

November 2013

Coming to Grips with SGML

Kathlene Karg

Association of American Publishers

Follow this and additional works at: <http://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Karg, Kathlene (1994) "Coming to Grips with SGML," *Against the Grain*: Vol. 6: Iss. 3, Article 37.

DOI: <https://doi.org/10.7771/2380-176X.1570>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.



COMING TO GRIPS WITH SGML

by **Kathlene Karg** (Assistant Director, Copyright,
Association of American Publishers)

As authors and publishers plunge into the digital age, they are grappling with issues ranging from rights to dissemination, from freedom of access to proprietary pricing structures. But one basic question confronts us all: If we convert from print to digital form, how can we be sure we'll only have to do it once? We don't want to have to keep re-keying the same document in order to meet the needs of each platform or device that's invented.

Increasingly, the publishing industry is looking towards SGML — the Standard Generalized Markup Language — for answers. But what is SGML and what effect could it have on publishing?

Almost a decade ago, the Association of American Publishers (AAP), established the Electronic Publishing Special Interest Group (EPSIG). Our purpose was and will be to provide a forum for the literary world to influence the development and implementation of SGML. Since then, we have aggressively pursued the development of electronic coding systems (called Document Type Definitions or "DTDs") for books, articles, and periodicals. Yet most people don't have a firm grasp of SGML and its publishing potential.

SGML is a standardized method of coding text — not the actual text, but parts of it — so that the format remains consistent throughout the entire document.

The coding can begin at any level of the publishing process. An author can submit a coded manuscript; an editor can code it during editing or a production person during production; in addition, previously published materials can be coded for future use.

You may well ask, why code it at all? Because coding it once in SGML means it's coded for life; the document and its pieces exist independent of the final format. An SGML-coded document is ready for the digital world and all the opportunity it offers.

SGML is a "language" that allows a document to be described by its parts, which can be used as a whole (as in a book or journal) or in pieces (as in a database or compendium). The user defines the parts, which typically include headings and footers, paragraphs, illustrations, and captions.

For instance, SGML places all journal articles in the same position on the page, makes the title of the articles consistent in type, font, and placement, and puts all page numbers in the same location. The concept aids publishers in developing uniformly formatted texts. Now, with the digital future staring

us in the face, the literary community is devising other applications of SGML. For years, the publishing industry's "techie" have been eagerly preaching their view of that future: a vision of grand information databases that would be delivered on demand anywhere in the world. Now everyone realizes that those visions are a reality.

Where does SGML fall into all of this? It allows publishers to create such databases — or other digital inventions — by placing the coded or "tagged" pieces of the text in a database. Queries from users can be answered with pieces of the document: an article instead of a journal, a chapter instead of a whole book. SGML's advantage is that it functions independently of the format of the final document. The text can be used in clearly defined form (e.g., what the publisher develops to publish a book) or in user-defined form (e.g., what fits on standard letter-sized paper).

Another usage example: at the start of the document, one can change the DTD from a book to a CD-ROM or, more simply, change the font from 10 characters per inch to four to serve a visually impaired sector of the market. Changing fonts, colors, styles, and setups thus becomes much easier than in manual translations, allowing for multiple publishing versions from a single tagged manuscript.

Although this description is extremely basic (anyone with a working knowledge of SGML is probably concocting a nasty letter to the editor), my intent here is not to prepare the reader to tag documents but to convey what the term SGML means in everyday language. Not all of us want or need to know more than that.

For those who want to roll up their sleeves and learn more about SGML (how to parse and compile), several groups are available to assist you. One, mentioned above, is **EPSIG**, which not only gives tutorials but also offers "Round Tables" on cost-effective ways to discuss SGML problems and success stories with others in the industry. The EPSIG office can be reached at (703) 519-8184.

Another friendly source of information is the **SGML Forum of New York**, (212) 512-3766. In addition, the many SGML support groups and gurus would love to demonstrate its wonders to you. Don't be shy; ask lots of questions.

As **Stewart Bland**, author of *The Media Lab: Inventing the Future of MIT*, once said, "Once a new technology rolls over you, if you're not part of the steamroller, you're part of the road." Don't let that happen. Learn to walk the walk and talk the talk. Technology is no longer just for techies. ☛