Structure & Data Records

Markup, XML, JSON

Fundamentals of Scientific Metadata: Why Context Matters

What is markup?

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

Making markup work

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

Types of markup

Punctuational markup

!.?

Presentational markup

bold

Descriptive or declarative markup

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

Referential markup

link text displayed to reader on screen

[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

Computer markup vs manual markup

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]

```
**make this text bold**
```

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

Types of markup

Punctuational markup

!.?

Presentational markup

bold

Descriptive or declarative markup

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

Referential markup

link text displayed to reader on screen

[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

Benefit of rigorous markup

(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

<language>SGML</language>

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: © Valery Sysoev on Unsplash

<language>XML and HTML</language>

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.

Image: © Valery Sysoev on Unsplash

<example>XML</example>

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/

{“format”:”JSON”}

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/

DISCLAIMER

This slide deck is part of the Lesson

Fundamentals of Scientific Metadata: Why Context Matters

published on **The Carpentries Incubator**.

Please cite this presentation as:

Gerlich, S., Strupp, A., Hofmann, V., Sandfeld, S. (2023). *Fundamentals of Scientific Metadata: Why Context Matters.* The Carpentries Incubator.

You can find more information about this course on **Github**.

image: https://c.pxhere.com/photos/35/f5/coffee_notebook_wooden_background_orange_work_table_office-1222115.jpg!d