

The Chicago Manual of Style *Online*

How Books and Journals Are Produced

[Overview](#)

[Markup](#)

[General Principles](#)

[XML](#)

[Design](#)

[The Electronic Workflow](#)

[Introduction](#)

[When to Introduce XML](#)

[Source Files](#)

[Options for Presenting Content](#)

[Print Technologies](#)

[Introduction](#)

[Paper](#)

[Printing](#)

[Binding](#)

Overview

This material is intended to supplement the latest edition of *The Chicago Manual of Style* by giving an overview of the production process and the technologies involved. Production—of a book or a journal article, for publication in electronic or printed form—generally begins the moment an edited manuscript is considered final (see chapter 2 of the *Manual*). The details of what happens after that depend not only on the formats in which the work will be published but also on the tools used to produce them. Steps taken earlier, starting with the author or manuscript editor, can also contribute to the production process.

With the right software—and at least a working knowledge of basic markup or, for print, of typography and layout—anyone can publish. But to produce works of the highest quality, often in a variety of formats delivered simultaneously via multiple content providers, publishers depend on specialists. Designers create a typographic and visual format that is appropriate for a particular work in a given medium. Production controllers coordinate in-house staff and vendors to ensure timely publication, within budget, and adherence to the highest standards. Editors and authors—once their main roles are done—continue to play a key role as reviewers during various stages of production. Meanwhile, programmers and other IT specialists develop and maintain systems for document conversion and typesetting as well as other aspects of publication, from revising and archiving to scheduling and communications.

The XML workflow outlined in this supplement stems in part from the experience of academic journal publishers, who were responding to a demand for simultaneous print and electronic publication, typically in PDF (the basis of printed formats) and HTML (for full-text web presentations). This demand dates to the 1990s—long before electronic formats for books became popular. A journal, because it tends to maintain the same format from issue to issue, is particularly well suited to an XML workflow (XML was in fact initially designed to facilitate HTML presentations). Once implemented, the same sets of tags and the same procedures (some of which have been developed specifically for publishers and enjoy broad support) can be reused and refined as necessary, leading to certain economies of scale.

Book publishers can streamline their production processes to accommodate a similar approach, and many publishers now offer a variety of e-book formats in addition to print. As in journal publishing, book publishers can use XML as the basis of publication in both print and, in the case of books, EPUB, a standard format based on XHTML (HTML defined as an application of XML) that can in turn be used as the basis of (or converted to) a number of commercial e-book formats. (XML can also be used to facilitate full-text web versions of books, like the one produced for *The Chicago Manual of Style*.) Though books are typically more complex than journal articles, the process of deriving XML from a book manuscript can be aided by the use of word-processing templates, or the markup can be added or derived later in the process. Many book publishers instead rely on tools that produce EPUB directly from the files used for print. Whatever the approach, a properly structured manuscript is essential, and even those who are not involved in production can benefit from an introduction to XML and to markup languages in general.

A note on terminology. Many design and production terms are derived from the technologies and processes used before the advent of computers. For example, the term *font* once referred to a specific size and cut (e.g., twelve-point italic) of a particular type family, or *typeface* (once literally the surface, or face, of a piece of type); now that type is rendered and scaled by software, the terms are often used interchangeably. Similarly, the terms *typesetting* and *composition* have become interchangeable, both referring to the process of arranging words and images on a screen more or less as they are intended to appear in published form. Even the term *markup*, which now often refers to the tags used in XML and other markup languages to identify the parts of an electronic document, has its antecedent in pencil markup on paper typescripts and proof. For an extensive list of design and production terms with definitions, see the glossary in *The Chicago Manual of Style*.

Markup

General Principles

Markup comprises the labels or annotations that are applied to a manuscript or other document in order to identify its parts, from chapter numbers and chapter titles to subheads, paragraphs for running text and block quotations, emphasized text, entries in a bibliography, and so forth. Markup is usually applied to a manuscript in one of three ways: (1) as generic labels added to a paper manuscript or typed directly into the document; (2) as word-processing styles; or (3) by means of a formal markup language such as XML. Generic markup is converted to styles or to a formal markup language for publication. The following discussion applies mainly to formal markup languages.

Markup that describes the parts of a document rather than their appearance is sometimes referred to as *semantic* markup. To take one example, though the title *Origin of Species* might be distinguished from the surrounding text by italics, more meaningful markup would label it as a book title. Such markup would not only differentiate book titles from other types of italicized text (such as names of species or emphasized terms) but would allow them to be presented in something other than the customary italics if desired—for example, boldface or underscored. Such an approach also makes content more accessible to readers who use text-to-speech and related tools to interpret the text. In fact, matters of presentation—the *appearance* of the text—are best defined in a separate document called a style sheet.

On the other hand, markup does not ultimately need to be associated with any typographic characteristic or other type of formatting. It can be purely functional. For example, the title *Origin of Species* might be tagged to include the keyword *evolution*—facilitating queries to search engines. Moreover, some markup may have formatting or functionality associated with it in one medium but not another. For example, markup for hyperlinked text—which can be tapped or clicked in electronic formats—may not be evident on the printed page.

The formal definition of a document's structure—among other things, the names of all of its elements and the rules that govern their markup—is encoded in a *document type definition* (DTD) or other schema. Documents marked up using XML or the like can be analyzed with the help of parsing programs to make sure they conform to a specific DTD, allowing for the ready identification and correction of some kinds of tagging errors. Journal publishers typically use industry-standard DTDs for publishing and archiving their articles or books. For example, some journal publishers rely on the family of DTDs originally developed for the National Library of Medicine and now used in conjunction with the Journal Article Tag Suite, a standard maintained by NISO (the National Information Standards Organization). Similar tools are available for book publishers. Many book publishers offer e-books that are based on EPUB, an open standard format developed by the International Digital Publishing Forum. EPUB files can be validated using a tool offered by the IDPF that, among other things, checks for valid XHTML.

While a DTD or schema defines a document's structure, a *style sheet* provides the formatting specifications that define and automate its presentation—that is, how it looks on the screen or on the page. A formally marked up document published in more than one medium—for example, print (or PDF), HTML, and EPUB—will need corresponding style sheets that specify typefaces, layouts, and other features appropriate to each medium. Then, as long as illustrations and other such accessories to the text have been properly prepared, the document itself—its markup and text—will not need to be altered for each presentation. This separation of content from presentation—though it is rarely perfect—is one of the primary benefits of an XML-based production workflow.

Style sheets are prepared (generally by IT specialists in collaboration with designers) in one of several programming languages. CSS (cascading style sheets) can be used to describe how the components of an HTML or EPUB document will look and function in a particular browser or other application or how the document will print. Another language, XSL (extensible style sheet language), not only defines presentation but facilitates the conversion of XML documents to the required formats for publication (through XSL transformations, or XSLT).

All formally marked up documents include *metadata*, a structured form of resource description. Publishers generally include certain types of bibliographic metadata—for example, creator, title, publication date, description, and keywords—some or all of which may be displayed within (or alongside) their publications. Similar metadata may be kept in a separate database and shared with libraries and other content providers. Such metadata, which is a lot like the information that once resided in a library’s card catalog, can facilitate discovery by search engines.

More specific categories of metadata may be appropriate depending on the publication. For example, description of file format, language(s) of publication, and security level (to specify who can access the content) may be encoded in a document as needed. With the help of software, metadata can be derived, or extracted, from the markup of a publication before, during, and after production, for a variety of purposes. For example, metadata about a document’s revision status can help track the versions of a document through the proofreading and testing stages. Other types of metadata—for example, the DOIs of journal articles in a reference list—can facilitate linking of source citations to information about the location of the texts to which they refer. For more information on metadata, including information about best practices, consult the website of the [Dublin Core Metadata Initiative](#), an organization responsible for developing standards related to electronic resource description.

XML

XML, an abbreviation for extensible markup language, is a detailed specification for defining and marking up the logical structure of a document—in other words, it is a metalanguage that provides rules for naming and defining the parts of a document and their relationship to each other. It has been in development since the mid-1990s by various working groups of the World Wide Web Consortium (W3C). XML is an open standard, meaning it is free to the general public (and available through [W3C](#)).

XML markup requires that each element in a document, including the document as a whole, be properly delineated—usually by a pair of opening and closing tags. XML does not contain a standard set of tags; rather, it allows users to create unique tags customized for any type of document: chapters in a book, journal articles, catalog entries, database records, and so forth. For an example of how XML is applied, see [figures 1 and 2](#).

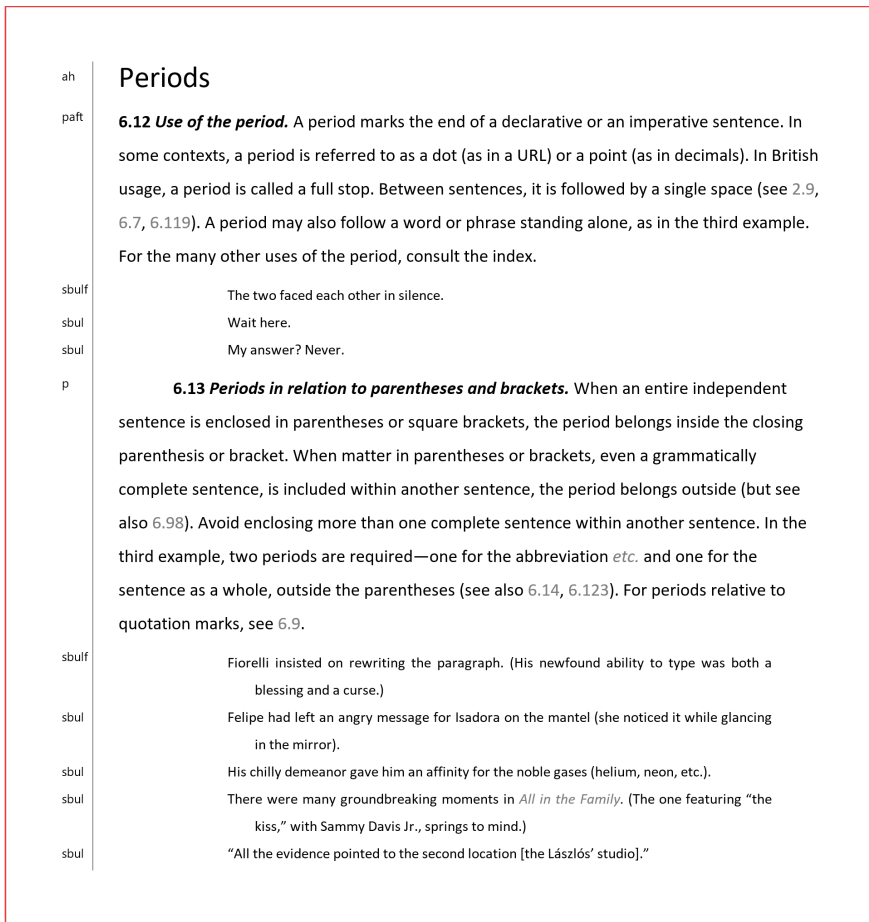


Figure 1. A portion of the Microsoft Word manuscript for chapter 6 of *The Chicago Manual of Style*. Paragraph-level style names are displayed in the left margin; character-level styles, applied for paragraph number and title, cross-references, and words intended to display in italics, are distinguished graphically. Compare figure 2.

```

<asection id="a06012">
<ah><xref idref="ra06012">Periods</xref></ah>
<psection id="pn06012">
<paft><pnum id="pnum06012">6.12</pnum> <hemb>Use of the period.</hemb>
A period marks the end of a declarative or an imperative sentence.
In some contexts, a period is referred to as a dot (as in a URL) or a
point (as in decimals). In British usage, a period is called a full
stop. Between sentences, it is followed by a single space (see <xref
idref="pn02009">2.9</xref>, <xref idref="pn06007">6.7</xref>, <xref
idref="pn06119">6.119</xref>). A period may also follow a word or phrase
standing alone, as in the third example. For the many other uses of the
period, consult the index.</paft>
<sidebar id="pn06012:sb01">
<list>
<sbulf>The two faced each other in silence.</sbulf>
<sbul>Wait here.</sbul>
<sbul>My answer? Never.</sbul>
</list>
</sidebar>
</psection>
<psection id="pn06013">
<p><pnum id="pnum06013">6.13</pnum> <hemb>Periods in relation to
parentheses and brackets.</hemb> When an entire independent sentence is
enclosed in parentheses or square brackets, the period belongs inside
the closing parenthesis or bracket. When matter in parentheses or
brackets, even a grammatically complete sentence, is included within
another sentence, the period belongs outside (but see also <xref
idref="pn06098">6.98</xref>). Avoid enclosing more than one complete
sentence within another sentence. In the third example, two periods
are required&#x2014;one for the abbreviation <i>etc.</i> and one for
the sentence as a whole, outside the parentheses (see also <xref
idref="pn06014">6.14</xref>, <xref idref="pn06123">6.123</xref>). For
periods relative to quotation marks, see <xref idref="pn06009">6.9</
xref>.</p>
<sidebar id="pn06013:sb01">
<list>
<sbulf>Fiorelli insisted on rewriting the paragraph. (His newfound ability
to type was both a blessing and a curse.)</sbulf>
<sbul>Felipe had left an angry message for Isadora on the mantel (she
noticed it while glancing in the mirror).</sbul>
<sbul>His chilly demeanor gave him an affinity for the noble gases (helium,
neon, etc.).</sbul>
<sbul>There were many groundbreaking moments in <i>All in the Family</i>.
(The one featuring &#x201c;the kiss,&#x201d; with Sammy Davis Jr., springs
to mind.)</sbul>
<sbul>&#x201c;All the evidence pointed to the second location [the
L&#xel;szl&#xf3;s&#x2019; studio].&#x201d;</sbul>
</list>
</sidebar>
</psection>
. . .
</asection>

```

Figure 2. Another view of figure 1, as plain text after the manuscript has been exported and converted to structured XML.

XML traces its origins to SGML, which stands for standard generalized markup language. SGML became an international standard in 1986, defined by the International Organization for Standardization as ISO 8879, after which it became the primary specification according to which many publishers marked up and maintained electronic versions of their documents. Such a standard was originally developed to ensure that governments and other organizations could maintain electronic archives that would be relatively impervious to changes in technology.

Like XML, SGML is a metalanguage—in the case of SGML, a comprehensive set of rules not just for naming and defining the parts of a document but for developing other, related markup languages. One such language is HTML (hypertext markup language), which was introduced in the early 1990s as a set of tags designed to facilitate web presentations. HTML tags (often in conjunction with a style sheet) work much like typesetting codes for print publications, determining how a document will appear, not on the page, but on the screen.

XML was initially developed as a simplified version of SGML that would facilitate conversion to HTML for publishing on the web. As electronic formats have proliferated, many publishers have found it advantageous to use XML as the basis of their publications in order to make them available for other purposes, such as electronic archives and various e-book formats. Like SGML but unlike HTML, XML can be adapted to describe any type of document and does not depend on any specific software or format.

This flexibility can facilitate the presentation of content in ways that were not possible at initial publication—a hedge against changing technologies. For example, in the early 1990s, when HTML for tables did not yet exist, tables on the web were presented only as static images rather than as text. Publishers who at that time were using SGML tags to arrange data cells into rows and columns were prepared to regenerate full-text HTML tables in their web publications as this functionality became available in later versions of HTML. A similar phenomenon is taking place for mathematical expressions marked up in MathML (mathematical markup language, an application of XML), which has only recently gained wider support across web browsers and other applications, including applications and devices for e-books.

An XML document, simply stated, consists of elements and attributes. An *element* is a specific part of the document—in other words, any part that has been labeled with a pair of XML tags. Elements might include such items as document title, section headings, paragraphs, titles of works, terms marked for emphasis (and destined to be italicized or otherwise distinguished from the surrounding text), names of authors, and cross-references. An XML *attribute* is included inside the opening tag for an XML element and provides additional information about that element, expressed as a *value*. For example, a document might include two types of lists, as specified by an attribute value for each type. A style sheet would recognize these values and format the lists accordingly—for example, assigning arabic numerals starting at 1 for the items in one of the list types and bullet points for the other. Another type of attribute might specify a URL as the value for a hypertext reference to facilitate the creation of a link.

In order to facilitate the use of special characters that may not be available in a particular application or from a particular keyboard or other input device, an XML document may also include plain-text placeholders known as *character references*. Each character reference is set off from the surrounding text by an ampersand and a semicolon. By default, XML supports character references defined by the Unicode standard. For example, a multiplication sign (defined by Unicode as U+00D7) can be represented in XML as the character reference `×`; (the `#x` signals the hexadecimal Unicode code point). In a program that interprets the XML document, the character reference would display as a multiplication sign (×). Alternatively, character references can consist of names taken from other ISO standard character sets—for example, `´`; or `&alpha`; (which resolve to `é` and `α`, respectively)—when these are specified in the DTD. In general, publications in electronic formats should display the proper Unicode characters, which are required by EPUB and supported in many other formats. For more on Unicode, consult the website of the [Unicode Consortium](#).

Design

In the most basic sense, *markup* describes the structure and content of a work, whereas a *design* determines how it looks—either on the printed page or on a screen. Whatever method has been used to mark up an edited manuscript—generic labels (on paper or in an electronic file), software-generated styles, or XML—each element must be mapped to a particular design specification. In some cases, editors work with designers to identify the design elements before a manuscript is submitted for production. For projects that are intended for publication in more than one format, a different design will be necessary for each. If XML or other formal markup is used, the design specifications are usually encoded for production in electronic style sheets.

CSS and other style sheet languages provide a means of rendering a formally marked up document according to a detailed set of design specifications. But style sheets cannot always control for certain factors that might affect the published document's appearance in print, such as line endings and page breaks. Style sheets for print are therefore usually supplemented with a publisher's rules for composition and page makeup (see [fig. 3](#)). These rules direct typesetters to make small adjustments in a page-layout program to fix such things as loose lines and uneven facing pages and to reposition images to fit the available space. Additional adjustments may be necessary to achieve precise alignment in tables or other textual environments with complex layouts. Such tweaks made in the files for print will not be reflected in the style sheets or tagged source files used for electronic formats. For formats with reflowable text, however, publishers have come to accept the inevitable variations in spacing and other aspects of presentation that might depend on a particular application or device and that cannot be controlled for in a style sheet. Where precise layout is mandatory, PDF or another fixed-layout format may be necessary.

House Style for Composition and Page Makeup

The University of Chicago Press, 1427 E. 60th Street, Chicago IL 60637

(Unless otherwise specified, these standing specifications apply.)

TYPOGRAPHY

1. **Vertical spacing** is specified as baseline to baseline.
2. **Letter spacing:** Text and smaller type are to be set at the standard character fit designed for the typeface unless specified otherwise.
3. **Word spacing:** Tight horizontal word space is the preference with no extra space after punctuation. Word space should average the width of a lowercase n, and be no tighter than the width of a lowercase i.
4. **Line breaks and hyphenation:**
 - ▶ Standards call for a minimum of two characters at the end of the line, with a minimum of three carried down.
 - ▶ No more than three consecutive end-of-line hyphens or punctuation (or a combination of hyphens and punctuation) are allowed.
 - ▶ Up to eight hyphens are allowed per 35-to-40-line page.
 - ▶ Hyphenated compound words should be broken only at the hyphen.
 - ▶ The final word of a paragraph is allowed to hyphenate except that a minimum of four characters (not counting periods, commas, and quotation marks) is required on the final line.
 - ▶ It is best not to have a hyphenated word as the last word on a page. However, it is allowable if the syllable or syllables leave little doubt as to the identity of the full word.
 - ▶ It is preferable not to break proper names but allowable when difficult to avoid. Line breaks are allowed after a first name or middle initial; names using two initials will not break between the initials.
 - ▶ More guidelines on word division are found in the CHICAGO MANUAL OF STYLE. Follow it for more detailed guidance.
5. **"Carding":** Adding extra space between lines or paragraphs to make facing pages align is not allowed.

Figure 3. Sample set of rules for composition and page makeup.

A complete design for a book or other work that will be published in print should include full type specifications and representative layouts for all the various categories of text. Common elements might include the front (or preliminary) matter; a chapter-opening page; two or more facing pages showing primary textual elements such as running text, extracts, subheads, footnotes, illustrations, running heads, and folios; and the back (or end) matter (including, as relevant, the appendix, endnotes, glossary, bibliography, and index). See figure 4 for an example of a partial text design for a printed book showing marginal type specifications. Before typesetting, the final, edited manuscript is checked by the editor against the design specifications to make sure that each element has been properly accounted for. If an electronic style sheet will be used to encode these specifications, it must include every element in every context that might require a different specification.

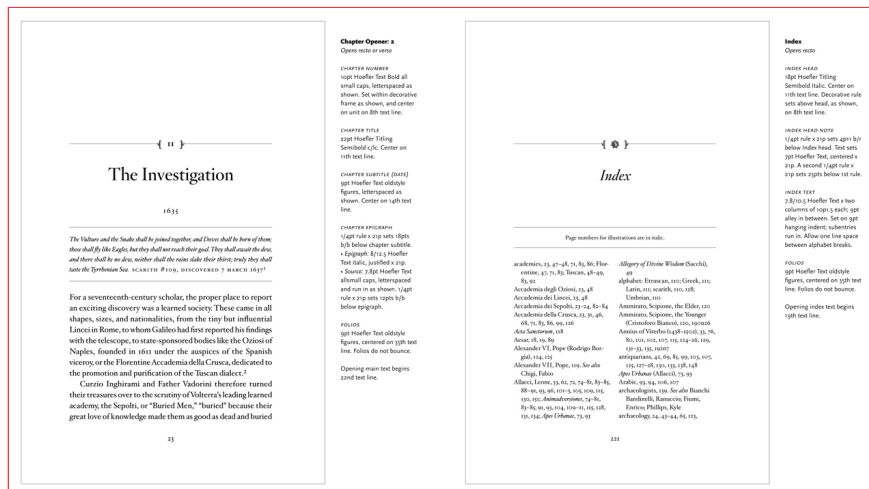


Figure 4. Sample design specifications for a book, showing a chapter-opening page and a page for an index.

Design specifications for electronic formats must take into account the applications and devices that will be used to view them. HTML for web pages and related formats and, more specifically, EPUB for e-book formats already determine much of the presentation based on markup. Style sheets can override and augment these default behaviors to a greater or lesser degree. For e-book formats especially, the overall look and feel is typically determined by the device or application used to read the book, and publishers may be limited in what they can or should do to override default settings. For publications offered as a custom web presentation or app, the design must take into account such things as navigational elements, additional copyright statements, and help screens.

For printed books, publishers will often specify a target page count (usually for a specific, often standard, trim size—e.g., 6 × 9 inches), taking into account the length of the manuscript. Designers must figure out type sizes, spacing, and margins that will result in a book that comes close to this target. The traditional way to determine this is to start by getting a character count, or castoff, and dividing this count by the target number of characters per printed page. Since each category of text will likely be given its own set of type characteristics, the designer needs not just a total character count for the work but a count broken down by type of material (e.g., text, extracts, appendix, notes, and bibliography). Some designers will prefer instead to look at the entire manuscript in a page-layout program, where decisions about type sizes and layout can be made and the resulting page count determined at the same time.

For the most accurate result, it is best to use the final, edited manuscript. Whatever method is used, it is important also to account for illustrations, tables, and other material presented separately from the text. For printed journals, which have fixed designs and page allotments, the total number of articles must be adjusted to fill each issue. For electronic formats, length is less of a consideration from a production standpoint, though it is always an editorial and design consideration.

Type sizes and line spacing for printed works are measured in *points*. The dimensions of the type page—the area occupied by the running head, the text proper, the footnotes (if any), and the page number (or folio)—are measured in *picas*. A pica is approximately one-sixth of an inch; there are twelve points in a pica. Type specifications for the screen may be given in the same dimensions used for print; alternatively, *pixels* may be specified for type sizes and line spacing. (A pixel is the smallest element making up an image on the screen. Earlier generations of computer monitors typically had a resolution of 72 pixels per inch, measured horizontally or vertically, such that a pixel was equal to a point. Modern screens display as many as 200 ppi or more.) Another option is to specify type size and line spacing in relative terms—for example, as a percentage of a web browser’s default size for a particular category of type. All of these dimensions have been incorporated into the language of electronic style sheets such as CSS.

Sometimes designers will specify a particular space between two elements—for example, between single and double quotation marks, where a thin space may be called for. Such spaces are usually defined relative to the size of the type in question. Some of the more common types of spaces—starting with the em space, on which all of them are based, and ending with the regular Space bar space (and its nonbreaking counterpart)—are as follows:

Unit	Definition	Unicode number
em space	the width of a capital <i>M</i> in a given font at a given size	2003
en space	half the size of an em	2002
thick space	one-third of an em	2004
medium (or mid) space	one-fourth of an em	2005
thin space	a fifth (or sometimes a sixth) of an em	2009
hair space	thinner than a thin space	200A
regular space (achieved with Space bar)	in unjustified text, same as a medium space (or, in some fonts, a thick space)	0020
nonbreaking space	same as regular space, but does not break over a line	00A0

All of these distinctions can be achieved for the printed page in page-layout or typesetting programs. They can also be specified in a limited way in electronic style sheets—for example, as an instruction to increase (or decrease) the spacing

between all the letters or words in a particular type of paragraph. To date, however, though these spaces have been defined for Unicode, they are not always supported by the devices or applications used to display electronic publication formats. The widely supported nonbreaking space, on the other hand, is an essential tool in preventing unwanted end-of-line breaks between words or other elements in any format.

The Electronic Workflow

Introduction

Manuscripts prepared using Microsoft Word or the like can be formatted for print or converted to a variety of electronic formats using the same application. But most publishers prefer to import word-processed manuscripts into a dedicated publishing program such as Adobe InDesign or QuarkXPress to take advantage of more sophisticated tools for refining layout and typography for the printed page. Such conversions have typically involved mapping generic typesetting instructions or word-processing styles to corresponding styles or codes used by the publishing software. Some of these same programs can now use XML tags in addition to their own systems of markup, and manuscripts can be converted to XML in order to take advantage of this downstream flexibility and to facilitate publication in multiple formats.

Publishers who introduce XML into the process at an early stage can make it the basis of their entire production workflow. Such a workflow, which is determined by the principles of markup discussed in the previous sections, might be summarized very briefly as follows:

1. XML tags are added to a manuscript in conformance with a DTD. Such tags can be derived from word-processing styles using a conversion utility after the manuscript-editing stage. In this case, manuscript editors can facilitate the process by working in a template that determines which styles to use in which contexts.
2. Meanwhile, any artwork is scanned and converted as necessary to the recommended file formats and resolutions for publication in print and electronic formats. XML tags that indicate the correct versions of each publication-ready file must be included in or added to the manuscript. These can be derived from consistently formatted figure-placement callouts in the edited manuscript.
3. The XML-tagged file—that is, the marked-up manuscript from step 1—can be converted (or “transformed,” using XSLT) to a variety of formats and, in conjunction with the artwork from step 2, typeset for print or published in one or more electronic formats (e.g., HTML or EPUB). In each case, a particular style sheet determines what the presentation will look like.

For an illustration of such a workflow, see [figure 5](#).

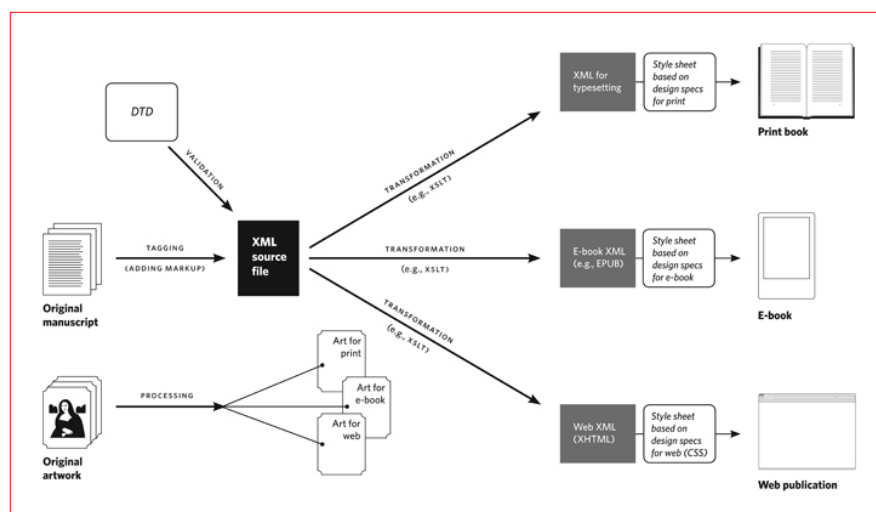


Figure 5. A simplified XML workflow, in which an XML-tagged book manuscript is used as the basis of print, e-book, and web versions of the publication. Any artwork is processed and converted as appropriate and integrated into each version before publication.

In any workflow, the final output—whether print or electronic—needs to be proofread, as detailed in chapter 2 of *The Chicago Manual of Style*. For simultaneous print and electronic publication, there are a number of options for doing this. Especially for longer works, it can be expedient to proofread the work as it is typeset for print first. Corrections made during the proofreading process are incorporated into the XML file used for typesetting. Then, the corrected XML (including any index produced during the proofreading stage) can be converted to HTML, e-book formats, and any other electronic version using a range of style sheets and conversion programs. This saves the trouble of having to proofread print and electronic versions from scratch simultaneously, though the electronic versions still need to be carefully reviewed and tested. For publications that routinely include supplemental materials not available in print (the case for many journal articles in the sciences), it may be expedient to proofread the content of the print and electronic versions simultaneously (after collating the material from each medium into a single PDF or other document for review).

The workflow described above, developed for books and journal articles published simultaneously in print and electronic formats, is essentially linear. Web-based publications and apps, which require the development and implementation of functional and navigational features, generally follow a different model—one that allows for design, engineering, testing, release, and maintenance of the website or app itself, in all of the contexts in which it will be used. Such workflows tend to be cyclical, allowing for the refinement of features to improve usability and optimize presentation. Editors, designers, and other publishing professionals who become involved in producing web-based publications or apps should be prepared to adapt to the processes and requirements of that environment.

When to Introduce XML

To date, few editors and even fewer authors work directly with XML markup. Most of the coding that goes into such markup—not to mention the DTDs, style sheets, and converters that are used to implement it—requires an immersion in software and programming that is generally the arena of IT specialists. Thus, it usually falls to the publisher or one of its vendors to apply the XML. This can be done at various points in the production process, each with its benefits and costs, as discussed below. Meanwhile, IT specialists are also typically responsible for developing DTDs (or ensuring compliance with existing or standard DTDs), writing style sheets (or making modifications to existing style sheets), and coordinating the systems that allow a publisher to maintain an archive of the latest versions of the electronic files used to publish every version of every work in its inventory.

Though few publishers do so, there are advantages to introducing XML markup to a manuscript before it is edited. Manuscript editors are in many ways closest to the content; they are therefore in a good position to ensure the quality of the markup. A publisher may invest in XML tools that facilitate traditional and new editing tasks by allowing editors who have little knowledge of XML syntax to modify content and structure without invalidating the underlying code. Such an approach is especially suited to a journal publisher, where each article can be marked up in essentially the same manner and according to the same DTD and style sheet for each issue. Book publishers, on the other hand, might develop sets of templates that correspond to different types of books.

Many publishers instead provide their manuscript editors (and, in some cases, their authors) with word-processing style templates that map to XML tags for a specific document type such as a book or journal article. Customized add-ins can help editors apply such styles and can enforce certain markup rules. The advantage of such an approach is that most editors are already familiar with the word-processing environment. Conversion to XML based on the final, edited manuscript, the components of which have been carefully scrutinized by the editor, can help to ensure consistent and accurate results. One primary disadvantage is that structural identifiers for chapters and the like, which are not available per se as word-processing styles, must be anticipated by the conversion (sometimes with the aid of textual signposts placed in the manuscript). Likewise, attributes to facilitate linking or for cross-references must be added, sometimes with the help of pattern matching.

Some publishers add XML later in the process, after a manuscript has been typeset for print. There are two scenarios for doing this: (1) the typesetter's electronic files are used to create the XML (for example, by deriving XML from a program such as Adobe InDesign); or (2) XML is derived from PDF files. The latter approach is recommended only when there are

no other suitable electronic files—for example, for an older work available only in print, from which pages have been scanned and converted to electronic text.

Many book publishers forgo XML entirely and send the files used to produce print to vendors who in turn produce e-book formats. Others choose to do this in-house. EPUB, which can serve as the basis of a number of other e-book formats, can be exported directly from print-ready InDesign or QuarkXPress files using tools native to those applications. Though such an export technically depends on producing the XML-compliant HTML (and related files) required for EPUB, publishers who use such tools do not need to worry about incorporating XML explicitly into their workflows.

Source Files

XML source files comprise the XML text files and any files for illustrations or audiovisual materials that, in conjunction with style sheets, converters, and an array of software programs, can be rendered in published form in a variety of media. As the basis of an electronic workflow, these files should accurately reflect the *content* of a published work, including any subsequent changes thereto. For example, if it is necessary to correct an error in the text of an HTML version of a journal article or a chapter in a book, it is not enough to make the correction in the HTML file; the correction must also be made in the corresponding XML source file to ensure that it will be reflected in any future versions of the work generated from that source. If extensive corrections are required, it may be better to make them in the source files and then reconvert them to HTML. Any changes to the way the content is presented, on the other hand, are usually made in a style sheet or, for some types of adjustments in printed works, with a page-layout or typesetting program.

Files for such nontext items as illustrations or audiovisual materials may be needed in various sizes or formats (thumbnails, medium resolution, high resolution, low bit rate, high bit rate, etc.) to allow for optimal presentation in various media. All of these versions should be derived either from high-resolution bitmapped graphics in a lossless format such as TIFF (e.g., for photographic images) or from vector-based graphics (e.g., for line art created in a program such as Adobe Illustrator). The richer the source, the higher the level of quality not only in print but on the screen, where image files are typically down-sampled (i.e., prepared at a lower resolution) and compressed for faster downloading and rendering. It should be noted that once information is discarded from a file (e.g., using a lossy compression format such as JPEG), it cannot be recovered. Though resolution requirements for the screen are typically lower than those for print, the original, high-quality source files should always be retained, as they are more likely to meet requirements for republishing the work in alternative formats in the future. Many publishers provide authors with guidelines for preparing and submitting artwork; Chicago makes [detailed guidelines](#) available on its books and journals websites.

Tables are generally marked up and included as part of the XML source files with the rest of the text. The HTML tables derived from this markup are sometimes supplemented by image files or PDF versions of the same (often derived from pages typeset for print). Though not strictly necessary, this can be desirable because, to date, HTML table markup does not include layout options sufficient to achieve all of the typesetting niceties that may be required, such as certain types of character alignment within a column. Meanwhile, the character data in the rough-hewn HTML tables derived from XML can be useful—for example, the data can be cut and pasted or otherwise exported into spreadsheets and other tools for efficient verification of results. It is important to keep track of each version of a table: any correction must not only be made in the XML source file but also reflected in any published version of the table.

Mathematical expressions may be tagged using MathML, which can allow it to be rendered either directly from MathML or from XML-based HTML. Though browser support for MathML continues to improve (and HTML and EPUB now both include support for MathML), math in web publications and e-book formats is still often presented using bitmapped images derived from typeset pages. (Alternatively, the images can be derived from math-editing tools in word-processing applications such as Microsoft Word or LaTeX.) As with tables, it is important to make sure that the MathML in the source file reflects any changes or adjustments made to the embedded images.

Options for Presenting Content

Print remains an important medium for many types of publications. From the standpoint of a typical publisher's electronic workflow, however, it has become one option among several. (Print-related technologies, including considerations for

paper selection and binding, are discussed [in the next section](#).)

For presenting publications on the web, two standards have emerged: full-text web presentations and PDF. Full-text presentations, usually in HTML, are especially common for journals, particularly those in the scientific, technical, and medical fields (so-called STM publications), and for reference works. By taking advantage of all the textual and audiovisual functionality of modern browsers, full-text presentations are extremely flexible. (Standalone apps offer a similar degree of flexibility, often tailored for a specific device.) On the other hand, such productions are costly and can require a lot of maintenance simply to adapt to changing technologies and evolving standards.

PDF, because it presents fixed layouts and typefaces that match the printed page, is easier to produce and maintain, especially as a supplement to print. In fact, printed publications generally must be submitted to the printer as PDF files, and optimizing these files for electronic publication is a relatively straightforward process (e.g., using a PDF-creation program such as Adobe Acrobat). Journals that publish on the web typically provide PDF versions of their articles accompanied either by full-text HTML or, to facilitate search engines and other services, HTML metadata (often accompanied by an abstract).

Most major book publishers now publish their content in print and e-book formats simultaneously. At the very least, PDF, the basis of print, can also be used as the basis of an e-book. But in part because PDF does not always translate well to smaller screens, many publishers offer dedicated e-book formats instead of or in addition to PDF. EPUB, which is supported by a number of e-book applications and the devices that run them, can also serve as the basis of conversion to other e-book formats, including commercial formats supported by Amazon and others. To manage the need for multiple formats across a proliferation of devices and applications, many publishers have turned to third-party services to prepare proprietary versions of their e-books and distribute them to booksellers and other content providers, whether or not they produce EPUB files in-house.

Publishers also have the option of offering book chapters, journal articles, and other components of a larger publication separately from the publication in which they originally appeared. This has always been true of print publications (which have long offered what are called offprints), but digital technology has greatly expanded the possibilities for such derivative works. For example, a medical publisher might group together all the articles that, according to their metadata, are about Tamiflu-resistant flu strains and sell them as a new publication. Such “virtual” publications can be assembled automatically from a publisher’s source files, or they can be curated by content experts. On the other hand, individual journal articles may be published in advance of, or independently of, the print publication. This approach facilitates the way readers tend to find articles: thanks to search engines and citation-linking networks such as CrossRef, much of the traffic coming into journal sites goes directly to individual articles.

To ensure that their content is accessible to people with disabilities, publishers will want to consider adding appropriate text alternatives for materials that require them. For illustrations, properly marked up text and captions will enhance accessibility; if an illustration is insufficiently described in either of these locations, alternative text will enable readers who use text-to-speech and related tools understand the nature and content of the illustration. For tables and math, proper markup is the best tool for ensuring the accessibility of the content. If tables or math must be presented as images, however, or for an especially complex table or expression, alternative text may be needed. Editors or authors would ordinarily be in the best position to decide what text is needed and how to write it, though not without recourse to detailed guidelines. For more information, see the [DIAGRAM Center](#), a Benetech literacy initiative; see also the [Web Accessibility Initiative](#) (WAI) of the World Wide Web Consortium (W3C) and, for EPUB, the [Accessibility Guidelines](#) provided by the International Digital Publishing Forum.

Publishers who offer their content in electronic form face challenges from those who might attempt to copy and redistribute this content in violation of copyright law. To ensure that copyright holders and publishers are fairly compensated for their intellectual property, on the one hand, while providing content to readers in an easily accessible manner, on the other, publishers can employ a number of strategies. These strategies—which include technologies such as encrypted passwords, user-authentication schemes, and digital watermarks—are collectively referred to as digital rights management (DRM).

Essentially, DRM can be seen as the intersection between copyright law, distribution models, and technology. Some of the legal aspects of electronic distribution and licensing are discussed in chapter 4 of *The Chicago Manual of Style*. The technological aspects of DRM will depend on the software used to implement it and on the medium of publication. For e-books, these considerations are usually incorporated into the agreements publishers make with various vendors and distributors, from Amazon and Apple to digital content providers for libraries, and implemented specifically for the applications or devices used to consume the publisher's content.

Print Technologies

Introduction

The principles of electronic markup occupy a central place in the production of publications intended for print. XML derived from the manuscript or, later, from the typeset files, **as outlined above**, can help publishers provide such content in a variety of electronic formats in addition to print. Some type of structured markup is essential in making even print publications visible to search engines and available to libraries and other content providers. The difference is that for print, the end product is physical. Whereas the appearance of an electronic publication might vary depending on the device and software used to view it, a printed book or journal is a permanent artifact with a fixed design.

To date, most publishers rely on PDF (portable document format) as the source of their printed publications. PDF presents fonts, images, layout, and pagination exactly as they will appear when printed. This makes PDF an essential tool in the proofing process. After a book or article is typeset, first proof is generated as PDF and supplied to authors, editors, and indexers either as a printout or for on-screen review. One or more rounds of revised proof can be offered in the same manner. Moreover, the files used by the printer are typically the same as those generated by the typesetter. Nonetheless, the PDF files themselves must be print-ready. Among other things, it is important to check that the latest versions of files are present; that they are properly and consistently named; that any specialized fonts have been embedded or otherwise included; that any images have been provided as high-resolution files appropriate for printing; and that all specifications regarding paper, trim size, resolution, and so forth have been provided.

Once the final files have been received by the printer, decisions related to paper, printing methods, and options for binding come into play.

Paper

Paper comes in varying weights, sizes, shades, coatings, and degrees of opacity and smoothness, or finish. In consultation with the printer or a paper merchant, the publisher must determine which type of paper best suits a particular publication and which will print on and run through a given printing press most efficiently. Other considerations include cost, availability, and durability of the paper. Publications printed on acid-free paper have a longer life expectancy than those that are not, and they may carry a notice on the copyright page indicating their compliance with the durability standards of the American National Standards Institute or another such body.

Paper is manufactured to standard roll and sheet sizes. Because printing presses and bindery equipment are set up to accommodate these roll and sheet sizes with minimal waste of paper, publishers usually find it most economical to choose one of a handful of corresponding trim sizes for their books and journals—in the United States, these are 5½ × 8½ inches, 6 × 9 inches, 7 × 10 inches, and 8½ × 11 inches. Publications requiring a nonstandard trim size will generally cost more. Note that the dimensions for the type page (that part of the page occupied by the text, the running head, and the folio, if any) must leave adequate margins for the given trim size and, if required, allow for illustrations to bleed.

Environmental issues have come to play a significant role in paper selection. Most works can be printed on recycled paper, generally a combination of virgin fiber and pre- and postconsumer wastepaper. Using papers certified by the Forest Stewardship Council (FSC) is another way to minimize environmental impacts. FSC is a nonprofit organization widely regarded as ensuring the best practices in forest management. Publishers who use recycled and FSC-certified papers not only contribute to the reduction of greenhouse gas emissions but also minimize the negative impact on endangered forests and forest-dependent communities.

Also important are the bleaching methods used to make papers, which employ varying amounts of chlorine and chlorine derivatives, the use of which contributes to the formation of dioxins and other hazardous substances. Processed chlorine-free (PCF) papers are recycled papers that have been produced with no chlorine or derivatives beyond what may have been used originally to produce the recovered wastepaper. A totally chlorine-free (TCF) process can be used on virgin fibers. Elemental chlorine-free (ECF) and enhanced ECF papers are bleached using a chlorine derivative (chlorine dioxide) in order to minimize hazardous by-products but are less safe than TCF or PCF papers. Publishers wishing to learn more about the environmental impact of paper purchasing should seek the advice of the [Forest Stewardship Council](#). Additional resources include the US-based [Green Press Initiative](#) and the Canadian-based [Canopy](#).

Printing

At least for larger print runs (more than a few hundred copies), the most common method for producing books continues to be offset printing, or offset lithography. This process involves the transfer of images (text, illustrations, and any other marks that will be distinct from the background color of the page) from metal plates to paper through an intermediate cylinder. The images are usually imposed directly onto the photosensitive plates using the typesetter's electronic files (referred to as computer-to-plate technology, or CTP) or, using an older process, by contact with film negatives (also typically generated from the typesetter's files). Ink is applied to each plate, and the inked images are offset onto the paper through the rubber-blanketed intermediate cylinder (see [fig. 6](#)). The printing press itself may be either sheet-fed, using sheets of paper that have been pre-cut, or web-fed, using rolls of paper that will be folded and trimmed at the end of the printing stage.

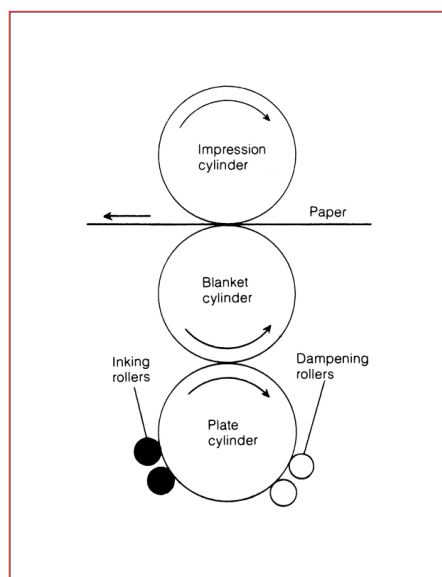


Figure 6. Principle of offset printing. A plate is wrapped around and fastened to the plate cylinder. Water applied by the dampening rollers adheres only to the background area of the plate; ink applied by the inking rollers adheres only to the dry image of the type on the plate. As the plate cylinder revolves, it transfers the ink to the rubber blanket of the blanket cylinder, which in turn transfers (or offsets) it onto the paper, which is held in place by the impression cylinder.

In digital printing, images are printed directly onto paper through ink jets or thermal transfer using either powder- or liquid-based toners. The quality of the reproduction is typically not as high as that achieved through the offset process (though the gap has closed), but digital printing makes it economically viable to print small quantities of a publication on demand (as few as one), preferably from archived electronic files, and can reduce the publisher's cost for warehousing unsold stock. Digital printing also makes it feasible to customize each copy. Short-run digital copies of books can be bound either as hardcover or paperback editions.

Before binding can begin, the press sheets, or printed sheets, that emerge from the printing press must be folded and trimmed. A press sheet has printed pages on both sides, and when it is folded in half, and then in half again—continuing until only one page is showing—all the pages fall into proper sequence in a process known as imposition (see [fig. 7](#)). The

folded sheet, called a signature, usually consists of thirty-two pages, but this number may vary depending on the bulk and flexibility of the paper and the size of the offset printing press. When all the signatures have been gathered in the proper order, they are referred to as folded-and-gathered sheets, or F&Gs. The F&Gs can then be bound in either hardcover or paperback format. (Ideally, publishers review the first set of F&Gs for completeness, accuracy, and proper order.)

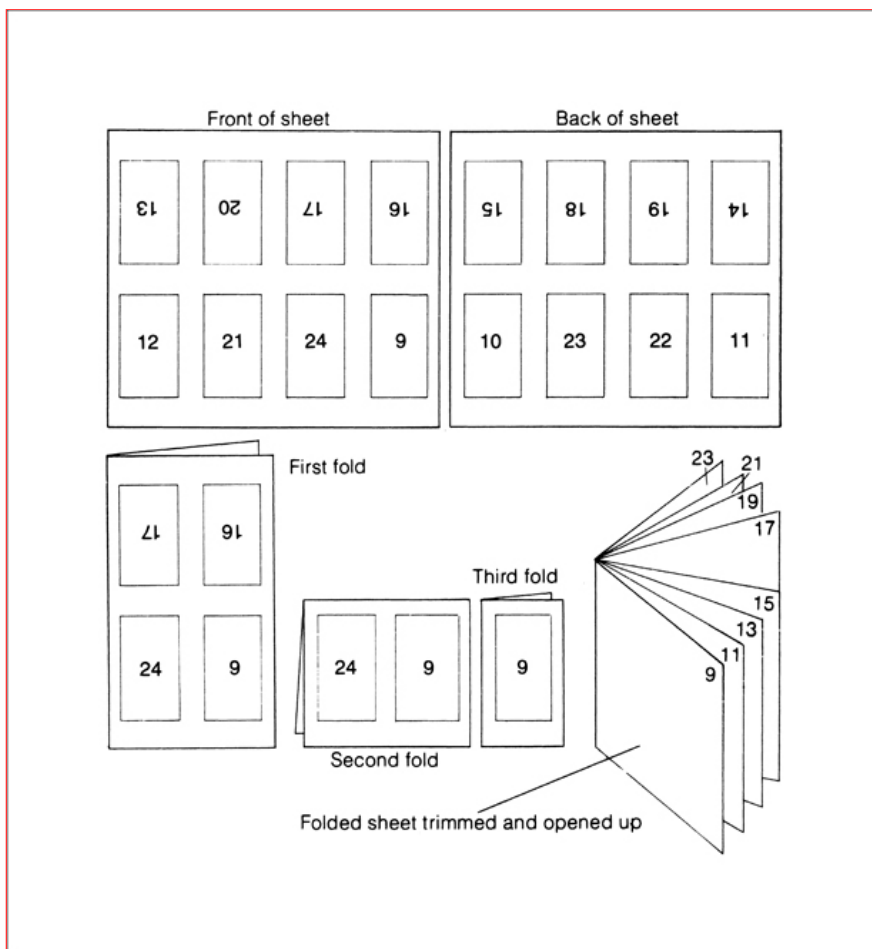


Figure 7. A sheet consisting of sixteen printed pages. After being folded, the pages fall into proper numerical sequence.

Binding

Binding hardcover books typically requires sewing the signatures together, usually by either Smyth sewing or side sewing (see fig. 8). An alternative method, adhesive binding, involves notching or fraying the folded edges and then applying adhesive to the signatures to hold them together. Smyth- or side-sewn books may have a sturdier binding and hold up better over time, but adhesive binding, which is faster and less costly, is often just as strong, thanks to improved polyurethane-based adhesives. Meanwhile, the hardcover case is fashioned by the application of cover material (such as cloth, synthetic fabric, leather, or paper) to boards. The case is then affixed to the body of the book through the application of glue to the endpapers, and a dust jacket may be wrapped around the case. (An alternative to the dust jacket is provided by the paper-over-board format, which allows full-color images and type to appear directly on the hard outer cover—including on the inside front and back panels.)

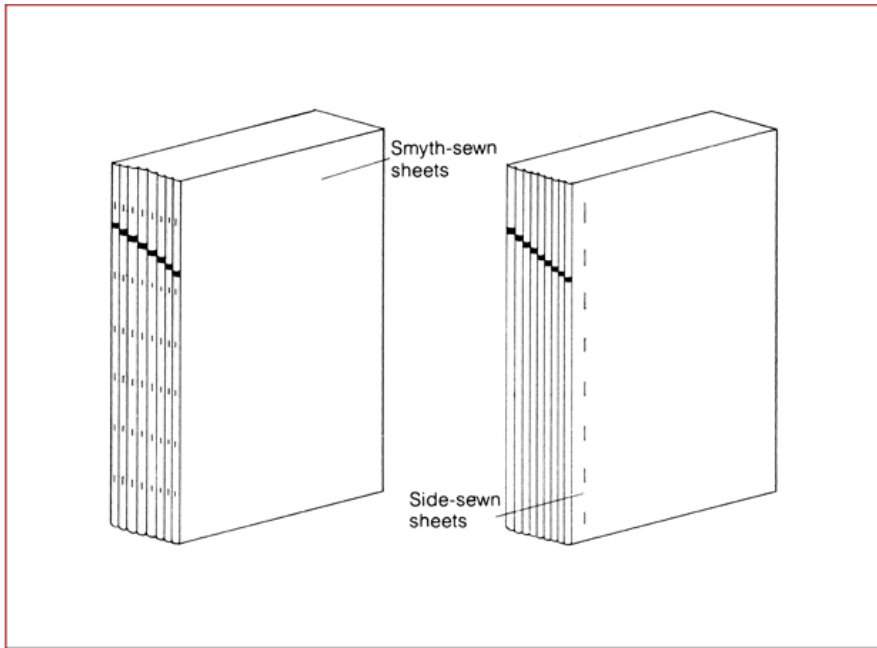


Figure 8. Two methods of sewing used in binding. In Smyth sewing the sheets are stitched individually through the fold; in side sewing they are stitched from the side, close to the spine. The black rectangles printed on the folds in both methods help the binder recognize whether a signature is missing, duplicated, out of order, or upside down.

Paperback books are almost always adhesive-bound, through one of three methods: perfect, notch, or burst binding (see [fig. 9](#)). In the perfect-binding method, about an eighth of an inch is mechanically roughened off the spine of the tightly gathered F&Gs, reducing them to a series of separate pages. The roughened spine is then coated with a flexible glue, and a paper cover is wrapped around the pages. In the other two methods, the spine is either scored by a series of notches (notch binding) or perforated (burst binding) and then force-fed with glue. Unlike perfect binding, these methods prevent the loss of part of the back margin and ensure that signatures remain intact, reducing the risk of pages coming loose. For paperbound books of higher quality, the signatures can be sewn and the covers (sometimes with flaps) then affixed, as with adhesive-bound books; this style of binding is known as flexibinding or limp binding.

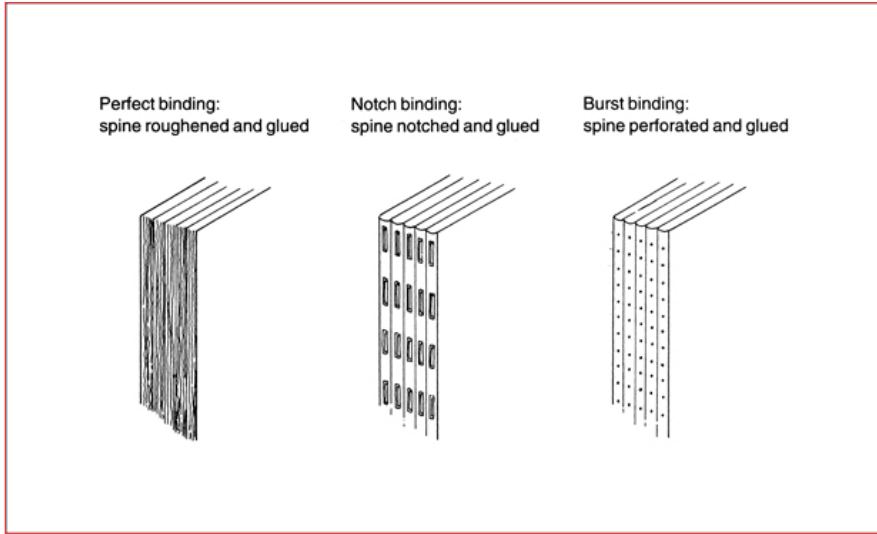


Figure 9. Three methods of adhesive binding: perfect, notch, and burst binding.