



# Information Retrieval Content

---

What is Biomedical & Health Informatics?  
William Hersh  
Copyright 2023  
Oregon Health & Science University



# Classification of knowledge-based content

- Bibliographic
  - By definition rich in metadata
- Full-text
  - Everything on-line
- Annotated
  - Non-text or structured text annotated with text
- Aggregations
  - Bringing together all of the above
- These categories are somewhat fuzzy, and increasing numbers of resources have more than one type

# Bibliographic content

- Bibliographic databases
  - The old (e.g., MEDLINE) have been revitalized with new features
  - New ones (e.g., ECRI Guidelines Trust) have emerged
- Web catalogs
  - Share many characteristics of traditional bibliographic databases
- Real simple syndication/Rich site summary (RSS)
  - “Feeds” provided information about new content

# Bibliographic databases

- Contain metadata about (mostly) journal articles and other resources typically found in libraries
- Produced by
  - U.S. government – most produced by National Library of Medicine (NLM, <https://www.nlm.nih.gov/>)
    - e.g., MEDLINE, omics information, etc.
  - Commercial publishers, e.g.,
    - EMBASE – part of larger SciVal
    - CINAHL – Cumulative Index to Nursing and Allied Health Literature
    - ACM Guide to Computing Literature – computer science and related areas
    - Google Scholar – <http://scholar.google.com>

# MEDLINE

- References to biomedical journal literature
  - Original medical IR database – system for searching MEDLINE launched in 1971 with literature maintained in MEDLARS system dating back to 1966
    - Name derives from MEDLARS On-Line – MEDLINE
  - NLM maintains policies for addition and maintenance – [https://www.nlm.nih.gov/medline/medline\\_policies.html](https://www.nlm.nih.gov/medline/medline_policies.html)
  - Free to world since 1997 via PubMed – <https://pubmed.gov/>
    - Now with links to full text of articles and other resources
    - Beginning to add preprints, e.g., BioRxiv and MedRxiv
    - PubMed contains some additional content not in MEDLINE, e.g., PubMed Central and books
- Statistics
  - [https://www.nlm.nih.gov/bsd/medline\\_pubmed\\_production\\_stats.html](https://www.nlm.nih.gov/bsd/medline_pubmed_production_stats.html)
  - Over 30M references to peer-reviewed literature
  - Over 5200 journals, mostly English language
  - Over 1M new references added yearly

# ECRI Guidelines Trust

- <https://guidelines.ecri.org/>
- Contains detailed information about guidelines
  - Including degree they are evidence-based
  - Interface allows comparison of elements in database for multiple guidelines
- Links to those free on Web and to producers when proprietary
- Successor to Agency for Healthcare Research and Quality (AHRQ) National Guidelines Clearinghouse

# Web catalogs

- Generally aim to provide quality-filtered Web sites aimed at specific audiences
  - Distinction between catalogs and sites blurry
- Some are aimed towards clinicians
  - Translating Research into Practice –  
<https://www.tripdatabase.com/>
- Others are aimed towards patients/consumers
  - MedlinePlus – part of larger consumer health site –  
<https://medlineplus.gov/>

# RSS – mostly defunct (Target, 2019)

- RSS “feeds” provide short summaries, typically of news, journal articles, or other recent postings on Web sites
- Users receive RSS feeds by an RSS aggregator that can typically be configured for the site(s) desired and to filter based on content
  - Work as standalone, in Web browsers, in email clients, etc.
- Forked into different versions but basically provided
  - Title – name of item
  - Link – URL of full page
  - Description – brief description of page

# Full-text content

- Contains complete text as well as tables, figures, images, etc.
- If there is corresponding print version, both are usually identical
- Includes
  - Periodicals
  - Books
  - Web sites – may include either of above

# Full-text primary literature

- Almost all biomedical journals available electronically
- Many initially published by Highwire Press (<https://www.highwirepress.com/>), which added value to content of original publisher
- Now also published by leading commercial scientific publishers, e.g., Elsevier, Kluwer, Springer, etc.
- Growing number available via open-access model, e.g., Biomed Central (BMC), Public Library of Science (PLOS)

# Full-text literature before publication

- Repository of full-text papers funded by NIH research
  - PubMed Central (PMC;  
<https://www.ncbi.nlm.nih.gov/pmc/>)
- Preprint servers – some journals maintain but also general sites
  - arXiv – <https://arxiv.org/>
  - Biology – <https://www.biorxiv.org/>
  - Medicine – <https://www.medrxiv.org/>
  - JMIR preprints – <https://preprints.jmir.org/>

# Books

- Textbooks
  - Most well-known clinical textbooks are now available electronically
    - e.g., Harrison's Principles of Internal Medicine
  - NLM Bookshelf
    - <https://www.ncbi.nlm.nih.gov/books>
- Compendia of drugs, diseases, evidence, etc.
- Handbooks – very popular with clinicians
- Many of above are bundled into aggregations by publishers
  - e.g., Access Medicine (McGraw-Hill), Elsevier, Kluwer
  - Also increasingly published on mobile devices

# Value added for electronic books

- Multimedia, e.g., skin lesions, shuffling gait of Parkinson’s Disease, etc.
- Bundling of multiple books
- Can be updated in between “editions”
- Linkage to other information, e.g., to references, self-assessments, updates, other resources, etc.

The screenshot shows the Merck Manuals Professional Version website. The header includes the Merck logo and 'Professional Version' text. A navigation bar contains links for Medical Topics, Drug Information, News & Commentary, Procedures & Exams, Quizzes & Cases, and Resources. A search bar is present with the text 'Find information on medical topics, symptoms, drugs, procedures, news and more, written for the health care professional.' Below the search bar, the breadcrumb trail reads 'Professional / Endocrine and Metabolic Disorders / Lipid Disorders'. The main content area is titled 'Dyslipidemia (Hyperlipidemia)' by Anne Carol Goldberg, MD. A note indicates this is the Professional Version and provides a link to the Consumer Version. The text describes dyslipidemia as an elevation of plasma cholesterol, triglycerides (TGs), or both, or a low high-density lipoprotein level that contributes to atherosclerosis. It mentions that diagnosis is by measuring plasma levels of total cholesterol, TGs, and individual lipoproteins, and that treatment involves dietary changes, exercise, and lipid-lowering drugs. A sidebar on the left lists sections: Dyslipidemia, Classification, Etiology (Primary and Secondary causes), Symptoms and Signs, Diagnosis (Lipid profile measurement, Other tests, Secondary causes, Screening), Treatment (General principles, Elevated LDL cholesterol, Elevated TGs, Low HDL, Elevated Lp(a), Secondary causes, Monitoring treatment), and Key Points. A diagram on the right shows 'Lipid Disorders' branching into 'Overview of Lipid Metabolism', 'Dyslipidemia', and 'Hypolipidemia'.

# Web sites

- Defined more narrowly here to refer to coherent collections of information on Web
- Usually take advantage of Web features, such as linking, multimedia
- Increasingly integrated with other resources and available on different platforms (e.g., integrated into electronic health records [EHRs], on smartphones, etc.)

# Some notable full-text content on Web sites

- Government agencies
  - National Cancer Institute
    - <https://www.cancer.gov/>
  - Centers for Disease Control – travel and infection information
    - <https://www.cdc.gov/health-topics.html>
    - <https://wwwnc.cdc.gov/travel/>
    - <https://www.coronavirus.gov/>
  - Other NIH institutes, e.g., National Heart, Lung, and Blood Institute (NHLBI)
    - <https://www.nhlbi.nih.gov/>

# Full-text Web sites (cont.)

- Physician-oriented medical news and overviews, e.g.,
  - Medscape – <https://www.medscape.com/>
  - Many professional societies provide to members, e.g., <https://www.acponline.org/clinical-information>
- Patient/consumer-oriented, e.g.,
  - MayoClinic.org – <https://www.mayoclinic.org/diseases-conditions>
  - WebMD – <https://www.webmd.com/>
- Many mobile apps provide health information, e.g.,
  - Epocrates – <https://www.epocrates.com/>
  - WebMD app for consumers

# Other types of Web content

- Wikipedia – <https://www.wikipedia.org/>
  - Encyclopedia with free access and distributed authorship
  - Medical content often retrieved in general Web searches (Laurent, 2009)
  - Making attempt to improve quality of medical content (Heilman, 2013; Shafee, 2017; Azzam, 2017)
- Body of knowledge
  - Software Engineering Body of Knowledge (SWEBOK, <https://www.computer.org/education/bodies-of-knowledge/software-engineering>) organizes knowledge of field
- Social media and beyond – Twitter, Facebook, LinkedIn, etc.

# Annotated

- Non-text or structured text annotated with text
- Includes
  - Image collections
  - Citation databases
  - Evidence-based medicine databases
  - Clinical decision support
  - Omics databases
  - Other databases

# Image collections

- Most prominent in the “visual” medical specialties, such as radiology, pathology, and dermatology
- Come and go, but well-known collections include
  - Open-I – <https://openi.nlm.nih.gov/>
  - Radiopedia – <https://radiopaedia.org>
  - Visible Human – [https://www.nlm.nih.gov/research/visible/visible\\_human.html](https://www.nlm.nih.gov/research/visible/visible_human.html)
  - WebPath – <https://webpath.med.utah.edu/webpath.html>
  - More pathology – PEIR, <http://peir.path.uab.edu/library/>
  - DermIS – <http://dermis.net/dermisroot/en/home/index.htm>
  - More dermatology, also a decision-support system – <https://www.visualdx.com/>
- Many have associated text, which assists with indexing and retrieval

# Citation databases

- Science Citation Index and Social Science Citation Index
  - Database of journal articles that have been cited by other journal articles
  - Now part of a package called Web of Science, which itself is part of a larger product, Web of Knowledge (Clarivate)
    - <https://clarivate.com/webofsciencegroup/solutions/web-of-science/>
- SCOPUS – <https://www.elsevier.com/solutions/scopus>
- Google Scholar – <https://scholar.google.com/>

# Evidence-based medicine databases

- Cochrane Database of Systematic Reviews – <https://www.cochrane.org/>
  - Collection of systematic reviews, kept updated
- Evidence “formularies”
  - JAMAevidence – <https://jamaevidence.mhmedical.com/>
- AHRQ Evidence Reports – <https://www.ahrq.gov/research/findings/evidence-based-reports/>
- Many resources part of aggregations

# Clinical decision support (CDS)

- Content used in CDS systems, usually part of EHRs
  - Order sets (usually “evidence-based”)
  - CDS rules
  - Health/disease management templates
- Growing and evolving commercial market for such tools, especially as EHR adoption increases; leaders include
  - Zynx – <https://www.zynxhealth.com/>
  - Provation – <https://www.provationmedical.com/order-sets/>
  - EHR vendors themselves and partners

# Omics databases

- National Center for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov/>; Sayers, 2023) collection links
  - Literature references – MEDLINE
  - Textbook of genetic diseases – On-Line Mendelian Inheritance in Man (OMIM)
  - Sequence databases – Genbank
  - Structure databases – Molecular Modeling Database
  - Genomes – catalogs of genes
  - Maps – Locations of genes on chromosomes
  - Clinical associations – ClinVar
- More in bioinformatics unit...

# Other databases

- ClinicalTrials.gov
  - <https://clinicaltrials.gov/>
  - Originally database of clinical trials funded by NIH
  - Also used as register for clinical trials (DeAngelis, 2005; Laine, 2007; Zarin, 2016; Zarin, 2017)
  - And for results of trials (Zarin, 2019), although reporting incomplete (Nelson, 2023)
- NIH RePORTER
  - <https://reporter.nih.gov/>
  - Database of all research grants funded by NIH
- biomedical and healthCARE Data Discovery Index Ecosystem (bioCADDIE)
  - Database of metadata about available biomedical data sets
  - <https://datamed.org/>
  - <https://www.covid19dataindex.org/>

# Aggregations – integrating many resources

- Clinical – growing tendency of publishers to aggregate resources into comprehensive products
  - Univadas – <https://www.univadis.com/>
    - Formerly Merck Medicus, a collection of many resources available to any licensed US physician
  - Up to Date – <https://www.uptodate.com/contents/search>
  - Essential Evidence Plus (includes InfoPOEMS, “Patient-oriented evidence that matters”) – <https://www.essentialevidenceplus.com/>
  - Dynamed Plus – <https://www.dynamed.com/>

# Other aggregations

- Biomedical research: Model organism databases, e.g., Mouse Genome Informatics
  - <http://www.informatics.jax.org/>
  - Combines genomics and related data, bibliographic database, gene references, etc.
- Consumer: MEDLINEplus
  - <https://medlineplus.gov/>
  - Integrates a variety of licensed resources and public Web sites