02 The Basics

## Layout and structure of a text

A text is a collection of words and words are composed of letters. In order to read a text, we make use of a variety of layout aids. The most simple and basic of these is the use of spaces between words. Classical Greek and Roman texts were written in 'continuous script', that is without spaces between words. In the course of time, an increasing number of design and layout aids were introduced: not only spaces between words but also capital letters at the beginning of sentences, line breaks, and a variety of punctuation marks such as commas, semicolons, colons and periods.[^hist] Also, concepts such as paragraphs, chapters and so forth, developed into a standardized system that allowed the structure provided by the authors to be transmitted to readers familiar with these standards, thus enabling ease of reading and reading in silence.

In the electronic handling of texts, this hierarchical ordering of words into sentences, sentences into paragraphs and so on, and further reading aids such as punctuation, bold or italic text is made possible by using certain codes. This process is called **\*\*markup\*\*** and the codes are called markup elements. It goes without saying that all these markup elements require stable definitions and clear relationships. In order to establish which markup is allowed and how it should be used, **\*\*markup languages\*\*** were defined.

With the emergence of computer networks and the increasing need to standardize texts for multiple usages, an international ISO standard was defined in 1982, known as Standardized General Markup Language (SGML). This logically structured markup language was important, as it established a split between the text structure as such and the final representation of that structure. SGML defines functions or roles, as opposed to markup languages used in word processors as Microsoft Word, WordPerfect or OpenOffice. What does that mean? In the latter, presentation and text structuring are mixed together: when we type a word in **\*\*bold\*\*** in the text is immediately is represented in its layout as **\*\*bold\*\***. What is happening in that case is a mixing of **\*\*layout\*\*** and structure. But using the word processor we in fact type 'start bold' -\> type the word -\> 'end bold'. This is most easily understood when you think of keyboard shortcuts such as ctrl-b (Windows) or cmd-b (Mac) for bold. However, a *\*structured\** text with markup such as SGML, that is cut loose from the layout, reads ‘<bold>the typed word</bold>’.

Only in the next stage the **\*\*plain text\*\*** ‘<bold>the typed word</bold>’ is assigned with a specific layout, namely **\*\*the typed word\*\***. SGML and its derivatives, the simpler **\*\*HTML\*\*** (HyperText Markup Language) and the expanded XML (Extensible Markup Language) thus make a strict distinction between structure and representation. Markup languages as such allow for concepts such as 'bold text', 'chapter heading' or 'quotation' to be linked to a specific layout or function type (described in a separate style sheet as type 1, 2, 3...). This makes it possible to define, for example, chapter headings as type 1 and quotations as type 4. For each specific output you can then define in the layout phase how exactly this will look like. For instance, chapter headings will be rendered in a certain font and font size and centered on the page (as described with type 1), while quotations will be rendered in the same font and size as the running text, but in italics (type 4 for example). For another output, for example a display screen, chapter headings could be defined differently as pink text and quotations as yellow text.

![](images/03\_4\_boldstrong.png)

The layout that is separated from the content and that can take on all these different forms is defined using what is known as a **\*\*style sheet\*\***, a list of elements linking layout definitions to markup definitions. With the style sheet the marked up content is displayed as a designed text with its specific layout. Using style sheets to define the layout of the markup of the text is an essential requirement if a source text is to be represented in several different ways on various media of various sizes, such as a Kindle, a print book and a smart phone. Combined with different style sheets for different output formats, only one structured text is needed to make different product in an efficient way. Therefore the golden rule in hybrid publishing is to *\*always make sure that the source text is as systemically structured as possible\**.

### What is an electronic text?

An electronic text is generally understood as a text which is represented on a screen of some sort. Of course this is sloppy language. The important difference between a paper and an electronic text when it comes to hybrid publishing is the notion that the text becomes a structured file. Due to this markup, we can make different layouts, expressions, onto different media. It is of great importance to note that electronic publishing introduces a big shift away from the page centered culture of book printing. Book printing allows for various printing sizes depending on the wishes of the author, designer and publisher. For example, when making an art book based on a collection of paintings or drawings, a decision can be made on what the ideal book size is and whether or not it will be printed in for example the oblong (landscape) format. In the world of screens these types of decisions are different as there are many different screen sizes. No screen can be cut to the demanded size like with paper book publishing.

Many electronic reading devices now support **\*\*reflowable\*\*** text, which means the size of the text automatically fits the size of the device. However, this mostly works for genres such as novels which generally consist mostly of text. When the overall design and the relation between illustrations and other elements such as footnotes or tables, and body text is important, the transposition from an existing print work to an electronic representation is rife with difficulties. In other words: especially for arts and design publications and works presenting research. In the following chapters these issues will be discussed further.

## What are some electronic possibilities?

The most widely used form of e-publishing is perhaps scanning paper books. Book scanning is a process that is still being used to convert physical books into digital media such as images, electronic texts, or an ebook. **\*\*Optical character recognition\*\*** (OCR) is often used to convert a book into a digital text format like **\*\*ASCII\*\***. After this conversion the user can reformat the text, search through it, or process it with other applications. Examples of book scanning by organizations on large scale are projects like: Project Gutenberg[^Project-Gutenberg], Million Book Project[^Million-Book-Project], Google Books[^Google-Books], and the Open Content Alliance scan books on a large scale.

New electronic technologies enable a great number of possible publication outlets. This abundance has pros and cons. Pagination and indexing, which have become so important since the printing press, become problematic in e-publications, where page numbers are unstable or not even used. A historical illustration of this problem is the Bible with its numbering of phrases, which made it possible for all readers to refer to the same passage, even when using different copies with different sizes (or even languages!). Fixed page numbers and indexes referring to them allowed for identical copies to be used by readers, writers and researchers spanning years and even centuries, independently of location. In an electronic world with non-fixed screen sizes this is complicated. The text might remain fixed, as the text file exists independent from its final format (**\*\*e-ink\*\***, **\*\*LCD\*\***, paper), but its presentation on this final format can vary substantially. The challenge becomes even greater if we widen our ambition to pictures, audio and video, hyperlinks, but also tables, references, notes etc.

As cultural standardization over the centuries rendered, for example, an octavo, pocket book or a collection art reproductions in a ‘coffee table book’ instantly distinguishable by its typography and layout, – this too will likely become a fact for different genres of electronic books. Apart from the electronic (or paper) book as a *\*object d'art\**, electronic art book publishing will have many commonalities as a genre despite its differences per book.

The electronic sub-structure makes all files in an ebook just bitstreams (streams of binary code: zeros and ones). The great new thing in the world of electronic art books is that based on standardized, though well-tailored structures, the creative content can be published in a great variety of ways. This not only depends on the capabilities of the output (reading/viewing/listening) device, but also on the function of the book, such as a dictionary, a study, a reference, a catalogue, or a leisure book. All these chances and possibilities demand a thorough and more labor-intensive editorial and production strategy. Not only because the same source can express itself in various output forms but even more so because once properly edited and stored electronically, information and its constituting parts can be reused and used in different ways, to be decided upon given a specific environment of goal, now and in the future.

[^hist]: Guglielmo Cavallo and Roger Chartier (eds.), A History of Reading in the West, Polity Press 1999.

[^Project-Gutenberg]: Project Gutenberg, https://www.gutenberg.org.

[^Million-Book-Project]: Million Book Project, https://archive.org/details/millionbooks

[^Google-Books]: Google Books, http://books.google.nl/

[^Open-Content-Alliance]: Open Content Alliance, www.opencontentalliance.org/