



# Information Retrieval Content

What is Biomedical & Health Informatics?  
William Hersh  
Copyright 2022  
Oregon Health & Science University



1

## 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**

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
  - Free to world since 1997 via PubMed – <https://pubmed.gov/>
    - Now with links to full text of articles and other resources
    - PubMed contains some additional content not in MEDLINE
- Statistics
  - [https://www.nlm.nih.gov/medline/medline\\_overview.html](https://www.nlm.nih.gov/medline/medline_overview.html)
  - Over 28M references to peer-reviewed literature
  - Over 5200 journals, mostly English language
  - Nearly 1M new references added yearly

WhatIs11

5



5

# 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

6



6

## 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

7



7

## 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

WhatIs11

8



8

## 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

9



9

## 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

10



10

## 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/>

WhatIs11

11



11

## 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

12



12

## 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 interface. The main content area displays the article for "Dyslipidemia (Hyperlipidemia)" by Anna Carol Goldberg, MD. The article includes a definition, classification, etiology, symptoms and signs, diagnosis, treatment, and key points. A sidebar on the right lists related topics such as "Lipid Disorders", "Disorders of Lipid Metabolism", "Dyslipidemia", and "Hyperlipidemia".

WhatIs11

13

13

## 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

14

14

## 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/DiseasesConditions/>
    - <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

15



15

## 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/>
  - WebMD – <https://www.webmd.com/>
- Many mobile apps provide health information, e.g.,
  - Epocrates – <https://www.epocrates.com/>
  - WebMD app for consumers

WhatIs11

16



16



## 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

17



17

## 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

WhatIs11

18



18

## 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

WhatIs11

19



19

## 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/>

WhatIs11

20



20

## 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

WhatIs11

21



21

## 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

22



22

## Omic databases

- National Center for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov/>; Sayers, 2022) 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...

WhatIs11

23



23

## Other databases

- ClinicalTrials.gov
  - <https://clinicaltrials.gov/>
  - Originally database of clinical trials funded by NIH
  - Now used as register for clinical trials, with results reporting for some (DeAngelis, 2005; Laine, 2007; Zarin, 2016; Zarin, 2017; Zarin, 2019)
- 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/>

WhatIs11

24



24

## 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/>

WhatIs11

25



25

## 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

WhatIs11

26



26