) were categorized in accordance with infrastructure resilience criteria, as shown in table two, below. The higher the level of resilience in the site, the more reliable would be the service and the more suitable the site will be to run high tier applications.

Bearing these tier requirements in mind, you might like to suggest possible requirements for site categories below – add any additional requirements you deem necessary.

**Table two – user site infrastructure categories**

| Requirement | Site A | Site B | Site C |
|---|---|---|---|
| Network | | | |
| Frame Relay | | | |
| ISDN Backup | | | |
| K per user (In and Out) | | | |
| File & Print Server on Site | | | |
| UPS for file and print server in communications equipment | | | |
| Dedicated Room for equipment | | | |
| Domain logon time | | | |
| SOE approved applications* | | | |
| SOE standard PC* | | | |

* SOE: Standard Operating Equipment.

The requirements for each application Tier were initially identified as shown below:

*Tier One*

| |
|---|
| Maximum 24K per user bandwidth requirements |
| RDBMS with roll forward/ roll back recovery implemented |
| Software/database vendor support – 24 hour x 7 day 4 hour response |
| Hardware maintenance agreement for 8 hour call to repair minimum |
| UPS – 30 minutes + generator |
| Computer room facility |
| Disaster recovery plan & facility |
| Backup & recovery procedures |
| Network redundancy |

*Tier Two*

| |
|---|
| Maximum 24K per user bandwidth requirements |
| RDBMS |
| Software/database vendor support – 24 hour x7 day 4 hour response |
| Hardware maintenance agreement for 24 hour x 7 day 4 hour response |
| UPS – 20 minutes + generator |
| Computer room facility |
| DR plan & facility |
| Backup & recovery procedures |
| Network redundancy |

*Tier Three*

| |
|---|
| Maximum 24K per user bandwidth requirements |
| Software/database vendor support – 12 hour x 5 day 4 hour response |
| Hardware maintenance agreement for 12 hour x 5 day 4 hour response |
| UPS – 10 minutes |
| Dedicated room |
| DR plan |
| Backup & recovery procedures |

*Tier Four*

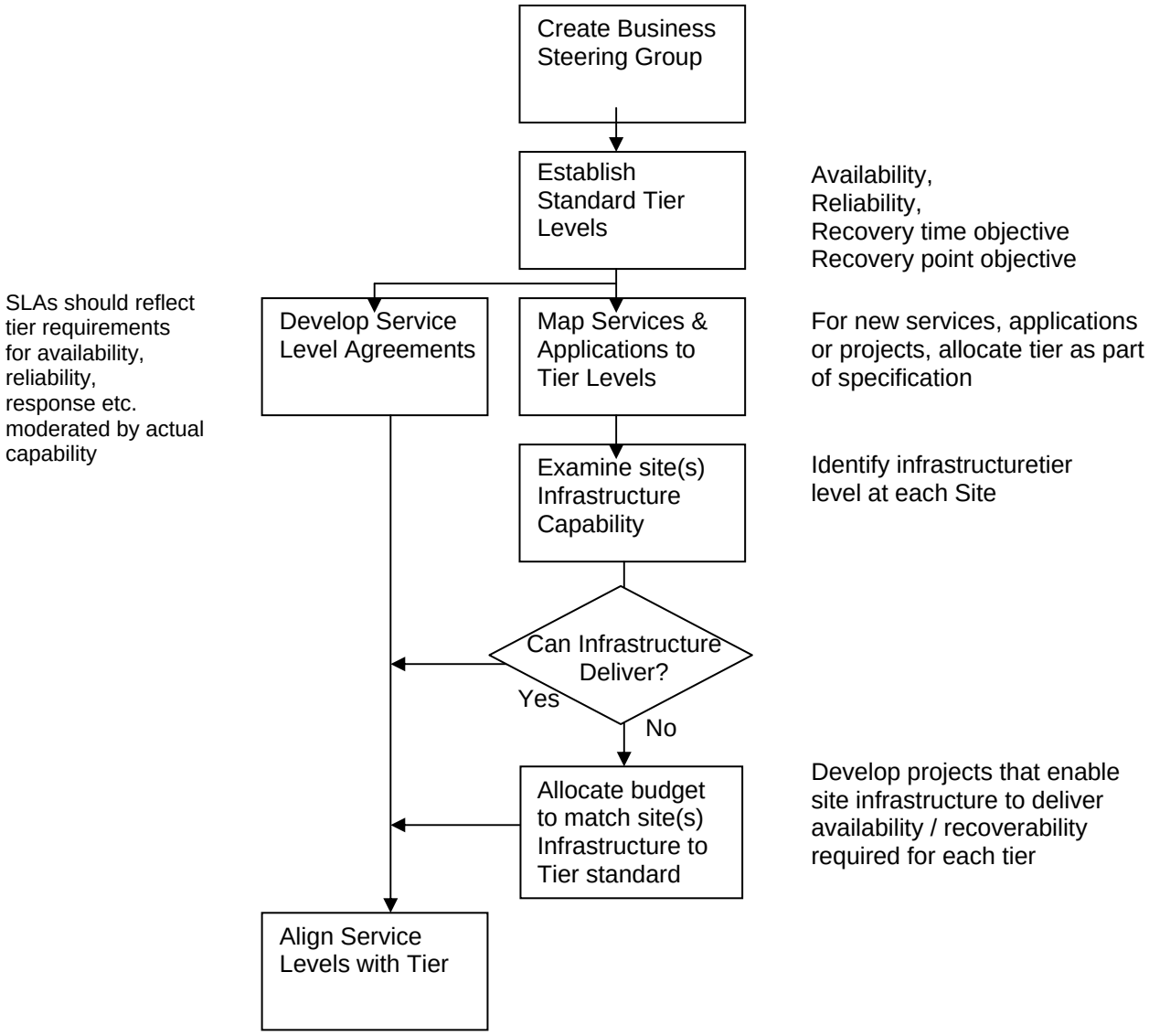
All others.

A gap analysis identified what site infrastructure is necessary to provide the necessary availability, resilience and response required by the application tiers. This could then be compared with the site category and risk assessment. Where there was a mismatch, either the site could not run the higher tier service or the investment had to be made to enable it to do so. Bearing in mind that the tiers were defined by the business steering group and the tier requirements were endorsed by this group, budget hassle was minimized since investment naturally had to follow the business decisions. So this approach helped to define ICT and facilities architecture and infrastructure strategy and management and embraced the business continuity requirements at the same time.

Problem management is also facilitated by this approach. The application tier was also considered in assessing the severity of an issue reported to the service desk. Tier One services are likely to have a more demanding support response service level than Tier Four services. Any loss of service or significant degradation of response to a Tier One service is likely to have a high priority, whereas a Tier Four service may never justify severity level one (i.e. the most urgent) response. Problem escalation timeframes followed the tier RTO and RPO requirements, with automatic escalation to business continuity management if it appeared that these could be infringed.

To summarize, tiered SLA were implemented as illustrated in figure five below.

*Figure 5: Tiered SLA Implementation Schematic*



By following this approach, the result was a compact 25-page SLA covering all ICT activities for the whole organization. This was the start: the same approach could then be applied to other business functions and external suppliers. More importantly, the Tiered SLA provided a coherent approach to service architecture and infrastructure planning, budgets, project and service specifications, business continuity, problem management and service level management.

**Author:**

Andrew Hiles is President of Kingswell International, a consulting company specializing in business risk management and service management. e-mail ahiles@kingswell.net www.kingswell.net .

He is the author of:

*The Complete Guide to IT Service Level Agreements, Matching Service Quality to Business Needs*. 0-9641648-2-5

*E-Business Service Level Agreements: Strategies for ISPs, ASPs, *SPs and CLECS.*

*Service Level Agreements, Winning a Competitive Edge for Supply and Support Services.* ISBN 0-9641648-4-1

*Business Continuity Management: Best Practice.* ISBN 0-9641648-3-3.

*Enterprise Risk Assessment & Business Impact Analysis – Best Practices* ISBN 1-931332-12-6

Editor and main contributor, *The Definitive Handbook of Business Continuity Management, 2nd Edition,* John Wiley & Sons. ISBN 978-0-4750-51638-6 (HB)

All books can be obtained from:

Rothstein Associates Inc
e-mail info@rothstein.com
www.rothstein.com