Appendix A

Service-Level Management

Say what you mean, mean what you say, and do what you said you would.
Barbara Colorosso

Service-level management (SLM)—it is a topic that is really simple and straightforward, but one that many people seem to have difficulty understanding or feel a need to make it more complex than it really is. The essence of SLM is embodied in the quote, although the associated processes are definitely more formal than that. Serviceable management is about ensuring that customers (internal or external) get what they are paying for and that what they receive matches their expectations. While the subjects of SLM and service-level agreements (SLAs) are most commonly associated with information technology (IT) and telecommunications, they are equally applicable to any kind of service. There are SLAs for every imaginable kind of service, ranging from baggage handling at large airports to weed control in Australian parks.

Much has been written about SLM (including the top-selling book on the topic by one of the authors, titled *Foundations of Service Level Management*). The objective of this chapter is not to address the subject in depth, but rather to provide the reader with a brief overview.

TERMINOLOGY

One of the big problems with SLM is that it is riddled with acronyms that are tossed about, without a proper definition of underlying terms. This ambiguity works to the disadvantage of the IT manager or service provider hoping to address it. Ambiguity fosters confusion. That confusion may lead to inaction or, worse, it may lead to inappropriate action (as in the case of purchasing the wrong tool to help with SLM). Before we can really begin our discussion of SLM, it is necessary to establish some common ground.

To begin, we must look at the term "**service**." Contrary to common supposition, SLM is not limited to IT and telecom organizations. It applies anywhere a "service" is delivered. However, this statement just further begs the question, "What is a service?" In the simplest terms, a service is any task performed by a person or group for another person or group.

SLM is often mistakenly equated with SLAs, but the two are very different. SLM is a process for delivering services that consistently meet client requirements. It entails defining, managing, and improving the levels of those services. This is true regardless of the type of service that is being delivered. It is just as applicable for an airline as it is for an MSP or the IT department in a large company.

An SLA is a component of an effective SLM program. Interestingly, it is possible to have an SLA without a program for managing service levels and conversely, an SLM program may be established

without creating SLAs. While both approaches are possible, neither is advisable. *The most fundamental aspect of an SLA is that it represents an agreement (or contract) between the service provider and the client, which defines the service to be provided and the level (or quality) of service that will be delivered.*

There are many facets to an SLA. It is a communication vehicle between the client and the service provider. Central to the SLA is the definition of terms in a way that will be meaningful to both parties. For example, the agreement must define the service that is covered, the level of service that will be provided, etc. It is this latter point that leads to one of the key benefits of an SLA: managing expectations.

Operational-level agreements (OLAs) are a specialized form of an SLA. OLAs are used within the service provider's organization to address subsets of the overall service. For example, an OLA related to an online billing system might address the network used by the client to communicate with the application. OLAs are more technical than an SLA with a client organization. Using the example of an online billing system, the SLA would focus on the end-to-end availability of the application and the overall response time or time to process a transaction. However, in an OLA, the network portion of that same service may include metrics for such things as: buffer miss ratio, buffer utilization, CPU utilization, line utilization, faults, discards, latency, error rate, packet loss, etc. These metrics are not normally meaningful to a client but are important to IT management, particularly those directly responsible for that network.

Next are two closely related terms: **service-level objective (SLO)** and **service-level guarantee (SLG)**. The main difference between these is contractual in nature. The SLO represents the level of service that the service provider and customer have agreed will be delivered during the term of the SLA. This means that the service provider agreed to meet that commitment. In order to agree to it, the service provider (including IT organizations) must be reasonably certain that it is possible to meet the commitment within the existing constraints (e.g., budget, staff, technology, etc.). In turn, the client's agreement means that they determined that they can function properly at this level of service.

The step from SLO to SLG is very small. It is often a legal difference. The term service-level guarantee (SLG) is far more common in formal, legally binding contracts between companies than in internal SLAs. An SLG is simply a guarantee by the service provider that the SLO will be met. It is common for the contract to include some form of penalty that will be imposed if the service provider fails to meet the guarantee.

SLM PROCESS

The minimal functions of SLM are to:

Define
Measure
Assess
Set objectives
Monitor
Refine and improve (Fig. A.1)

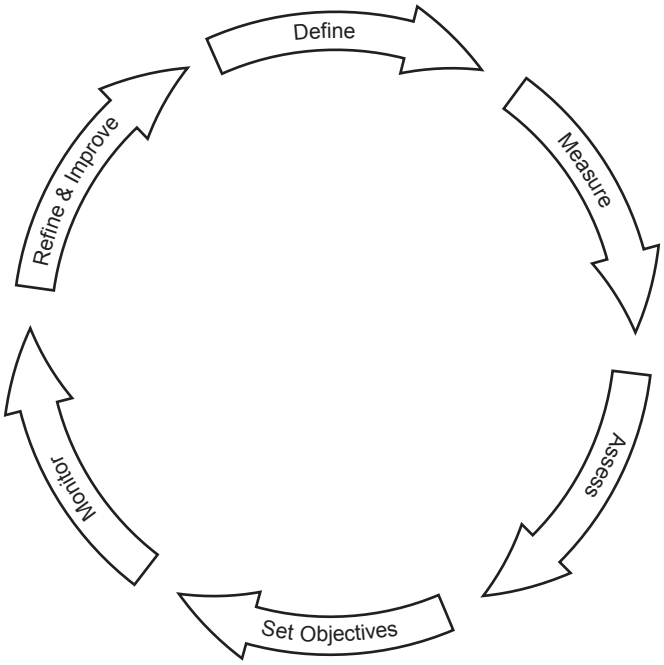

FIGURE A.1

SLM functions.

These six functions assume that there are other associated activities taking place in parallel (e.g., reporting, client dialogue, negotiating SLOs, negotiating SLAs, etc.). Also, this is an iterative process with feedback loops incorporated.

Define: Before any work can be done on SLM, you need to first determine which service you wish to manage. You also need to decide what metrics are available, or may be captured, that will give an indication of the level of service that is being provided.

Measure: The starting point for the process of SLM is to establish a set of baseline data. Of course, this should include the data that was identified in the definition phase as being representative of the level of service. However, that is probably not going to be sufficient. Data should also be captured that will allow system and network experts in your organization determine what would need to be done to improve the service. Some organizations find that the management tools they already have are sufficient for this part of the process, at least initially. Others find that they simply do not have the ability to acquire adequate information. They are then faced with the choice of abandoning the project or purchasing the additional tools. However, it is rare that an organization does not have access to any of the data necessary to assess the level of service provided. If none of the necessary information is available, then that service provider probably has much more serious problems than implementing a formal SLM process. They are probably facing the problem of sheer survival.

Assess: Once captured, the data needs to be analyzed to understand the current state. It also needs to be studied to identify opportunities for improvement. With these two pieces of information (current level of service and the potential for improvement), management is now in a position to engage in a serious dialogue about levels of service with clients. For now, management knows what it is delivering, what improvements can be made, and the cost of those improvements. Without this information, it is not possible to confidently negotiate SLAs.

Set Objectives: Negotiating SLAs is normally the best way to go about setting objectives for service levels. Through the SLA process, the objectives are set jointly. The result should be something that is realistic, attainable, and meaningful to the client. The targets that are set also must be affordable.

Monitor: Once the SLOs are established, the service provider must take steps to ensure that the SLOs are at least met. There follows an ongoing process of monitoring the actual level of service being delivered and measuring it against the SLO. Here again, ideally existing management tools will be sufficient. However, in practice, many organizations find that the monitoring and reporting part of the SLM process is improved (or made possible) with the addition of some specialized tools aimed at that function.

Refine and Improve: A service provider (whether it is an IT organization, carrier, ISP, or any other type of service provider) should never be satisfied with any given level of service, even if that level completely satisfies its obligations to its clients. There needs to be an effort aimed at continuous improvement of the level of service being delivered. This can be for defensive reasons (e.g., staying ahead of the competition). It may also be seen as a way to add more value to the service and ultimately derive more revenue.

Finally, the unstated seventh step of the SLM process is **repeat**. The process must be regularly scrutinized for relevance, accuracy, etc. The manager or executive responsible for providing the service covered by the SLM process must always be scrutinizing both the process and the service, looking for ways to improve both.

SERVICE-LEVEL AGREEMENTS

An SLA is a contract between the service provider and the client. If the service provider is part of the client's enterprise, the agreement can be less formal. Certainly, it will not normally be a formal, legally binding contract.

As shown in Table A.1, an SLA has several key components.

Table A.1 SLA Components

Parties	Reporting
Term	Reviews
Scope	Revisions
Service-level objectives (metrics)	Responsibilities
Nonperformance	Approvals
Exclusions	

Parties: This part of the SLA simply identifies the client and the service provider(s).

Term: An SLA does not run indefinitely. That is because the changes that occur over time will eventually make any SLA obsolete. Therefore, the SLA needs to have a duration specified. Having an expiration date forces periodic reviews of requirements, capabilities, and costs.

Scope: The scope section is where the services covered by the SLA are defined.

Service-Level Objectives: The SLOs represent the level of service that the parties to the SLA have agreed will be delivered. The level of service is expressed and measured by metrics. It is those metrics that are used in the SLOs. In the context of IT, metrics are generally about either availability of an application or its performance. Availability is usually represented by such details as total hours of availability, hours or minutes of outage, mean time to repair, etc. However, whatever metrics are chosen must indicative of the clients' actual experience.

Performance is significantly more difficult to measure than availability. The challenge is determining which aspect of performance is appropriate for the SLA being written. As with availability, strive for simplicity in performance metrics and try to make them as indicative of the client experience as possible. Usually, performance has a speed dimension to it. Speed may be represented by such things as response time, data transfer rate, etc. For online systems, the SLA should always include a metric for response time. Another common performance characteristic is volume. Volume may be reflected in such statistics as the amount of data transferred, the number of transactions processed, etc. There is no place in an SLA for such details as error rate, packet loss, percent of frames delivered, network latency, etc. Also, like availability, keep performance metrics as simple as possible. There is something seductive about performance metrics. They make us want to drill down into more and more detail. However, when writing an SLA, resist that temptation and go only as deep as necessary to accurately reflect the user **experience**.

Metrics for SLAs must be:
- Attainable
- Measurable
- Meaningful/understandable
- Controllable
- Mutually acceptable
- Cost-effective

METRICS

Metrics should always be defined from the perspective of the user rather than from that of the service provider. That is because SLAs are always about the users' experiences. Therefore, availability and performance of individual components together comprise the service but are not relevant to the user. The user sees and cares about their ability to access the functionality afforded by an application. That is what the SLA should reflect.

Attainable—The target service levels that are defined in the SLA have to reflect reality and not just wishful thinking. The service provider must know, or reasonably expect, that the objectives defined in the agreement can be achieved. To do otherwise is simply foolish.

Measurable—There is absolutely no point of including a metric for something that cannot be measured. In that case, how will you be able to assess whether or not you met the commitment?

For example, it might seem like a good idea to include end-to-end response time as a metric. However, if you do not have the ability to measure the actual response time that is delivered, then it is useless.

Meaningful/Understandable—There is a tendency by IT organizations to proposed metrics for an SLA that are not meaningful to the client. The metrics need to be defined in terms that reflect the clients' experience and that the clients can relate to.

Controllable—The service provider must be able to control any of the factors reflected in the SLA. If the service provider is not able to control a factor, then that factor should not be reflected in the SLA.

Mutually Acceptable—SLAs are supposed to be created with a sense of equality and partnership. Therefore, any metrics need to be acceptable to both parties. While it may be possible for the client to dictate terms to the service provider, or vice versa, it is a mistake to do so. Dictating terms defeats the purpose of the SLA.

Cost-effective—Restraint must be exercised when defining a set of metrics for an SLA. It can be tempting to define a wonderful set of metrics that would be prohibitively expensive to actually measure. Thus, when defining metrics, there is often a tradeoff between cost and precision of the data.

NONPERFORMANCE

A nagging question regarding SLAs is what to do about situations in which the service provider does not deliver the level of service promised in the agreement. This is an important question and it must be addressed in the SLA, but the question remains: what is the appropriate action in such a situation?

Thinking about this question usually flows along a path of reasoning that assumes that there should be penalties for failing to meet service-level commitments (nonperformance). It is also assumed that the penalties should be financial. However, at this point, popular thinking starts to bog down. If there is to be a penalty, how large should it be? Also, there is the problem of in-house service providers. If they are penalized financially, will that simply result in making it harder, or impossible, for them to meet the service commitments? Won't they still need to have adequate funding to operate?

If there is a penalty for nonperformance, should it be financial? For external groups, a financial penalty is appropriate, though not the only option. Service providers often like to offer free service as compensation for poor service. That can be compared to the experience of being served a bad meal at a restaurant and then being offered a free meal in the future as compensation. Does that actually mean that I will have a better meal or just another bad meal for free? Sometimes it may make sense to accept that offer, but more often it is not.

The key point of any penalty for nonperformance, whether for an in-house or an external group, is that the penalty must be significant. One must be pragmatic and realize that the purpose of a penalty is not to compensate to offset the impact of the substandard level of service. It is a virtual certainty that it will never be possible to negotiate such a penalty. Instead, the penalty must cause enough discomfort or pain to the service provider that avoiding it acts as an incentive to meet the service commitment.

There have been cases in which a service provider made service-level guarantees. Subsequently, the service provider decides that it is cheaper to pay the penalty than to provision to meet the service. This is obviously an unethical approach to business and one that is ultimately self-limiting, as customers

desert when their contracts expire. However, setting aside the ethical aspect, the key point to notice is that in such a case, the penalty does not cause enough pain for the service provider to want to avoid it.

How can penalties be made more painful? One obvious answer is to make the dollar amount of the penalty very large. Naturally, service providers are usually unwilling to accept agreements with very large financial penalties. The secret is to be creative. Pain can be caused by many things, not just the size of a penalty check. Things such as escalation or creating obligations for their executives, etc., should be explored to find possible alternatives.

Applying a financial penalty may be a viable alternative with an external service provider, however, with an in-house provider (such as IT), it is a dubious proposition. If money is taken away from an IT group, that may very well guarantee that they won't have enough money to meet the guarantees in the future. Just as with the external service provider, the secret lies in being creative. Remember, the goal is to cause pain, not to hinder the group's ability to provide services in the future.

Behavioral psychologists have long known that positive reinforcement is far more effective in modifying behavior than negative reinforcement. That is worth considering when negotiating service-level agreements. Consider rewards for meeting or exceeding guarantees rather than penalties for nonperformance. Bonuses for in-house staff can be an option. More radical are incentive payments to external service providers that deliver a high level of service.

Remember that when contemplating the creation of consequences for nonperformance, it is essential to be creative and remain flexible. Think outside the box. Consider both positive and negative alternatives and always remember that financial penalties may not be the best answer for your situation. Penalties for nonperformance cannot ensure that SLOss will always be met; however, they can significantly increase the likelihood of those objectives being met.

EXCLUSIONS

While the client and service writer can come to an agreement on the SLOs in other major aspects of an SLA, there are times or circumstances that should be excluded from consideration when calculating whether the SLA commitments were met. Those exclusions need to be defined and incorporated into the SLA. For example, holidays may not warrant the same level of service, or it may not be possible to deliver the same level of service on some holidays. Likewise, it may be necessary to make a provision for maintenance windows for the network, the systems, or even the applications. Acts of God, excess volume, third-party errors, problems caused by the client, etc. are other examples of things that may warrant exclusion from the agreement.

REPORTING

An SLA requires reports so that both the service provider and the client can see whether or not the SLOs are being met. The SLA should define the reports that will be produced in support of the agreement and who will be responsible for producing those reports. Normally, the reports will be the responsibility of the service provider. At a minimum, there should be periodic reports that identified the service covered by the SLA, the date the report was created, the period of time covered by the report, and for each metric included, the actual result versus the objective.

REVIEWS

Periodically, the SLA should be reviewed by both parties. Reviews should take place once or twice per year. That frequency should be spelled out in the SLA.

REVISIONS

The SLA should include a description of the process for making revisions to the agreement. The environments through which the services are delivered are never static. They are constantly changing. Likewise, the needs of the client will change and evolve over time. Therefore, the SLA needs to be adjusted to reflect those changes.

RESPONSIBLE PARTIES

Although really just a formality, as with any other contract, the parties to the agreement need to be identified in the agreement. This actually defines who must approve the SLA.

APPROVALS

Finally, the agreement needs to be signed (approved) by a client representative and a representative of the service provider, each with the necessary authority to do so.

SUMMARY

Within the context of IT and its clients, SLM is not just about applications, but applications are almost always part of SLM. Having an SLM program in place does not guarantee that clients will experience higher levels of service, but it does provide a mechanism for setting realistic expectations. It also defines what will happen if the level of service falls below what is promised in the SLA. In short, it provides the basis for an objective and dispassionate relationship between the service provider and the client(s).