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 admittedly 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., National Guidelines Clearinghouse) have emerged
- Web catalogs
  - Share many characteristics of traditional bibliographic databases
- Real simple syndication/Rich site summary (RSS)
  - “Feeds” provide information about new content

Bibliographic databases

- Contain metadata about (mostly) journal articles and other resources typically found in libraries
- Produced by
    - e.g., MEDLINE, genomics 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
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 –
    http://pubmed.gov
    • Now with links to full text of articles and other resources
    • PubMed contains some additional content not in MEDLINE
  • Statistics
    – Over 24M references to peer-reviewed literature
    – Over 5600 journals, mostly English language
    – Nearly 900,000 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
  - HONSearch for professionals – https://www.hon.ch/HONsearch/Pro/
  - Translating Research into Practice – https://www.tripdatabase.com/
- Others are aimed towards patients/consumers
  - HONSearch for patients/consumers – https://www.hon.ch/HONsearch/Patients/hunt.html

RSS

- 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.
- Two versions (1.0, 2.0) but basically provide
  - 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)
  – Another source of full-text papers is PubMed Central (PMC; https://www.ncbi.nlm.nih.gov/pmc/)
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.
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/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/
Full-text Web sites (cont.)

• Physician-oriented medical news and overviews, e.g.,
  – 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.,
  – WebMD app for consumers

Other interesting types of Web content

  – Encyclopedia with free access and distributed authorship
  – Some concerns about manipulation (McHenry, 2004) but
    • Health information quality is reasonably good (Nicholson, 2006)
    • Content retrieved prominently in most Web searches (Laurent, 2009)
    • Making attempt to improve quality of medical content (Heilman, 2013)

• 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/Web 2.0 and beyond
Annotated

- Non-text or structured text annotated with text
- Includes
  - Image collections
  - Citation databases
  - Evidence-based medicine databases
  - Clinical decision support
  - Genomics 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
  - WebPath – https://webpath.med.utah.edu/webpath.html
  - 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)
- Google Scholar – [https://scholar.google.com/](https://scholar.google.com/)

Evidence-based medicine databases

- Cochrane Database of Systematic Reviews
  - [https://www.cochrane.org/](https://www.cochrane.org/)
  - Collection of systematic reviews, kept updated
- Evidence “formularies”
- 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

Genomics databases

  - Literature references – MEDLINE
  - Textbook of genetic diseases – On-Line Mendelian Inheritance in Man
  - Sequence databases – Genbank
  - Structure databases – Molecular Modeling Database
  - Genomes – Catalog of genes
  - Maps – Locations of genes on chromosomes
- More in bioinformatics unit...
Other databases

- ClinicalTrials.gov
  - [https://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)

- NIH RePORTER
  - [https://projectreporter.nih.gov/reporter.cfm](https://projectreporter.nih.gov/reporter.cfm)
  - Database of all research grants funded by NIH
  - Replaced the CRISP database

- biomedical and healthCAre Data Discovery Index Ecosystem (bioCADDIE)
  - Database of metadata about available biomedical data sets
  - [https://datamed.org/](https://datamed.org/)

- Online Registry of Biomedical Informatics Tools (ORBIT)

Aggregations – integrating many resources

- Clinical – growing tendency of publishers to aggregate resources into comprehensive products
  - Univadas – [https://www.univadis.com/](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/home](https://www.uptodate.com/home)
    - Very popular among clinicians
Other aggregations

- Biomedical research: Model organism databases, e.g., Mouse Genome Informatics
  - Combines genomics and related data, bibliographic database, gene references, etc.

- Consumer: MEDLINEplus
  - [https://medlineplus.gov/](https://medlineplus.gov/)
  - Integrates a variety of licensed resources and public Web sites