

*National Digital Strategy Advisory Board Meeting*  
*Library of Congress, June 28, 2004*

**National Library of Medicine**  
*Digital Preservation Update*

**Betsy L. Humphreys**  
**Associate Director for Library Operations**  
**NLM, NIH, HHS**  
**[blh@nlm.nih.gov](mailto:blh@nlm.nih.gov)**

# NLM Statutory Functions

(NLM Act, 1956)

---

- ◆ Acquire and preserve library materials pertinent to medicine
- ◆ Organize the materials by appropriate cataloging, indexing, and bibliographical listing
- ◆ Disseminate the catalogs, indexes, and bibliographies
- ◆ Make the library materials available through loans, copying procedures, or otherwise\*
- ◆ Provide reference and research assistance\*

\*NLM is national backup to other members of the National Network of Libraries of Medicine (MLAA, 1965)



# Relevant NLM policy

---

- ◆ Preserve intellectual content if preserving *both* content *and* original format not feasible
- ◆ As first priority, preserve content of the journals indexed by NLM and its predecessors
- ◆ Encourage publication in permanent media to reduce future preservation costs

# NLM Board of Regents

## Long Range Plan 2000-2005

---

- ◆ Organization of electronic information
  - .....
- ◆ **Permanent access** to electronic information
  - Take a leadership role in ensuring permanent access to important digital materials in health and biomedicine, **using NLM's own electronic output and services as test-beds**
  - Work with national libraries and other organizations to develop, test, and implement standards and strategies for permanent access to digital information

# Types of Data

---

- ◆ Databases – e.g., bibliographic, genetics/microbiology, toxicology and environmental health, vocabulary (UMLS), etc.
- ◆ Text (with some multimedia) – published literature, Web documents
- ◆ Image data – e.g., historical pictures and films, Visible Human, 3-D structures



# PMC Journal List

[About PMC](#)    [Journal List](#)    [Search](#)    [Utilities](#)

## PubMed Central Journals — Tabbed List Show Full List

Search by part or all of a journal name.

Find Journals

- New ... tab** lists journals added to PMC in the past 60 days.
- Title** links to a list of **all issues** of the journal in PMC.
- Latest Volume** links to the **most recent issue** available in PMC.
- Free Access** says how soon after publication the journal's articles are made free.

- A-B   
 C-H   
 I-M   
 N-S   
 T-Z   
 New ...

Search this Journal	Title	Volumes in PMC		Free Access
		Latest	First	
<a href="#">Search</a>	<a href="#">Infection and Immunity</a>	v.72(6) Jun 2004	v.1 1970	After 6 months
<a href="#">Search</a>	<a href="#">International Journal of Behavioural Nutrition and Physical Activity</a>	v.1(1) 2004	v.1 2004	Immediate
<a href="#">Search</a>	<a href="#">International Journal for Equity in Health</a>	v.3(1) 2004	v.1 2002	Immediate
<a href="#">Search</a>	<a href="#">International Journal of Health Geographics</a>	v.3(1) 2004	v.1 2002	Immediate
<a href="#">Search</a>	<a href="#">Journal of the American Medical Informatics Association</a>	v.10(6) Nov 2003	v.1 1994	After 6 months



# PubMed Central Growth

---

	<u>April 03</u>	<u>April 04</u>	<u>% Change</u>
<b>Journals in PMC</b>	121	148	22%
<b>Articles in PMC</b>	84,637	302,053	257%
<b>Unique Users</b>	273,207	736,000	169%
<b>Articles Accessed</b>	701,046	1,745,202	148%



## International Agreement to Expand PubMed Central

**Additional medical journals, some dating back more than 125 years, will be made freely available on the Internet.**

The Wellcome Trust, in partnership with the Joint Information Systems Committee (JISC), and the U.S. National Library of Medicine (NLM) are joining forces to digitize the complete backfiles of a number of important and historically significant medical journals. The digitized content will be made freely available on the Internet – via PubMed Central (<http://www.pubmedcentral.gov/>) and augment the content already available there.

With funding of £1.25 million [£750,000 from the Trust, £500,000 from the JISC] the project plans to digitize around 1.7 million pages of text. The NLM will manage the project, host the archive and ensure that the digital files are preserved in perpetuity.

The list of journals to be digitized will include the *Annals of Surgery*, *Biochemical Journal*, *Journal of Physiology* and *Medical History*. Digitization will commence in Summer 2004 and the first titles will be online early in 2005.

Once the backfiles are digitized, a significant new body of medical literature will be freely available on the Internet. Examples include:

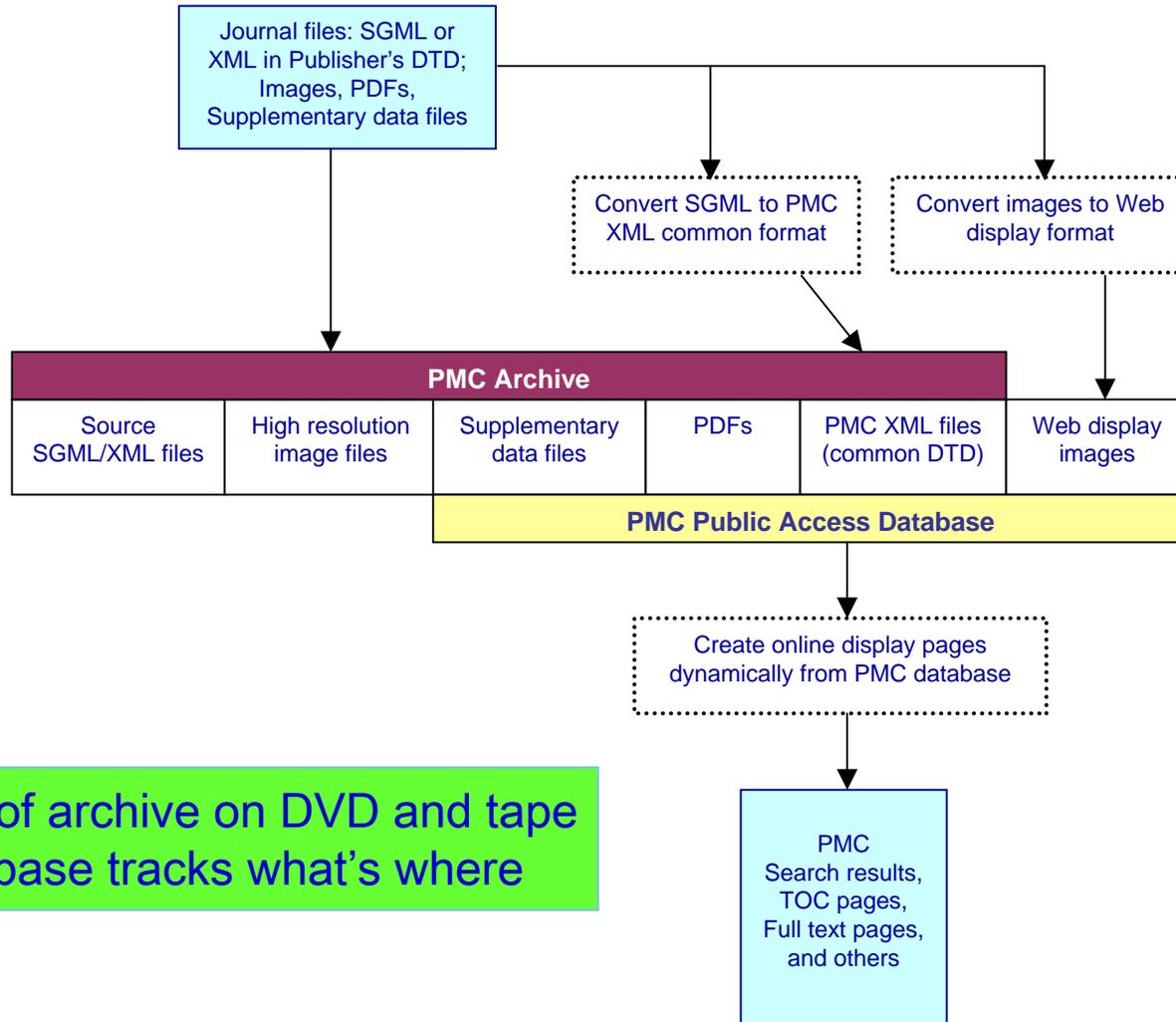
- Sydney Ringer's original research, (published in the *Journal of Physiology* in the 1880's), on the actions of inorganic salts on living tissues: work that provided the theoretical basis for the

[Printer-friendly Version](#)

**National Institutes of Health**  
FOR IMMEDIATE RELEASE  
June 25, 2004

National Library of Medicine  
Robert Mehnert  
Kathy Cravedi  
(301) 496-6308  
[publicinfo@nlm.nih.gov](mailto:publicinfo@nlm.nih.gov)

# PubMed Central Archiving Model



Multiple copies of archive on DVD and tape  
Catalog database tracks what's where





# Archiving and Interchange DTD



## Links

### Introduction

### Documentation

[Tag Library](#)  
[FAQ](#)

### Getting the Files

[FTP Directory](#)  
[DTD Zip File](#)  
[Schema Zip File](#)

### Updates

[Version 1.1](#)  
[W3C Schema](#)

### Feedback

[e-mail](#)  
[Feedback Form](#)

### Related DTDs

[Publishing DTD](#)  
[PubMed DTD](#)

## Introduction

The National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM) created the Journal Archiving and Interchange Document Type Definition (DTD) with the intent of providing a common format in which publishers and archives can exchange journal content.

This DTD was created from the Journal Archiving and Interchange DTD Suite, which provides a set of XML modules that define elements and attributes for describing the textual and graphical content of journal articles as well as some non-article material such as letters, editorials, and book and product reviews).

## The Suite of Modules

The intent of this DTD Suite is to preserve the intellectual content of journals independent of the form in which that content was originally delivered. The Suite has been written as a set of XML DTD modules, each of which is a separate physical file. No module is an entire DTD by itself, but these modules can be combined into a number of different DTDs.

The Archiving and Interchange DTD may be used as is, or the Suite can be used to construct DTDs for authoring and archiving journal articles as well as DTDs for transferring journal articles from publishers to archives and between archives. Details on

# DTD Technical Advisory Group

to PubMed Central National Advisory Committee

---

- ◆ Adriana Aredeleau, Microsoft
- ◆ Matt Cockerill, Biomed Central
- ◆ Mike Eisen, UC Berkeley
- ◆ Beth Friedman, DCL
- ◆ Gerry Grenier, Publishing Technologies IEEE, Inc.
- ◆ Clifford Lynch, CNI
- ◆ Evan Owens, Ithaka Electronic Archiving Initiative
- ◆ Diana Robinson, Highwire
- ◆ Bruce Rosenblum, Inera
- ◆ Nick Twyman, PLoS
- ◆ B. Tommie Usdin, Mulbery
- ◆ Roy Williams, National Virtual Laboratory, CalTech

# On the agenda

---

- ◆ DTD suite enhancements, including expansion to book chapters
- ◆ XML authoring tool – journal articles, book chapters
  - Microsoft Word, plus NLM/NCBI template
  - Submission to NLM/NCBI website for validation
- ◆ Functional specifications, database structure for “lite” journal archives



United States  
**National Library of Medicine**  
National Institutes of Health

Search NLM Web Site

Go

[NLM Home](#) | [Contact NLM](#) | [Site Map](#) | [FAQs](#)

# History of Medicine

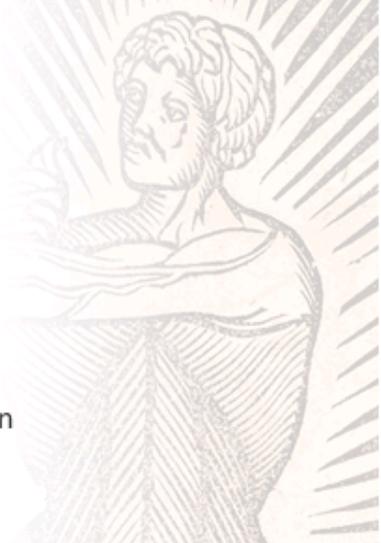
[History Home](#) | [About Us](#) | [Visit Us](#) | [Contact Us](#) | [Tours](#) | [FAQs](#) | [Copyright](#) | [HMD Site Map](#)

## Historical Collections

- ▶ Books and Journals
- ▶ Archives and Manuscripts
- ▶ Prints and Photographs
- ▶ Films and Videos
- ▶ Digital Collections

## Using the Collections

- ▶ Online Catalogs
- ▶ Printed Catalogs and Guides
- ▶ Reference Services and Information



Profiles in Science ▶



Quick Links:

Choose One

Go

Especially for:

- [First Time Visitors](#)
- [Scholars & Historians](#)
- [Image & Film Researchers](#)
- [Teachers & Students](#)

What's Happening

- [Seminars, Lectures, Films](#)
- [Exhibitions](#)
- [Tours](#)
- [Recent Acquisitions](#)

New Digital Project at NLM



On Exhibit at NLM



CELEBRATING AMERICA'S WOMEN PHYSICIANS

Last updated: 02 May 2004

First published: 01 January 1998

Permanence level: Permanent: Dynamic Content

# NLM Permanence Ratings

---

- ◆ Four categories of permanence have been defined:
  - 1) Permanent, unchanging content: NLM has made a commitment to keep this resource permanently available. Its identifier will always provide access to the resource. Its content will not change.
  - 2) Permanent, stable content: NLM has made a commitment to keep this resource permanently available. Its identifier will always provide access to the resource. Its content is subject only to minor corrections or additions.

# NLM Permanence Ratings

---

- ◆ 3) Permanent, dynamic content: NLM has made a commitment to keep this resource permanently available. Its identifier will always provide access to the resource. Its content could be revised, replaced.
- ◆ 4) Permanence not guaranteed: NLM has made no commitment to retain this resource. It could become unavailable at any time. Its identifier could be changed.



Go

Search Term(s): 

Search

[Help](#)

Search results for "UMLS"

## Search results found in:

- [National Library of Medicine](#) (241)
- [Health Information - MedlinePlus](#) (0)
- [Profiles in Science](#) (1)
- [Exhibits](#) (0)
- [Archives](#) (8)

**From the Archives**[Unified Medical Language System Fact Sheet](#)<http://www.nlm.nih.gov/archive/20031120/pubs/factsheets/ucls.html>

Last updated: 2003-11-18

First published: 1988-01-01

[Previous Version](#)[UMLS Metathesaurus Fact Sheet](#)<http://www.nlm.nih.gov/archive/20031120/pubs/factsheets/uclsmeta.html>

Last updated: 2003-11-26

First published: 1994-01-01

[2003 License Agreement for Use of the UMLS Products](#)<http://www.nlm.nih.gov/archive/20040106/research/ucls/license.html>

Last updated: 2003-11-12

First published: 1997-01-01

[NLM Newsline 2000 January-March Vol. 55, No. 1](#)[http://www.nlm.nih.gov/archive/20040423/pubs/nlmnews/janmar00/jm00\\_print.html](http://www.nlm.nih.gov/archive/20040423/pubs/nlmnews/janmar00/jm00_print.html)

Last updated: 2000-11-30

First published: 2000-01-01

[NLM News, March-April 1996](#)<http://www.nlm.nih.gov/archive/20040422/pubs/nlmnews/marapr96.html>

Last updated: 1996-10-01

First published: 1996-03-01

[NLM NEWSLINE, JANUARY-MARCH 1999: Part 8](#)[http://www.nlm.nih.gov/archive/20040423/pubs/nlmnews/jm99/jm99\\_print.html](http://www.nlm.nih.gov/archive/20040423/pubs/nlmnews/jm99/jm99_print.html)



## Fact Sheets

[Contact NLM](#) | [Site Index](#) | [Search Our Web Site](#) | [NLM Home](#)

[Health Information](#) | [Library Services](#) | [Research Programs](#) | [New & Noteworthy](#) | [General Information](#)

### Fact Sheet

## Unified Medical Language System<sup>®</sup>

### Background:

In 1986, the National Library of Medicine (NLM), began a long term research and development project to build a Unified Medical Language System (UMLS<sup>®</sup>). The purpose of the UMLS is to aid the development of systems that help health professionals and researchers retrieve and integrate electronic biomedical information from a variety of sources and to make it easy for users to link disparate information systems, including

**Last updated:** 18 July 2003

**First published:** 01 January 1988

**Permanence level:** Permanent: Stable Content

[Replaced by](#)

---

U.S. National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894 Copyright, Privacy, Accessibility  
National Institutes of Health, Department of Health & Human Services



## Fact Sheet

# Unified Medical Language System<sup>®</sup>

[Printer-friendly Version](#)

### Purpose:

The purpose of NLM's Unified Medical Language System<sup>®</sup> (UMLS<sup>®</sup>) is to facilitate the development of computer systems that behave as if they "understand" the meaning of the language of biomedicine and health. To that end, NLM produces and distributes the UMLS Knowledge Sources (databases) and associated software tools (programs) for use by system developers in building or enhancing electronic information systems that create, process, retrieve, integrate, and/or aggregate biomedical and health data and information, as well as in informatics research. By design, the UMLS Knowledge Sources are multi-purpose. They are not optimized for particular applications, but can be applied in systems that perform a range of functions involving one or more types of information, e.g., patient records, scientific literature, guidelines, public health data. The associated UMLS software tools assist developers in customizing or using the UMLS Knowledge Sources for particular purposes.

### UMLS Knowledge Sources and Software Tools:

There are three UMLS Knowledge Sources: the [Metathesaurus<sup>®</sup>](#), the [Semantic Network](#), and the [SPECIALIST lexicon](#). They are distributed with flexible [lexical tools](#) and the [MetamorphoSys](#) install and customization program.

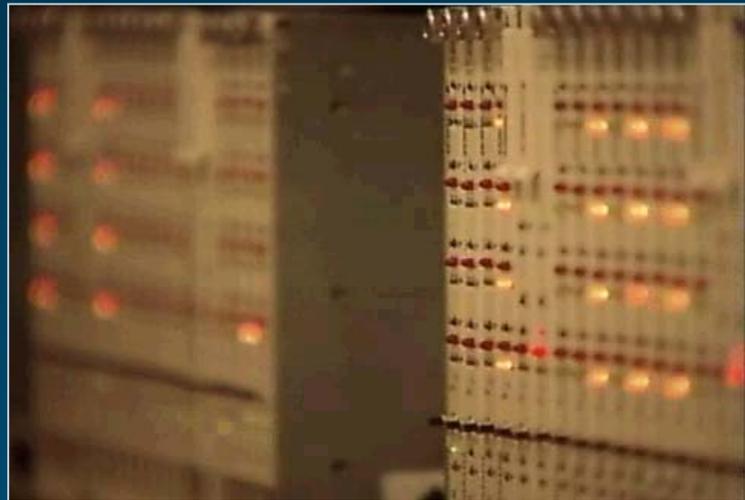
The Metathesaurus provides a uniform, integrated distribution format for more than 100 biomedical and health-related vocabularies, classifications, and coding systems (some in multiple languages) and links

# On the agenda

---

- ◆ Archiving non-HTML files, such as PDF, video and audio clips, etc.
- ◆ Archiving resources from NLM areas which do not use TeamSite
- ◆ Archiving test with other NIH components

# Back-up Computer Facility



# Information System Service Levels

