



Chapter 4. Writing for the Web Environment

Our Web pages must communicate our ideas clearly and effectively, and the visual aspects are as important as the text. Navigation must be easy, and some thought must be given to the interaction experience of the user.

Going Public

NCBI has an international audience. Because humor does not translate well into other languages, keep instructions clear and concise.

Avoid jargon, slang, and nonstandard usages.

- Do not use “**flavors**” when you mean styles, formats, or varieties. Instead of writing “GenBank submission tools come in three flavors”, write “NCBI provides three GenBank submission tools”.
- Avoid using “**allow**” to mean that computer resources or controls provide a service. Instead of writing “My NCBI allows users to store searches”, write “My NCBI stores searches for users”.
- Do not write that a computer “**crunches**” data---“processes”, or a similar term, is more appropriate.

Read about [anthropomorphism](#) and think about how computer jargon may relate to it.

Before making your Web pages public, have your colleagues, a person naive to the project, and an editor review your pages. Use the most accurate representations of symbols, with consideration of the capabilities of various Web browsers. Consider layout and color, as well as content, as means of making your ideas understandable and engaging.

Constructing Searchable Web Pages

All static and dynamically generated Web pages on public servers must be designed to maximize usability and searchability. HTML documents must have title and meta tags in the <head> section, and links must be unbroken:

- **TITLE** The document title (of 50 characters or fewer, including spaces) must be unique within NCBI. In addition to the document title, consider adding a title attribute to each meaningful image on the page. This is important if you use an image with graphical text as the heading of your page. The title should briefly describe what is in the image. Most visual browsers preferentially show image title text as a tooltip over the image. If the image is absent, the browser uses the ALT tag. If the ALT tag is absent, there is no pop-up. Voice browsers also speak the title tag, if any, or the ALT tag, if any, or nothing. Document titles are displayed on the search results page of a search engine, so the title needs to be human readable and should clearly convey what the page contains.
- **AUTHOR** Use your group’s project name for the author value.
- **KEYWORDS** Although some search engines do not use keywords, keywords are still helpful in intranet searches, such as the NCBI Site Search. Use no more than 25 single keywords or some combination of 2-3-

word phrases that are specific to the content of the page in a comma-delimited list. Think of different ways to describe your content; people often use different words and phrases to search for the same content (e.g., medical school or school of medicine, vitamin C or ascorbic acid). Start with a base of "national center for biotechnology information, ncbi, national library of medicine, nlm, database, archive" (searches of keywords are case insensitive) and then add words more specific to the content of your NCBI pages. Use a thesaurus or the [Medical SubHeading \(MeSH\)](#) browser to explore additional keywords, as well as a dictionary to ensure correct spelling and to find common alternate spellings. This approach will be useful for both internet and intranet searches by search engines.

- **DESCRIPTION** The description, like the title, must be unique. The best descriptions are short and specific sentences with fewer than 25 words, with the most relevant words near the beginning. Descriptions are displayed on the search results page of a search engine.
- **LINKS** Links may take you out of the present document or take you to another part of the same document. Links out of the document must reference existing content. These links include hyperlink "href" attributes, such as `label`, image src tags, such as ``, and linked stylesheets and Javascript files, such as `<link rel="stylesheet" type="text/css" href="styles.css"/>` and `<link rel="javascript" type="text/javascript" href="myscript.js"/>`. All links within the document must refer to existing anchors.

Example

```
<head>
```

```
<title>Single Nucleotide Polymorphism (SNP)</title>
```

```
<meta name="author" content="snpdev">
```

```
<meta name="keywords" content="national center for biotechnology information, ncbi, national library of medicine, nlm, national institutes of health, nih, database, archive, single nucleotide polymorphism, SNP, sequence variation, genome, gene mapping, population structure, dbSNP, evolution, microsatellite repeat, phenotype">
```

```
<meta name="description" content="The Single Nucleotide Polymorphism database (dbSNP) is a public-domain archive for a broad collection of simple genetic polymorphisms.">
```

```
</head>
```

Make sure that the content on your page is visible on browsers that have JavaScript disabled. Not all search engines can index content contained in scripts.

Keep Your Web Pages Current

Dates are an important part of your Web document because they help readers understand where your information fits in the timeline of relevant events. In some Web documents, dates serve merely to let readers know when the pages were checked. In other Web documents that report on fast-changing information, dates assume more importance.

When information is time sensitive, place the date at the top of the Web document. Otherwise, place the date at the bottom of the page.

It is also helpful to distinguish these two types of dates. "Last Updated" conveys some sense of information that is time sensitive, whereas "Revised" seems to suggest more of a "maintenance" approach. For the sake of consistency in the presentation of these dates, use the following forms:

- Last Updated: November 21, 2003.

- Revised: November 21, 2003.

When using “Last Updated”, place the date at the **top** of the Web document. When using “Revised”, place the date at the **bottom** of the page.

Note the periods at the end of the entries.

Glossaries on the Web

There are two styles for glossaries on the NCBI Web pages:

word to be defined - fragment and/or sentence definition.

word to be defined

fragment and/or sentence definition.

The first one saves a little space, if the definition is short enough. Note that both definitions begin with a lowercase letter and end with a period.

Itemized Lists

Many NCBI Web pages contain itemized lists. It is important to consider the elements of an itemized list and then decide on the type of list that you will need. For more information, also see [Lists in Style Points and Conventions](#).

There are three considerations for making lists:

1. **Type of list:** Is the **order** of the list items important? (If Yes, use a **Numbered List**; if No, use a **Bulleted List**.)
2. **Whether list items are fragments or complete sentences:** Will the list items be fragments or complete sentences? (Sentence-style punctuation is used for complete sentences.)
3. **Capitalization:** When should list items be capitalized? (Action verbs that begin lists and lists that follow headings without additional explanatory text begin with capital letters.)

Below are six examples of the two types of lists. There are two style choices in the **Numbered List** and four style choices in the **Bulleted List**. Please apply your style choice to your Web document consistently.

Numbered List

In a Numbered List, the order of execution of the items is important and/or emphasis on the number of items is being made.

When providing instructions or directions that must be followed in a particular order, capitalize the first word of the entry and end it with a period. Each item of this list begins with an “action” verb (sometimes called “commands”), meaning that the person reading the list is being told what to do and must take action to get the wanted results. See Example 1.

Example 1:

1. Press the Validate and Continue button.
2. Enter the number of CDS features.
3. Scroll down to the CDS Feature box.
4. Complete the CDS Feature subsections.

When the items in the list are not complete sentences, no punctuation is used at the ends of the items. See Example 2.

Example 2:

For each complete submission you have made to us, you will receive by email:

1. an automatic preliminary GenBank flatfile
2. a GenBank Accession number
3. a completed GenBank flatfile

Bulleted List

In a Bulleted List, the order of execution of the items is not important (it is not a step-by-step procedure), and “counting” the items is not necessary.

In both Examples 3 and 4 that follow, the list items merely continue the idea of the introductory sentence fragment. Contrast the use of the verbs in Example 4 with the action verbs in the Numbered List, Example 1. Note that these verbs in Example 4 are not capitalized because they are not giving directions for actions as commands. Note also that there are no periods, semicolons, or commas at the ends of the entries in Examples 3 and 4, because each item entry presents a part of the thought (each item entry is only a part of the whole idea, with the “idea” being the introduction to the list by the introductory sentence fragment and the list entries themselves). Because of the layout of the list itself, ending punctuation is not needed; the list presents the idea clearly.

Example 3:

Use BankIt if:

- you have one or a few sequence submissions
- you prefer to use a WWW-based submission tool
- your sequence annotation is not complicated
- you do not require sequence analysis tools to submit your sequence(s)

Example 4:

Failure to recognize foreign segments in a sequence can:

- lead to erroneous conclusions about the biological significance of the sequence
- waste time and effort in analysis of contaminated sequence
- delay the release of the sequence in a public database
- pollute public databases with contaminated sequence

When a list of complete sentences immediately follows a heading, use paragraph indents with filled bullets at the first level. (Your Browser may control the types of symbols used with nesting.) Begin with a capital letter and end the sentence with a period. See Example 5.

Example 5:

BankIt: GenBank Submission by WWW

- GenBank provides annotation examples and descriptions for several types of sequence submissions.
- Each sequence must be submitted individually and will receive its own BankIt number.

When a list of fragments immediately follows a heading, use paragraph indents with filled bullets. Begin with a capital letter and do not use periods, semicolons, or commas as closing punctuation for the fragments. A

combination of fragments and complete sentences is fine. If a complete sentence is used, use a closing period. See Example 6 for use of this combined form.

Example 6:

Requirements for Every Submission

- Contact information
- Release data information
- Input DNA sequence
 - A contiguous nucleotide sequence of at least 50 base pairs
 - Type of molecule sequenced
 - Description of the sequence:
 - Click on the specific examples in the left column.
 - If the sequence submitted does not fall into one of the categories on the left, please use the available features or provide a complete description.

Outlines in Web Help Documents

There are many ways to present information clearly and effectively. Consider using an outline when writing your Web Help documents. If you do plan on using an outline, an outline at the beginning of your Web Help document is a good way to orient your readers so that they can quickly understand the layout of your information. An outline is another visual aid to supplement the information supplied by the sidebars and toolbars. One good outline method, shown below, uses an indent system shown with specific symbols at each level of indent, a system used by Word and other wordprocessors.

This is an example of the indent system:

- Overview
 - Pubmed Searches
 - Author Names

An alternative method uses a combination of numbers and symbols:

1. Overview
 - PubMed Searches
 - Author Names

Both methods present Help documentation clearly and effectively.

Many people will print out documents to read off-screen. Unless the outline is on a Web page by itself, the assumption is that all of the material contained in the Outline will print out as one document. If sections of the Outline are links to other pages, add a parenthetical statement to the Outline section, such as “(open and print this link separately)”. Note that the words are lowercase. In general in Outlines, use lowercase for parenthetical information. A good example of this is shown on the [OMIM Help page](#).

Web Pages without Reference Lists: Making Citations

Link every citation in the text to PubMed, PubMed Central, and Bookshelf, where possible. Where this is not possible, all information must be provided for the reader in [National Library of Medicine \(NLM\)](#) format:

- ... was described previously ([Lipman et al. 2002](#); [Benson et al. 2000](#)) ([Abramovitz et al. Chemosphere 1990;21:1221-1229](#)) ([Salton G. Automatic text processing. Reading \(MA\): Addison-Wesley; 1989](#)).

In the example above, both the Lipman journal citation and the Benson journal citation are linked to PubMed. The journal article by Abramovitz is not in PubMed, but enough information is given in this format for the reader to find it from other sources. Likewise, the book by Salton contains enough information for the reader to contact the publisher or bookseller. When a reference list is not being provided on the Web page, complete citations must be provided in the text itself. The four citations given in the sample above are separated because of punctuation issues and concern for readers to be able to more easily understand the amount of information presented with the citations.

NCBI Standard Abbreviations and Frequently Used Word Forms

NCBI has an international audience of all educational levels. As an author, you must consider your audience, meaning that if you are presenting a biology paper of great complexity that is not likely to be read by a general audience, it is probably not necessary to write out DNA. If there is any doubt, a good method would be to provide the written-out form of the word, followed by the abbreviated form in parentheses:

- ... is called deoxyribonucleic acid (DNA).

Thereafter, use the abbreviated form only. Begin with your written-out forms again for every page that loads on the Web site, unless the pages that follow are a clear continuation of the initial page. Obviously, you must use your judgment on this.

Frequently Used Word Forms

a, an - article used according to the sound of the abbreviation (e.g., an STS; a sequence tagged site); this style point is an exception to the CBE rule, which states that the article is used according to the sound of the written-out form.

Jump to: [A-E] [D-F] [G-L] [M-O] [P-R] [S-Z]

[A-E]

AIDSLINE - AIDSTrials.

ASN.1

AVLINE

BAC-end sequence - (note hyphen).

BankIt

base pair - (two words).

BLAST - PHI-BLAST, PSI-BLAST, QBLAST, RPS-BLAST, IgBLAST, MegaBLAST.

BLink - BLAST link.

BLOSUM62

Bookshelf

Boolean - (capitalized).

by-product(s)

C terminus, COOH terminus, C-terminal

catalog - (no “-ue”).

CATLINE

CCAP - Cancer Chromosome Aberration Project.

CD-Search

CDD - Conserved Domain Database.

CGAP - Cancer Genome Anatomy Project.

CGH - comparative genomic hybridization.

check box

ChemID plus

cis - (italic).

ClinicalTrials.gov

ClustalW

Cn3D

co - co-dominant, co-inheritance, co-segregation, cofactor, cooperative.

COGnitor

COGs - Clusters of Orthologous Groups.

cross - crossing-over (recombination); but crossover (events), crossovers, crosslink, crossbreeding.

cutoff - (noun).

cut off - (verb).

[D-F]

3D - (for three-dimensional; not 3-D).

DART - Domain Architectural Retrieval Tool.

data - this is a plural word (datum is singular).

database-mining

dataset

dbEST

dbGSS

dbSNP

dbSTS

DDBJ

DDBJ/EMBL/GenBank - (alphabetical order as agreed by collaborators).

DDD - Digital Differential Display.

DGED - Differential Gene Expression Displayer.

drop-down menu

email - consider putting the email address in boldface text.

EMBL

employs - change to “uses” unless specifically referring to the employment of a person.

Entrez

Entrez Genomes

Entrez Nucleotides

EST - expressed sequence tag.

eukaryote - (no “c”).

ExPASy

fa2htgs - not to be capitalized, even beginning a sentence.

false-positive results

FASTA

FISH - fluorescence *in situ* hybridization.

flatfile - (one word).

fruit fly

FTP

[G-L]

GDS - GEO DataSets.

GenBank

GeneMap'99

GenPept

GEO - Gene Expression Omnibus.

GI - “Geninfo Identifier”.

GLS - Gene Library Summarizer

Help desk - NCBI Help desk

homepage

homolog - (no “-ue”) .

HomoloGene

Human–Mouse Homology - (note en dash).

ID - sequence ID, PubMed ID, MEDLINE UID.

in vitro - (italic).

in vivo - (italic).

kb - kilobase.

LOCATORplus

LocusLink

log in - (verb).

login - (noun).

look-up table

lowercase - (one word).

[M-O]

MAGEML - MicroArray Gene Expression Markup Language.

Map Viewer

matrices

Mb - megabase.

MD - (no spaces, no periods; “Medical Doctor”).

MEDLINE

MEDLINEplus

MGC - Mammalian Gene Collection.

MGD - Mouse Genome Database.

MGED - Microarray Gene Expression Database.

MinSeq - Minimal Sequence.

MMDB - Molecular Modeling Database.

Model Maker

multi - multi-step.

N terminus, NH₂ terminus, N-terminal

NEXUS

NLM Gateway

non - non-coding, non-cumulative, non-identical, non-redundant, non-synonymous, nonhuman.

OLDMEDLINE

OMIM

online

ONO - order and orientation information

[P-R]

pairwise

paralog - (no “-ue”).

PDB - Protein Data Bank (Brookhaven).

PDBeast

Pfam

PhD - (no spaces, no periods).

PHYLIP

PopSet

post - post-translational.

pre - precalculate, precomputed, predetermine, pre-existing.

PREMEDLINE

ProbeSet

prokaryote - (no “c”).

PROW - Protein Reviews On the Web.

psort2

PSSM - position-specific scoring matrix.

PubMed

PubMed Central

pull-down menu

RCSB - Research Collaboratory for Structural Bioinformatics.

RefSeq - Reference Sequence.

RepeatMasker

RFLP - restriction fragment length polymorphism.

[S-Z]

SAGE - serial analysis of gene expression.

SAGEmap

Sequin

SERLINE

SKY/CGH - Spectral Karyotyping and Comparative Genomic Hybridization database.

SMART

SNP - single nucleotide polymorphism.

SOFT - Simple Omnibus Format in Text.

species - spp. (plural; note period).

Spidey

SSLP - simple sequence length polymorphism.

stand-alone program - (note hyphen).

STS - sequence tagged site.

SWISS-PROT**TaxBlast****Taxonomy Browser****TaxPlot**

tbl2asn - not to capitalized, even beginning a sentence.

TinySeq - Tiny Sequence.

ToolKit**ToolBox**

trans - (italic).

UniGene**UniSTS****UniVec**

uppercase - (one word).

UserID

VAST - Vector Alignment Search Tool.

VecScreen

versus - (italic); OK to abbreviate as “vs.” in tables.

VNTR - variable number of tandem repeats.

website**Whole Genome Shotgun sequencing****wild card**

XML

xProfiler

X-ray

ZFIN - Zebrafish Information.

Be consistent with these frequently used word forms.