

Mastering the Content Revolution

After decades of evolution, the technology for managing and publishing the content that organizations create and exchange has arrived at the threshold of being a core enabler for the way business is conducted. All of the pieces are now in place and all that is now needed is for these pieces to be made to work together seamlessly in support of enterprise solutions.

This last step of integration is the one that will determine whether organizations adopting XML really achieve all that it makes possible. Historically, however, it has been this last step that is the most challenging. The reasons why this last step of integration has proven so difficult is not itself hard to explain.

The Web Revolution

When the web burst onto the scene in the early 1990s, the notable simplicity of HTML and the web connectivity framework sparked an explosive rate of adoption. This in turn saw the numbers of people and businesses actively using content technologies grow by several orders of magnitude. The lesson of the web revolution was clear - *simplicity works*. But there are other lessons to be drawn from the success of the web, namely that simplicity always gives rise to new complexities. In the case of the web, among the complexities introduced was an accelerating growth in the volume and variety of content being created and exchanged.

The XML Revolution

With the late 1990s, a second revolution started with the appearance of XML. This revolution returned to one of the secret, and undervalued, ingredients in the success of the web – a venerable standard for encoding open information, the Standard Generalized Markup Language (SGML). XML was chartered to be a simplification of SGML for the purposes of enabling a more sophisticated web, one in which applications can interoperate as well as people. Among the goals the XML community set for itself, when creating a trimmed-down version of SGML, was called the ‘desperate PERL hacker test’. A programmer, it was declared, working in an unheated basement and without the benefit of doctoral studies in computer science, must, in less than two weeks, be able to develop a meaningful application that used XML as its content source. The target, then, was once again simplicity.

And XML was successful in achieving this objective, at least initially. But as had happened with the Web, simplicity begat complexity and suddenly organizations found themselves able to exchange real data with partners and to actually integrate systems that had been developed with blissful disregard for any such possible integration. Again, the volume and variety of content being created and exchanged took a dangerous leap forward. In reaction to this eventuality, the XML recommendation quickly became the family of XML recommendations, as new capabilities were added feverishly to keep up with the demands being placed on XML. In fact, before long the family of XML recommendations began to look like a large and boisterous family reunion where more seasoned standards mixed freely with the new and unproven.

The Scalability Challenge

Meanwhile organizations had real jobs to do. The public, emboldened by an increasing familiarity with a web-enabled marketplace, was now very much online, and expectations for effective online services were escalating quickly. It became critical that these organizations master the web and XML revolutions and keep up with this growing demand. Among the first places these organizations could turn was the XML community who, to their credit, had developed an attractive array of open source tools for performing common tasks with XML. There was also a large and growing community of developers, many of them being web developers who had recently added XML to their skill sets, who became accomplished at using these tools. And so it was that one of the stated objectives behind the XML revolution, of empowering the 'desperate PERL hacker', was achieved. Organizations were reasonably successful at engaging this community of developers, and these open source tools, as they ventured into this XML-enabled web.

It should be clear, however, that there is a key problem that remains unsolved thus far. Large organizations are typically focused on delivering mission-critical services reliably to millions of customers. At the heart of this outstanding problem is the question of scalability. How can the XML family reunion of recommendations, with all these open source tools and with this emergent community of XML developers, be used to field *enterprise solutions*?

Many organizations that had made their initial investments in XML-enabled content management, e-Publishing and interchange solutions in fact found themselves asking a different, but related, set of questions. Where did all the simplicity go? Why are my applications so slow? Why are my infrastructure costs going through the roof? What do I do with all this XML content? And, most tellingly, why are my users complaining?

Now it should be emphasized that none of this really amounts to a criticism of XML itself. XML set out to be an extensible mark-up language that was easy to process. The problems being listed are a result of its profound success. That organizations in fact have lots of content and have many, many things they want to do with that content to satisfy their customers is not a failing of XML. It is the challenge before these organizations to adopt, and the providers of content technologies to deliver, methodologies and tools that can scale to meet these inescapable demands.

The Web 2.0 Revolution

Perhaps unsettlingly, there are new revolutions on the horizon. Specifically, there is a growing level of excitement around the rather nebulous concept of Web 2.0. However nebulous the meaning of the term may be, organizations and technology providers would be wise to take it seriously because its key attribute is public adoption. The rapid rise in popularity of *social applications*, such as personal publishing using blogs, is too important a phenomenon to be ignored. And it can be highlighted that once again the root of the success seen in this new phenomenon is simplicity. Web 2.0 applications, if they share any qualities at all, exhibit a compelling simplicity in the way users interact with them and in the way developers create and maintain them. It should be noted that once again a judicious use of structured mark-up standards has been the secret and unappreciated ingredient in this success.

If previous history has taught us anything, the success of the Web 2.0 revolution, which is assured because in large part it has already happened, will introduce new complexities in the form of another order of magnitude increase in the volume and variety of content organizations must deal with.

The Semantic Web Revolution

Unfortunately the pace of change shows no signs of abatement and those who are already weary can see another revolution brewing on the horizon. A subject of growing interest is the possibility of adding to the web a level of semantic informativeness and logical precision that will enable applications to operate across the web with a fair degree of autonomy. While its adoption has been more gradual, the impact of the semantic web is already being felt in specific areas of web functionality. Interestingly, one of the areas of immediate utility is in harvesting the profusion of content being spawned by the Web 2.0 revolution.

What is sometimes referred to as the semantic wave is somewhat different from the revolutions surveyed thus far in that simplicity is not a defining feature of the semantic web. On the contrary, the semantic web is currently distinguished by a sobering level of conceptual and technological sophistication. What can be foreseen however is the fact that the semantic web is pursuing a valid, even inevitable, goal and that it will just take one element of simplicity to set off the semantic web revolution. Simplicity, as it has in the past, will act as the catalyst.

And when this revolution commences in earnest, organizations can expect to be confronted with another leap forward in the volume and variety of content that they must find ways to manage and leverage. The ongoing evolution of the semantic web does provide some insight into what types of content will be introduced. Whereas the content associated with Web 2.0 is typically voluminous and structurally undifferentiated, the content associated with the semantic web is made up of extremely dense ontologies, themselves comprised of precisely controlled structures and semantic values. Ontology instances are frequently very large, at times grotesquely so, and on top of this they are subject to constant change and to the requirement for synchronization across communities of applications that need to inter-operate. Perhaps controversially, it can be said that it will be the coming of age of the semantic web that will make the overall vision of an adaptive Service Oriented Architecture (SOA) a practical reality and the way all future *enterprise solutions* will be constructed.

Surviving the Content Revolutions

What should be clear from the foregoing, to reduce things to an over-simplification, is that 'desperate PERL hackers' and a collection of open source tools, however capable each of these may be, will not be enough for organizations to survive, let alone master, these successive content revolutions.

What are needed are enterprise-class methodologies and tools that can genuinely scale to meet the demands posed by a constantly increasing growth in the volume and variety of content that organizations must create and exchange. As if it needed stating, these methodologies and tools are not something that can be concocted overnight. They must in fact evolve over time and they must reflect the accumulated experience of many implementation efforts. And they must have emerged within the specific context of building, operating and evolving *enterprise solutions* where robustness, reliability and performance are essential criteria.

Not surprisingly, tools and techniques that work marvellously for processing small sets of web pages do not scale very effectively to meet enterprise requirements. Somewhat surprisingly, many an organization has deployed production environments for processing high-volume content transactions using development tools conceived of for laying out web pages. These organizations initially took pride in being able to deploy purpose-built hardware accelerators to meet the performance demands. The pride typically lasted until the space for hardware expansion, or the hardware acquisition budget, or the application maintenance allocation ran out – and frequently all of these ran out at the same time.

Succeeding in a Brave New World of Intelligent Content

Although there might be a growing sense of foreboding associated with the emergence of a new world of intelligent content, there should also be a real sense of excitement. Associated with the culmination of these content revolutions will be entirely new fields of technology and business innovation, delivering massive benefits to those that successfully adapt. It is an underlying premise in this story that one of the organizational life-skills in this new world will be the ability to efficiently manage and process content on hitherto unimaginable scales of volume and variety.

And it should be one of the sources of optimism when organizations look towards this rapidly approaching future that there are technology providers who have been working on this problem from the beginning. These technology providers have focused all their energies on building and evolving methodologies and tools for efficiently processing content within *enterprise solutions*.

If there is a message in this story, it is that organizations needing to progressively improve the ways in which they manage and process content will have to move beyond the tools and techniques that they relied upon during their initial XML explorations. So in addition to continuing to make the most of what the XML community can provide, these organizations will need to take a closer look at technology providers that have invested heavily, over the decades in which these revolutions have occurred, in advancing methodologies and technologies that can genuinely scale to meet the types of content processing demands confronting organizations both today and in the future.

Joe Gollner
Vice President e-Publishing Solutions - Stilo North America
19 March 2007

info@stilo.com

www.stilo.com