

STRUCTURE & DATA RECORDS: MARKUP, XML, JSON

Markup is not part of the text or content but tells something about it ...



To make markup work, the writer and the interpreter
of the marked up content have to **agree on the interpretation
of the markup symbols.** [1]



[1] Cynthia Zender (2005). Markup 101: Markup Basics. SAS Institute. <https://www.lexjansen.com/pharmasug/2005/Tutorials/tu12.pdf>
Interrobang punctuation mark: <https://www.merriam-webster.com/dictionary/interrobang>

Punctuational markup

!.

Presentational markup.

```
**bold**
```

Descriptive or declarative markup

```
<h1>The most important headline per page</h1>
```

Referential markup

```
<a href="url">link text displayed to reader on screen</a>
```

[1] James H. Coombs et al. (November 1987). Markup Systems and the Future of Scholarly Text Processing. Communications of the ACM 30.

<http://xml.coverpages.org/coombs.html#Note1>

Marking up a manuscript or page proof is usually a manual process.

In computer files, markup includes formatting instructions and additional information to the natural text so that software can format the text or a printer can print the document. [1]



[1] Cynthia Zender (2005). Markup 101: Markup Basics. SAS Institute. <https://www.lexjansen.com/pharmasug/2005/Tutorials/tu12.pdf>

Punctuational markup

!.

Presentational markup.

```
**bold**
```

Descriptive or declarative markup

```
<h1>The most important headline per page</h1>
```

Referential markup

```
<a href="url">link text displayed to reader on screen</a>
```

[1] James H. Coombs et al. (November 1987). Markup Systems and the Future of Scholarly Text Processing. Communications of the ACM 30.

<http://xml.coverpages.org/coombs.html#Note1>

(Meta)data exchange formats need to be read and processed by humans and computers.

Descriptive & referential markup makes natural text more accessible for computer analysis. [1]



[1] Charles F. Goldfarb (1990). The SGML Handbook. Clarendon Press. <https://books.google.com/books?id=RilvKya0EnwC>

SGML (Standard Generalized Markup Language) was one of the first industry standards for electronic publishing – a meta-language for generalized, descriptive markup languages – first accepted as an ISO standard in 1986.



Image: HTML source code, https://unsplash.com/photos/Wyd_PkCa1BY, note angle brackets

Both, HTML (1989) and XML (1998) are based on SGML. HTML (HyperText Markup Language) is the standard markup language for web pages. In contrast, the main purpose of XML (eXtensible Markup Language) is the transfer and storage of arbitrary data on the World Wide Web.

```
</div>
<div class="wrap_2stroke">
  <div class="second">
    <div class="second_content">
      <div class="block">
        <h3>...
        alt="..."
      </div>
```

[1] <https://www.iso.org/standard/16387.html>

Image: HTML source code, https://unsplash.com/photos/WYd_PkCa1BY

XML is software- and hardware-independent. It is considered human-readable and allows for hierarchical (tree-like) structures. Data elements are wrapped in start and end “tags”. [1]

```
<example>  
  <title>This is the example title</title>  
  <description>A simple XML example</description>  
  <wordCount>1</wordCount>  
</example>
```

[1] “XML Tutorial”. © 1999-2022. Refsnes Data, W3Schools. <https://www.w3schools.com/xml/>

JSON (JavaScript Object Notation) is not a markup language. It is a lightweight, human-readable, hierarchical format to store and transport data. JSON syntax is inspired by JavaScript object notation. [1] Like XML, JSON is software- and hardware-independent.

```
{  
  "key":"value",  
  "aString":"string",  
  "anInteger":5,  
  "aBoolean":true,  
  "anArray": ["item1", "item2", "item3"]  
}
```

[1] <https://www.ecma-international.org/publications-and-standards/standards/ecma-404/>