Standards Update—1995

James David Mason
Lockheed Martin Energy Systems
Oak Ridge National Laboratory
Bldg. 2506, M.S. 6302
Oak Ridge, Tennessee 37831-6302
masonjd@ornl.gov, http://www.ornl.gov/sgml/WG8/wg8home.htm

Oak Ridge National Laboratory is operated by Lockheed Martin Energy Systems, Inc. for the U.S. Department of Energy
under contract DE-AC05-84OR21400
The submitted manuscript has been authored by a contractor of the U.S. Government under contract DE-AC05-84OR21400.
Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this
contribution, or allow others to do so, for U.S. Government purposes.

James D. Mason, originally trained as a mediaevalist, is a systems architect in the
Information Management Services organization of Lockheed Martin Energy Systems. Dr. Mason
has been active in standards development since 1981 and has convened ISO/IEC
JTC1/SC18/WG8, the group responsible for SGML, since 1985. For his work on SGML, Dr.
Mason has received the Gutenberg Award from Printing Industries of America and the Tekkie
Award from GCA.

Keywords:
SGML, DSSSL, HyTime, ISO, WG8,

An Active Year (Whew!)

What a year this has been! Not since 1986, when we were finishing SGML in the first
place, has there been so much activity in the SGML world. In ISO, our long-standing home, there
are new standards being completed and old ones (some of which are not really all that old) being
revised. As you'll be hearing, there is lots of SGML activity in the applications
world—particularly on the Internet—and that's causing other kinds of standards activity.

Much of what happens in the standards world, particularly in a mature organization like
the ISO, is formal and rigorous, not to mention highly politicized. So I don't want to talk a lot
about procedures. If you really want to see how we work, come join us through your national
standards body. You can also look at the sorts of documents we handle if you look at our Web

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED
site: http://www.ornl.gov/sgml/WG8/wg8home.htm. There you can find most of our official papers, going back to the dark ages before the big year, 1986. You can see the evolution from using troff and word processors (though almost always with some sort of generic markup) to the latest documents, which are in SGML (both the HTML application and one written especially for our standards committee). Our formal statement about our activities is in the annual report I presented to our parent body at its meeting in Dublin back in July: http://www.ornl.gov/sgml/WG8/DOCS/1789ann.sgm.

Just for the record, the body responsible for primary SGML standardization is ISO/IEC JTC1/SC18/WG8. What all that means is that Working Group 8, Document Description and Processing Languages, belongs to Subcommittee 18, Document Processing and Related Communication, of Joint Technical Committee 1, Information Technology, which is sponsored by both the International Organization for Standardization and the International Electrotechnical Commission. If you really want to follow up on the official side, take a look at ISO’s Web site: http://138.81.2.245/.

WG8 divides its work into five “Rapporteur Groups” (or “RGs”) for DSSSL, Font Description and Interchange, SGML, SPDL, and Hypermedia Languages. Since you’re probably most interested in DSSSL, SGML, and Hypermedia Languages, I’ll only mention that the other groups have been active, too. The Fonts group has been doing amendments to its standards, ISO/IEC 9541 and ISO/IEC 10036. The Fonts group has been active in providing support for ISO/IEC 10646, the massive character coding standard that has drawn a lot of attention in the SGML world. The SPDL group has at long last finished its standard, the Standard Page Description Language (ISO/IEC 10180) and is about to publish it. I’ll be glad to tell you more about these projects if you contact me.

**DSSSL**

Much of the activity in WG8 this year has been stirred up because of our efforts to complete ISO/IEC 10179, Document Style Semantics and Specification Language, commonly known as DSSSL. We started on this project because we realized that SGML is not enough to specify a complete specification: SGML is good at defining structures and document grammars to represent those structures, but it isn't designed to convey the semantics of the structures. DSSSL is a standard that defines semantics. For a starter, it addresses presentation semantics, the pouring of text onto a display medium, but it's not limited to that. The real center of the standard is a set
of tools for transforming SGML document structures. The basic design work on DSSSL was
done years ago; indeed, the standard passed its ballot as a Draft International Standard back in
1991, and we could have pushed it on to completion then. (DSSSL work always seems to be in a
mad rush. Back in 1991, the job was helped to completion by hotel staff who took pity on Sharon
Adler, Anders Berglund, and their editing team at 4 a.m. and found ice cream to keep them
functioning.)

But the time was not then right for completing the standard. There's no doubt that
publishers have understood the semantics of layout for centuries and have been able to turn those
semantics into computerized composition for decades. But finding the right expression tools to do
that for SGML documents has not been easy. We got over one hump about two years ago when
we adopted the Scheme dialect of Lisp to handle such things as arithmetic expressions. However,
describing the interface to SGML structures, which becomes, in effect, a query language for a
database, has not been so simple. Complicating the picture was the presence of another standard
under our purview that also needed to query structures—HyTime. These two standards had
slightly different views of the world of documents. After all, DSSSL is clearly constrained to
dealing with SGML documents, and HyTime just as clearly includes dealing with more than
SGML data.

The breakthrough came right at the first of this year. Just before the WG8 meeting in
February, James Clark, representing the DSSSL RG, and Charles Goldfarb, representing the
Hypermedia Languages RG, spent a week working over the problems. The result was a
realization that an alignment of the standards seemed within reach.

The two key areas of alignment are a common tree model and a common query language.
You'll be hearing a lot about these at SGML '95 because they'll be changing how we think about
SGML and SGML documents and applications. Many of the principal developers of the
changes—Sharon Adler, Anders Berglund, James Clark, and Charles Goldfarb—will be speaking.
Since they have even later breaking news than I have, I'll leave most of the details to them.

The tree model is very important to DSSSL because the transformation of SGML trees is
central to everything DSSSL does. Whether an application is doing one of the most basic DSSSL
tasks, like specifying typesetting, or something more abstract, like converting an instance
document from one DTD to another, understanding tree structure and the location of elements in
trees is essential. In the case of typesetting, the original logical structure of the document must be
rearranged into what DSSSL calls flow objects on a presentation surface. That both of these
views of a document involve substantially the same data and that both are trees should be
obvious. Under the new terminology, such collections of trees will come to be known, logically enough, as “groves.”

Locating objects in trees depends on what has come to be known as a “query language.” One of the persistent issues in WG8 for several years has been the perception that HyTime and DSSSL used different query languages. Indeed, concern for this issue may have been the principal driving force to bring about alignment between the two standards. Although HyTime had a query language, HyQ (defined in A.3 of the standard), it will now adopt an expanded DSSSL query language, to be called “Standard Document Query Language” or “SDQL.”

**Hypermedia Languages**

Since its completion just three years ago, HyTime has received a lot of attention as a new standard in its own right. We’re delighted to have our work put to use, particularly after the incredible pressure of finally getting it published (just ask one of the editing committee about the all-night conference calls).

HyTime has also changed how we think about SGML and SGML standardization. One of its most useful concepts is that of “architectural forms.”

It is these SGML metastructures that HyTime contributes to the alignment with DSSSL. Much of DSSSL’s notation is being reconstituted in terms of HyTime architectural forms. For its part, HyTime will adopt an expanded version of the tree and grove model from DSSSL and will adopt SDQL as a replacement for HyQ, HyFunk, and HyOp. This substitution of query and operational languages is actually quite reasonable, because the original HyTime languages had an admittedly Lisp-like syntax, and DSSSL was using a syntax adapted from Scheme.

But the changes to HyTime aren’t the only things that are happening in WG8’s Hypermedia Languages group.

As some who have followed the development of SGML and related standards will remember, HyTime itself is a by-product of a project to develop a “Standard Music Description Language” or “SMDL.” HyTime began by generalizing concepts like measurement and synchronization that are needed to represent musical concepts. SMDL (DIS 10743) progressed in parallel with HyTime in the early stages, but when it became apparent that HyTime would have greater overall significance and could be completed more rapidly by itself, WG8 let SMDL sit for a while. Now, with three years of experience with HyTime, WG8 has turned back to SMDL, and Steve Newcomb and Charles Goldfarb have updated the notation to conform to HyTime and sent
it out for DIS ballot.

Another project coming out of the Hypermedia Languages group is the “Standard Multimedia Scripting Language” or “SMSL.” Like SMDL, SMSL will be cast in terms of HyTime constructs. Right now it is in early stages of development, but Brian Markey has put out a first working draft for comment by the standards community.

SGML

The main event! SGML is the core of almost everything that WG8 does. And now our development and revision activity is leading us rapidly back towards SGML.

Almost since SGML was first adopted nine years ago, WG8 has been getting the question, “When are you going to revise it?” At first the developers could answer, “At least let us try it in production for a while and see whether it really needs to be revised.” Now, everyone has tried it. We all know it works. We’ve needed only one amendment (unlike some other standards could be named). But by now many of us also have little lists of things we’d like tweaked. When the fifth birthday of SGML came up, we started looking seriously at a revision process.

WG8 is not ready to issue a revision yet. But we’ve learned a lot about what we might want to do in a revision. We began with collecting suggestions for revisions; by now we have a list of things involving many details of the standard, such as RS/RE processing. The most significant work lately, however, has concerned some of our overall concepts of what SGML does and how SGML goes about doing it. Our efforts have been brought into focus by the work on DSSSL and HyTime.

Over the course of about two years, Lynne Price and Dave Peterson have been working on a class-and-object model for SGML. Although some have seen this only as an intellectual exercise, its goals have been helping us see how an SGML processor must work and giving us a more precise vocabulary for discussing both the standard and the things that apply and support it. Now this work is merging with the results of the DSSSL/HyTime alignment.

Many things we have discussed—from object-oriented models to tree-and-grove models—find expression in SGML property sets. The notion of a property set has been around for at least five years. One of the earliest property sets to be recognized was the “Element Structure Information Set” or “ESIS,” which WG8 described in a 1990 paper. The need to recognize other sets was acknowledged, too: ESIS is a reasonable model for some kinds of batch processing, but it’s not enough for an SGML editor, for example. DSSSL development has been a
driving force towards expanding the set of properties than can be described for an SGML document and its processing. At WG8’s most recent meeting in October at Broomfield, Colorado, the discussions moved rapidly from settling the properties needed for the alignment of DSSSL and HyTime to the general topic of SGML property sets.

There are many other areas of discussion that will have an impact on SGML revision. Momentum seems to be building behind the new character-encoding standard, ISO/IEC 10646 and its application that is commonly known as “UNICODE.” Many SGML applications, notably HTML, want to use this standard, and SGML product developers, at the very least, will feel its impact. We in WG8 are already considering its implications for our revision process.

Moving SGML documents through the Internet has brought a number of WG8 members into contact with the Internet Engineering Task Force (IETF) over the past few months (JTC1, our parent body, has official liaison with the IETF). A lively discussion in the IETF’s MIME-SGML working group may result in at least two RFCs on the packaging of SGML documents as registered MIME body parts.

I will not predict when WG8 will send a revision to JTC1 for processing. Charles Goldfarb will have more to say about what WG8 is developing—and how those things might turn out—in several presentations later in this conference, and I wouldn’t dream of stealing his thunder.

The Future of the Standards Process

The voluntary standards process promoted by ISO and its national member bodies has served us well: after all, we have SGML and HyTime and will shortly have DSSSL as well. The ISO process is not ideal, but we’ve learned to use it to get our work done. There are other models for standards development, however. About a year ago the IETF process was being held up to ISO as a model for speed and efficiency. I’ve been watching (and occasionally contributing to) that process since SGML ‘94. I’m not convinced that it’s actually getting its work done any quicker than the ISO process—it’s simply doing it differently.

The ISO process is changing, too. With its emphasis on face-to-face meetings and distribution of numerous ballots and responses, it is simply becoming too expensive to maintain. Over the years I’ve built a fortress in my office: rows of filing cabinets to hold the thousands of documents that have been circulated in SC18 and its working groups. I’m running out of space—time to go all electronic!

Amid the flurry of activity this year, one of the most encouraging developments in WG8
has been the increased practicing of what we preach. We actually use SGML to develop SGML-related standards. Of course, way back in 1986, ISO 8879 set the course: SGML was published using an SGML application. HyTime and DSSSL (and SPDL and the fonts standards) are also stored as SGML master documents. But suddenly SGML technology is traveling everywhere with us on our laptop computers. There was a sudden pause in WG8's summer meeting in Dublin when we looked around the room and realized that almost everyone there was sitting at a computer looking at some sort of SGML document. Although it wasn't the only system in use, SoftQuad's Panorama was on nearly every machine. We spent most of the week with a PC hooked up to a projector, working on live SGML documents.

The next week at the SC18 plenary, the ODA developers never quite recovered from my delivering WG8's annual report live, using several SGML display systems. I was asked to chair a session at SC18 on the application of information technology to standards development. One of our recommendations was that we use SC18 standards to distribute SC18 documents. Thirty seconds after that recommendation was approved, I was able to deliver my report to SC18—as an SGML document, of course. The ODA folks still don’t know what hit them.

WG8 has essentially stopped distributing paper documents. We have several WWW and FTP sites, and we put up everything from meeting notices to discussion papers there. Even in our meetings we print hardly anything. If the ISO process is to survive, the standards bodies would be well advised to look carefully at what WG8 is doing.

1995 has been a thrilling year for SGML standardization. But the work is not over. I invite everyone interested in the future of SGML to participate in the process. After all, it may be only a mouse click away.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.