



Information Retrieval Content

What is Biomedical and Health Informatics? - <http://informatics.health/>
William Hersh
Copyright 2024
Oregon Health & Science University



1

Classification of knowledge-based content

- Bibliographic
 - By definition rich in metadata
- Full-text
 - Everything on-line
- Semi-structured
 - 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

WhatIs11

2



2

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

WhatIs11

3



3

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>

WhatIs11

4



4

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
- 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
 - Contains some additional content not in MEDLINE, e.g., PubMed Central and books

WhatIs11

5



5

MEDLINE statistics

- https://www.nlm.nih.gov/bsd/medline_pubmed_production_stats.html

	FY2022	FY2021	FY2020	FY2019	FY2018
MEDLINE Citations Indexed (Annual)	1,369,611	1,291,807	952,919	956,390	904,636
MEDLINE Citations Cumulative Total	29,807,639	28,444,654	27,149,277	26,196,358	25,239,968
MEDLINE Journal Titles	5,282	5,282	5,274	5,243	5,251
PubMed Citations (Annual)	1,714,780	1,733,089	1,514,199	1,366,447	1,329,148
PubMed Citations Cumulative Total	34,693,538	33,136,289	31,563,992	30,178,674	28,934,389
PubMed Searches	2.58 Billion	2.57 Billion	3.3 Billion	3.1 Billion	3.3 Billion
Web/Interactive	1.283 Billion	1.186 Billion	1.076 Billion	896 Million	831 Million
Script/E-Utilities	1.303 Billion	1.391 Billion	2.2 Billion	2.2 Billion	2.5 Billion

WhatIs11



6

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

WhatIs11

7



7

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/>

WhatIs11

8



8

RSS – bibliographic feeds (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

WhatIs11

9



9

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

WhatIs11

10



10

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)

WhatIs11

11



11

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/>

WhatIs11

12



12

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

WhatIs11

13



13

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.

MERCK MANUALS Professional Version

Search

Find information on medical topics, symptoms, drugs, procedures, news and more, written for the health care professional.

Professional / Endocrine and Metabolic Disorders / Lipid Disorders

Dyslipidemia

Dyslipidemia (Hyperlipidemia)

by Anne Carol Goldberg, MD

NOTE: This is the Professional Version. **CONSUMERS:** Click here for the Consumer Version

Dyslipidemia is elevation of plasma cholesterol, triglycerides (TGs), or both, or a low high-density lipoprotein level that contributes to the development of atherosclerosis. Causes may be primary (genetic) or secondary. Diagnosis is by measuring plasma levels of total cholesterol, TGs, and individual lipoproteins. Treatment involves dietary changes, exercise, and lipid-lowering drugs.

There is no natural cutoff between normal and abnormal lipid levels because lipid measurements are continuous. A linear relation probably exists between lipid levels and cardiovascular risk, so many people with "normal" cholesterol levels benefit from achieving still lower levels. Consequently, there are no numeric definitions of dyslipidemia; the term is applied to lipid levels for which treatment has proven beneficial. Proof of benefit is strongest for lowering elevated low-density lipoprotein (LDL) levels. In the overall population, evidence is less strong for a benefit from lowering elevated TGs and increasing low high-density lipoprotein (HDL) levels.

HDL levels do not always predict cardiovascular risk. For example, high HDL levels caused by some genetic disorders may not protect against cardiovascular disorders, and low HDL levels caused by some genetic disorders may not increase the risk of cardiovascular disorders. Although HDL levels predict cardiovascular

Lipid Disorders

- Disorders of Lipid Metabolism
- Dyslipidemia
- Disorders of Lipid Metabolism

WhatIs11

14



14

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.)

WhatIs11

15



15

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/>

WhatIs11

16



16

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

WhatIs11

17



17

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.

WhatIs11

18



18

Semi-structured content

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

WhatIs11

19



19

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
 - 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

WhatIs11

20



20

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/products/scientific-and-academic-research/research-discovery-and-workflow-solutions/webofscience-platform/>
- SCOPUS – <https://www.elsevier.com/products/scopus>
- Google Scholar – <https://scholar.google.com/>

WhatIs11

21



21

Evidence-based medicine content

- 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

WhatIs11

22



22

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

WhatIs11

23



23

Computational biomedical knowledge (CBK)

- Digital objects representing biomedical knowledge in machine-interpretable structures
 - <https://mobilizecbk.med.umich.edu/>
- Requires technical infrastructure to support and maintain (McCusker, 2023)
- Metadata categories describe type, domain, location, technical characteristics, authorization, and evidence basis (Alper, 2022)
- Small but growing number of CBK repositories (Platt, 2023)

WhatIs11

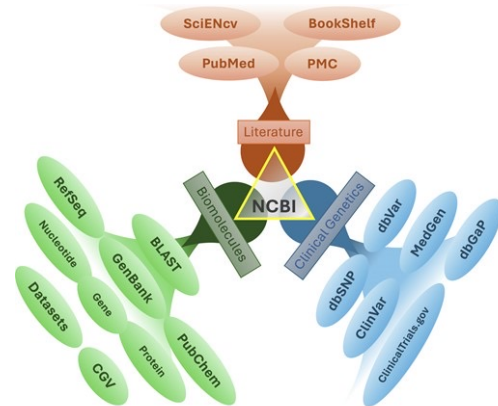
24



24

Omics databases

- National Center for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov/>; Sayers, 2024) collection links
 - Literature references – MEDLINE
 - Textbook of genetic diseases – On-Line Mendelian Inheritance in Man (OMIM)
 - Sequence databases – Genbank
 - Structure databases
 - Genomes – catalogs of genes
 - Clinical associations – ClinVar
- More in bioinformatics unit...



WhatIs11

25



25

Data repositories

- With growing publication of data, need methods to discover and access metadata
- Journal articles and reports with published data may
 - Make data available as supplement on journal site
 - Provide instructions to request from author
- One of largest publishers of data is US government, e.g.,
 - <https://data.gov/>
 - <https://www.datasetcatalog.nlm.nih.gov/>
- Many other sites for data; some include
 - biomedical and healthCare Data Discovery Index Ecosystem (bioCADDIE)
 - <https://datamed.org/>
 - <https://www.covid19dataindex.org/>
 - New York University (NYU) Data Catalog (Yee, 2023)
 - <https://datacatalog.med.nyu.edu/>
 - Hugging Face – datasets and models for machine learning research
 - <https://huggingface.co/>

WhatIs11

26



26

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

WhatIs11

27



27

Aggregations – integrating many resources

- Clinical – growing tendency of publishers to aggregate resources into comprehensive products
 - 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/>

WhatIs11

28



28

Aggregations (cont.)

- 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

WhatIs11

29

