MANAGING LONG-LIVED INFORMATION ASSETS ... EMERGING STANDARDS AND TOOLS CAN $AVE BIG $$.

Kent M. Taylor, ISOGEN International Corp, Colorado Springs, CO

Abstract

The digital revolution has enabled rapid advances in product/systems capability and functionality that can only be described as *astounding*. Corresponding advances in the complexity of systems and even individual components is equally astounding, and occasionally *confounding*. Add to this the ability to fundamentally change the behavior of software-driven products and/or systems in near real time, and you have rapidly *compounding* growth in the amount of information that is needed to support today's products and systems over their long lifespans.

Effective management, distribution, and use of this continually growing stockpile of information is becoming a significant effort, and a plethora of tools have been introduced to address various aspects of the information management challenge. While the individual tools may streamline the variety of information creation, storage, change, management, production, and distribution processes, they also have the potential to compound the overall problem ... especially in the support of complex, long life cycle products. Major enterprise-level inefficiencies can be introduced if individual information handling systems cannot communicate *perfectly* with each other. And major disruptions can occur over the life of a product when information handling systems are replaced or upgraded ... and legacy data is found to be incompatible with the new system.

The Problem

It is generally agreed that the technical information is being generated at an exponentially increasing rate, as illustrated in Figure 1, below.

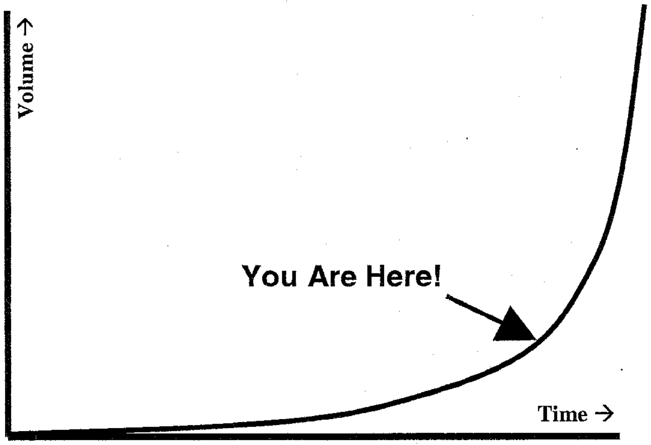


Figure 1 – Growth in Technical Information

The growth rate appears to be relatively independent of industry, with key drivers being

- Shorter Product Life Cycles (Technology)
- Software Driven Products (Easily Upgradable)
- More Suppliers (Global economy)
- Easier to Create (Computer-based Publishing Tools)

We can debate the relative position of the "you are here" marker, but it is my contention that we are at or near the knee of the curve; the real "information explosion" is just around the corner. While many enabling technologies are in place, I believe that what many have been calling the "information explosion" will barely qualify as the "fuse" when we look back from ten years out.

Today's situation feels like an explosion largely because we are either using old tools and concepts to manage and distribute the rapidly increasing volume of information ... or, we are applying new tools to manage information the old way.

0-7803-5086-3 /98/$10.00 ©1998 IEEE

Most of the volume increase depicted in Figure 1 comes from the automation and/or computerization of existing processes. It's not coincidental that the shape of the information volume curve looks a lot like the shape of computer chip performance curves; incremental increases in Workstation performance has contributed directly to increased ability to create/compile information. But we're reaching the point of diminishing returns by applying new technology to "Business as Usual" processes ... we need to look elsewhere for the next order-of-magnitude improvements.

Locally Optimized Solutions

Information is created, stored, and used in a variety of ways, by a number of different organizations within a typical business enterprise. For any given product,

- Engineering develops requirements and specifications
- Marketing/Product Management develops product plans
- Tech Pubs develops "User Documents"
- Sales develops advertising collateral
- "Everyone" develops administrative information.
- And so on for other types of organizations and the information they develop.
-

Good News: Over the last five to ten years, each of these functional organizations has discovered the joy of word processing, computer graphics, and/or desktop publishing ... and greatly improved their personal and departmental productivity. Many have also discovered alternate (other than paper) means of storing and distributing information.

Bad News: What should be a smooth pipeline of information from initial product definition through final User Documents, is now full of roadblocks. The tools used by Engineering are not compatible with those used by Marketing, which are not compatible with those used by Tech Pubs ... After all the information outputs from each of these departments *looks* different. They *"have to"* use different tools to get the desired output, despite the fact the fact that much of the content may be very similar, or even identical.

Summary: While each of the departments is efficient/productive when viewed independently, the efficiency of the Enterprise suffers as a result of the multiple translations/transformations that must take place at organizational boundaries.

"Obvious" Enterprise Solutions

Translation Programs: Because word processing and desktop publishing systems (and their users) are focused on how the *output looks,* and because there are multiple ways to make printed output *look* identical, the "obvious" solution of an automated/software translation rarely exceeds 80% accuracy. And a human being must then examine 100% of the output to find the 20% errors. Bottom line – it is frequently more effective to rekey or rewrite information that already exists in another department.

Make Everyone Use (fill in your favorite Word Processor): Remember why different departments have already chosen different tools ... they each have different *output* requirements. Sales brochures *look* a lot different than product specifications, which look different than Operators' Manuals ... Tools that are effective at creating jazzy brochures aren't real effective for creating specs or multi-volume User Manuals, and vice-versa.

This "solution" also leaves you with a legacy problem ... all of the active existing documents must be converted to the system of choice ... at 80% accuracy. And they will likely need to be converted again, when a new system comes along that promises a 20% productivity improvement (or the old system becomes obsolete/unsupported).

0-7803-5086-3 /98/$10.00 ©1998 IEEE

Object-Oriented Solutions

Given the right mindset, text and graphic objects can be created, stored, and manipulated in much the same way as software. And, just as the software development community has migrated from large linear programs to modular programs to object-oriented techniques, content creators are moving from large linear documents to modular documents, to object-oriented techniques.

A number of progressive companies are rethinking their strategies for managing their information assets, and reengineering their information development, storage/management, and distribution processes to take advantage of emerging information standards and tools.

Emerging Standards

There are a series of ISO standards that deal specifically with text, tabular, graphical, and multimedia information objects, and methodologies for describing them *independent of their presentation/formatting.* In other words, these standards focus on the logical structure and nature of content objects, and their relationships to one another.

This paper addresses the primary structural standard, ISO 8879-1987 *Standard Generalized Markup Language (SGML).* ISO 8879 defines structural and semantic rules for the descriptive markup of human readable text. Note once again, that the rules *do not address presentation/format of the text!* Also note that once the rules have been established, text files can be *parsed* to ensure that they conform to the rules.

This approach enables a new systems-level view separating Authoring/Editing, Storage/Management, and Presentation/Distribution processes, as illustrated in Figure 2.

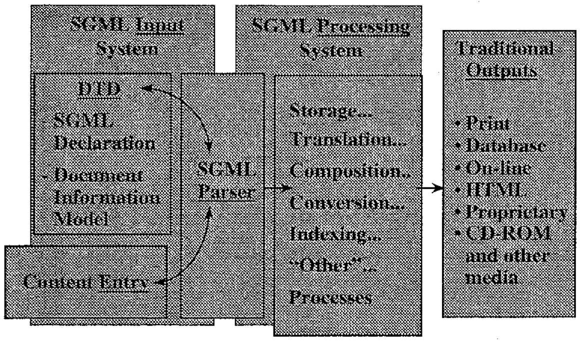


Figure 2 – SGML Systems-Level View

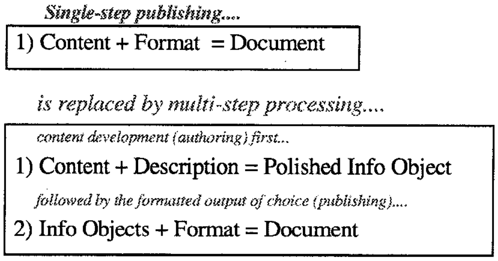
This view of the world results in features that can be extremely attractive in life-cycle management processes for information supporting large, complex systems:

- Authoring/editing tools can enforce structural and semantic rules through interactive, guided editing and/or after the fact parsing/reporting … all the way from initial Engineering Requirements through final Customer-deliverable documents
- Parsed Information can be interchanged easily across organizational and tool boundaries
- Parsing of text files against established, consistent rules enables accurate processing, translation, and/or transformation of text files
- Parsed files can be automatically formatted for multiple outputs (eg, paper, CD-ROM, Web, etc)
- Authors can focus on the quality of their content (versus format)
- Configuration management/control can be effectively applied at a more granular level
- Descriptively marked-up content can serve multiple business needs (eg, informational needs, control needs, ordering/inventory control)
- Information can be effectively reused (boiler plate, common procedures, and descriptions, admonitions, etc)
- Information can be effectively repurposed, customized, translated/localized

Sounds Good … What's the Catch?

Catch #1 – Significant Process Change

SGML-based information management enables a basic paradigm shift … away from the familiar/comfortable world of document-oriented single-step publishing where the author simultaneously addresses both content and format issues, toward a new object-oriented world where authors are concerned primarily with content and logical structure at a much more granular level. Figure 3 describes the conceptual shift that authors will face.



Single-step publishing....

1) Content + Format = Document

is replaced by multi-step processing....

content development (authoring) first...

1) Content + Description = Polished Info Object

followed by the formatted output of choice (publishing)....

2) Info Objects + Format = Document

Figure 3 – Difference in Value Added by Authors

This seemingly minor change in viewpoint can be traumatic for authors; after a lifetime of controlling both content and form of the final deliverable, authors will have to adjust to dealing almost exclusively with content. They may never even see many of the final products built from their individual information components. (Note: This "catch" can be overcome with appropriate doses of training and communication.)

Catch #2 - Significant up-front costs.

An effective SGML application requires a good deal of analysis, planning, systems engineering, and training of all participants, in addition to hardware and software costs.

Catch #3 – Legacy Conversion

The legacy problem still exists, and still has the problems outlined earlier. However, once converted to consistent, parsable SGML, content can easily be used by any conforming tool (authoring, editing, production, presentation), or translated for use by non-conforming tools and systems.

So, Why Bother?

In general, there are 4 overwhelming reasons to implement an SGML solution.

Reason #1 -- Customer Demand

SGML applications are gathering wide acceptance as Interchange Standards in many industries, including

- Aerospace
- Automotive
- Defense
- Telecommunications
- Semiconductor Manufacturing
- Pharmaceuticals

If your trading partner(s) demand that product information be delivered in an industry standard exchange format, it's probably a good idea to provide it …

Reason #2 – Customer Demand

Alternatively, if you have several large customers, all of whom demand *different* interchange formats, an internal SGML application is also likely to be the most cost-effective solution. Likewise if different market segments require different information products (eg, paper, CD-ROM, WWW, etc.), customized or translated/localized information products.

Reason #3 – Breakdowns in Current Systems and Processes.

As large-scale systems become more sophisticated, complex, and software driven, the information required to operate, administer, maintain, and repair them has grown exponentially. Document sets of tens or even hundreds of thousands of pages are not unusual in aerospace, defense, telecommunications, and other industries. Requirements for the simultaneous release of product and supporting

User Manuals in 20 or more custom versions, in as many languages is not unusual in Consumer-oriented industries. Requirements for "built-in" electronic "manuals" are becoming commonplace in everything from space vehicles to household appliances.

Old publishing systems and processes designed to put ink on paper just aren't capable of keeping up with today's requirements for the "first release" ... much less managing the on-going changes and updates over the life of the product.

Reason #4 – Cost, Quality, and Time-to-Market Improvements

If you are in a market niche where your Customers aren't looking for higher quality, lower prices, quicker time to market, and/or greater flexibility/customization, you may skip this section.

However, if you're among the other 99%, you may be interested to know that *serious* SGML applications typically pay back in 2 - 3 years, while significantly reducing cycle times and improving quality. I stress the word *serious* because the exceptional savings require a serious commitment to taking advantage of the paradigm shifts enabled by a well thought out SGML application. A "business as usual" SGML application generally results in (incremental) "savings as usual ..."

Show Me ...

SGML applications have been implemented by Fortune 500 companies in many industries, generally for one or more of the four reasons cited above. While Customer demand or impending breakdowns may have provided the initial motivation for exploring SGML-based solutions, implementation approval and funding has generally been based on expected financial benefit and/or market advantage.

And SGML applications have been demonstrated to deliver – big time – on both accounts.

We've worked primarily with Technical Publications Departments of companies both large and small, across many different industries, to migrate to SGML-based systems and processes, and have seen authoring productivity improvements of 10% to 50%, and production interval reductions from months to hours. Add to this the added bonus of noticeable improvements in the quality of the content, while enabling the production of innovative new information products, and the front-end *cost* starts to look more and more like a *wise investment.*

It is our expectation that even greater benefits will be realized by SGML applications at the enterprise level, addressing (at least) Hardware/Software Engineering, Technical Publications, Technical Training, and Customer Technical Support organizations.

Want to Know More?

Independent Case Studies

$GML – The Billion Dollar Secret, Chet Ensign, © 1997 by Prentice Hall PTR, Upper Saddle River, NJ. Presents five detailed case studies of large companies that have implemented major SGML systems. Includes a management-oriented overview of SGML.

Comprehensive SGML/XML Bibliography

The most comprehensive SGML/XML bibliography in the known universe is maintained by Robin Cover, at:

http://www.oasis-open.org/cover/

Additional information my be found at the host site,

http://www.oasis-open.org