Chapter 5 Discovering scientific information

Note: This is a non-edited version of this chapter. This is the official citation of the published version:

Baykoucheva, S. (2022). 5 - Discovering scientific information. In S. Baykoucheva (Ed.), Driving Science Information Discovery in the Digital Age (pp. 85-110). Chandos Publishing. https://doi.org/10.1016/B978-0-12-823723-6.00002-1

Abstract

With so much information available on the Internet, it is difficult for users to figure out which resources are the most appropriate to use. Before going to a bibliographic indexing database, they often do a preliminary search in Google Scholar or in some other popular search engine. A large volume of searches, though, are now also occurring on Baidu Scholar, Dimensions, so.com, x.mol.com, and The Lens—newcomers that are beginning to play a critical role on the information landscape. As the integration of diverse information resources on a single platform is becoming a norm, bibliographic literature databases are sharing "home" with resources for finding properties of chemical compounds, patents, book catalogs, and other resources. This chapter presents different scenarios and strategies for discovering scientific information more efficiently. An overview of the current scholarly communication formats is followed by examination of the major science information resources. The process of discovering scientific information is presented through specific examples of searching for literature, properties of chemical compounds, research datasets, and other scientific information; refining and analyzing search results; and using bibliographic management tools to export, store, and cite references. The chapter also looks at some informal ways of keeping up with the most recent research, using social media and networking.

Keywords

bibliographic databases, bibliographic management, chemical information, datasets, discovering scientific information, indexing, information retrieval, patents, scientific journals, scientific information, search engines

Introduction

> *Selecting relevant information and suppressing details is the sort of pragmatic fudging everyone does every day. It's a way of coping with too much information. For almost everything you see, hear, taste, smell, or touch, you have the choice between examining details by scrutinizing very closely and looking at the "big picture" with its other priorities.*
>
> Warped Passages: Unraveling the Mysteries of the Universe's Hidden Dimensions (Randall, 2005)

With the advances in digital technologies, the number of resources we can choose from has grown significantly. Search engines, indexing services, and library catalogues are competing in offering more sophisticated 'smart tools' for discovering and managing scientific information. In the past, these resources were either not existent or could not be accessed the way we do it today. Data Planet, for example—a statistical data repository—has collated and organized billions of datasets into a common structure in a user-friendly environment, providing immediate access to data presented in charts, maps, graphs, and tables that can be searched from multiple points of entry (DataPlanet, 2020). Without these new technologies, researchers would have spent many hours trying to answer a question that can easily be answered today.

How we go about discovering scientific information depends on many factors—how quickly we need it, how comprehensive it should be, and which resources and discovery tools are available to us. We are accustomed to using search engines and bibliographic databases to find literature. In some cases, we need information about a chemical compound that is summarized in one place rather than dispersed in many different papers. We may need to find full text articles or obtain some specific information buried in a dataset.

To use an analogy from Lisa Randall's quote above, we could live with one-dimensional information—a bibliographic description of an article (title, authors, abstract), or we could look for other dimensions (full text, references cited in a paper, or similar articles). We could even use Web of Science to perform a cited reference search for an article published in the past and see how often this paper was cited and by whom, how the concepts expressed in the article have evolved through time, and in which areas of science the paper has made an impact.

This chapter presents different scenarios and strategies for discovering scientific information more efficiently. An overview of the current scholarly communication formats is followed by examination of the major science information resources. The process of discovering scientific information is presented through specific examples of searching for literature, properties of chemical compounds, research datasets, and other scientific information; refining and analyzing search results; and managing them with bibliographic management tools. The chapter also looks at some informal ways of keeping up with the most recent research, using social media and networking.

Communication formats

How researchers communicate their findings varies and depends on the discipline and other factors. The common scholarly communication formats and some of the alternative ways they use to share their work are presented below.

Science communication formats:

- Books, Book chapter
- Clinical trials
- Conferences
- Datasets
- Dissertations
- Journal articles
- Patents
- Preprints
- Technical reports

Alternative communication formats:

- Blogs, Web sites
- News articles
- Social Networks (Facebook, Instagram, Reddit, Twitter, Mendeley)
- Works posted on an institutional repository
- Academic social sites (Academia.edu, ResearchGate, ResearcherID)
- Researchers' profiles (on Academia.edu, Google Scholar, ResearchGate, Instagram, LInkedIt, ResearcherID)
- Listservs (such as CHMINF-L, the chemical information sources discussion list, which was started in 1991 (Baykoucheva, 2015c)

Research articles (also called *primary research articles*) report results obtained through direct or indirect observation and research. They include quantitative or qualitative data and analysis and are usually organized in the following sections: Abstract, Introduction (may be called "Background"), Materials and Methods (also called "Methods" or "Experimental"), Results, Discussion, Conclusion, and References. The Results and Discussion sections are often combined in one section (Results and Discussion), and the Conclusion section may also be added to them. Many journals now require that authors provide raw data and methodology details as supplementary material, which becomes part of the publication. Unfortunately, the supplementary material is usually not indexed and is not parsed by the search engines. Peer-reviewed (or refereed) articles go through a rigorous review process, often including revisions to the original manuscript, by peers in the same area of research, before publication in a scientific journal. In addition to the research and review articles, scholarly journals also publish editorials, letters, and other items that are not peer reviewed.

Review articles present an overview of already published primary research articles. They summarize, analyze, and critically evaluate the important literature and give us the "big picture" on a particular topic. Sometimes, review articles may be peer-reviewed, and they usually do not include an experimental section.

Systematic reviews are a special kind of reviews. They summarize the existing clinical research on a particular topic and include meta-analyses of the effects of the treatment, judgments about the quality of the studies and recommendations on the potential application of the results in healthcare.

Clinical trials are studies in biomedical or behavioral research designed to address specific questions about interventions such as new treatments (e.g., new vaccines, drugs, nutrition, dietary supplements, and medical devices).

Conferences and departmental seminars allow researchers to share their work with their peers and get a feedback about their work before submitting a paper to a journal. In some disciplines (computer science, engineering), conferences are the first (and sometimes the only) way of reporting and publishing their newest research findings. Some of the conference proceedings are peer reviewed, but others are not. Many bibliographic indexed databases cover conference abstracts and conference proceedings (full text).

Patents report innovative techniques, new drugs, and therapeutic effects.

Preprints are manuscripts posted by scientists on *preprint servers* prior to submitting them for publication in a journal. Making results available more quickly accelerates the research process and improves the quality of the journal literature, because authors can receive feedback from peers prior to official publication. A list of preprint servers is included later in this chapter.

Selecting information resources

Some of the most important considerations for choosing a resource are the coverage and the options for filtering and managing the retrieved literature. As science is becoming more interdisciplinary and data-rich, publishers and libraries integrate several databases on a single platform—literature databases are sharing a platform with chemical property and reaction databases, patent sources, and book catalogs. When looking for properties of a chemical compound, a product of a chemical reaction, or another specific information, having this information summarized in one place rather than trying to extract specific data from different papers would save us a lot of time. Users can manage the retrieved results with bibliographic management tools such as EndNote, Mendeley, and Zotero, easily accessible from the same platform.

Aggregation of resources on a single platform allows accessing diverse content at the same time. For example, when searching for properties of a chemical compounds in PubChem or SciFinder, users can find also articles discussing this compound. Academic search engines and bibliographic indexed databases are connecting users to full text articles. Some academic social networks, such as Academia.edu, Mendeley, ResearcherID, ResearchGate, also play the role of information resources, where authors can post their articles in full text. Many commercial publishers are taking note about something called Open Science/Open Access and are giving up more and more of their proprietary turf, for which they were previously charging significant fees.

Knowing which resources would be the most appropriate will save time to researchers. Those working in the life sciences and the biomedical field, for example, are not aware that they can find more information using SciFinder than using only MEDLINE or PubMed. They benefit from performing the searches in SciFinder is that they will be searching simultaneously two very large databases—MEDLINE (the indexed part of PubMed) and CAplus (the large Chemical Abstracts databases), which complement each other in terms of content. SciFinder covers more document types (patents, books, and dissertations) than PubMed.

Figure 5.1 summarizes the most important science information resources, grouped in seven categories: Search Engines, Bibliographic Databases, Full Text Databases, Open Access Resources, Data Repositories, Chemistry Resources, and Bibliometric Databases. This organization is conditional and is made just for convenience, as some of these resources fall into more than one of these categories. For example, a resource that is promoted as Open Access may not provide all of its content for free.

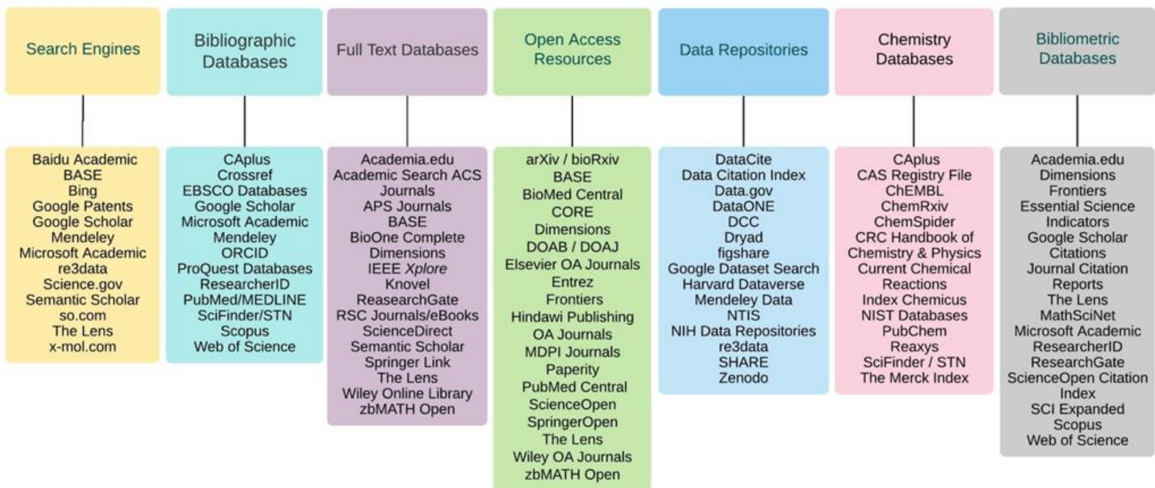

Figure 5.1 Science information resources, classified in the following categories: Search Engines, Bibliographic Databases, Full Text Databases, Open Access Resources, Data Repositories, Chemistry Databases, and Bibliometric Databases. BASE (Bielefeld Academic Search Engine); SHARE (SHared Access Research Ecosystem); re3data (Registry of Research Data Repositories); NTIS (National Technical Reports Library); ACS (American Chemical Society; APS (American Physical Society); RSC (Royal Society of Chemistry); SCI (Science Citation Index) Expanded; CORE (COnnecting REpositories); DCC (Digital Curation Centre.

Search engines

Search engines use crawlers—programs that parse the Web by following hypertext links from page to page—to retrieve information. Even the most extensive general search engines cannot keep up with the extensive proliferation of Web pages, and they leave many documents uncovered. The information on the search engine pages is of varying quality and is based on weighting and ranking the documents by certain criteria. Academic search engines provide information on a range of topics—from engineering and technology—to biology and natural sciences. They also allow customizing the way to search and retrieve information on a given topic. While there are many academic search engines available, there are some that are more trusted than others. It is a good idea to examine the features of the search engines to be able to use them more efficiently.

The retrieved results depend significantly on the keywords of the query and the order in which they are entered in the search box of the search interface. Most search engines allow the user to join terms with *and*, *or*, and *not* to refine queries. Search engines and indexed databases automatically suggest possible search terms based on what the user has already typed in the search box and some of them allow downloading content directly from the search results. Some search engines also provide free alert services that keep users up to date for newly published papers and other documents in their areas of interest. Knowing how search engines rank results would benefit users allowing them to retrieve relevant results more efficiently. Search Engine Optimization (SEO), a process aimed at improving the quality and quantity of website traffic, has become of great interest in the last decade. This technology is discussed in Chapter 3 of the book. More detailed information about search engines can be found elsewhere (Leung et al., 2019).

When looking for scientific information, researchers usually start with a search in an academic search engine such Google Scholar or Microsoft Academic, but a considerable volume of searches also occurs on regular search engines such as Google and Bing. Some of the academic search engines are presented below:

BASE is an academic search engine hosted at Bielefeld University in Germany and that is where its name comes from (Bielefeld Academic Search Engine). BASE combines entries from thousands of institutional repositories in one place.

CORE (COnnecting REpositories) is an open access academic search engine dedicated to open access research papers. For each search result a link to the full text PDF or full text web page is provided. It provides abstracts, cited by, and links to full text. All articles in CORE are open access. CORE is pursuing the goal of aggregating all open access content from different systems (repositories, open access journals) and increase the value of this content by adding text and data mining capabilities to find more relevant information. These features are what distinguishes CORE from other academic search engines like Google Scholar and Microsoft Academic Search, whose services remain only at the search and access level and are not capable of supporting the further use of the open access content for such purposes as text and data mining.

Directory of Open Access Journals (DOAJ) is a free search engine for scientific and scholarly resources. The directory offers a huge range of topics within scientific areas of study. All the journals that it covers are thoroughly peer-reviewed.

DuckDuckGo is a general search engine that retrieves very relevant and "right to the point" scholarly publications and other scientific information without the noise other more popular search engines are pushing to users.

Google Scholar (GS) is the number one choice when it comes to academic search engines. It is multidisciplinary search engine for finding scientific literature. It not only lets you find research papers for all academic disciplines for free, but also often provides links to full text PDF files. Authors can create a personal profile, which will appear at the top of the list of the search results when people search for this author's name. GS Includes several citation metrics and shows graphically citation analysis data.

Microsoft Academic is a search engine that uses a semantic search technology to understand what users are looking for. It does not just match keywords to content but understands the meaning of words. It indexes a range of scientific journals—from computer science and engineering—to social science and biology and generates for each paper that is indexed an overview page that allows exploring the top citing articles and references of the article.

Semantic Scholar is a search engine for academic publications which has the goal of providing more relevant and impactful search results. It is backed by AI-powered algorithms that find hidden connections and links between research topics and uses recent advances in natural language processing to provide summaries of scholarly papers.

Regional search engines such as Baidu, Baidu Scholar, so.com, and x.mol.com (discussed below) are becoming important information channels and are playing a critical role today on the science information landscape.

Baidu Scholar. The first time I heard about Baidu Scholar was a few years ago, when a graduate student came to me with a chemistry question. As she opened her laptop, I noticed that she was using an unfamiliar to me database to find some articles. The name she mentioned was "Baidu." Baidu is the dominant Chinese internet search engine company—a rival of Google in China. And as Google has Google Scholar—Baidu has Baidu Scholar, a free research platform providing access mainly to Chinese, but also to foreign literature. The interface of Baidu Scholar is available only in Chinese, but users can enter queries in English or other languages (Figure 5.2-Figure 5.4).

Figure 5. 2 The main interface of Baidu Scholar is in Chinese, but queries can be entered in English. Date of access: March 29, 2021. [Copyrighted material removed]

Figure 5.3 Example of Baidu Scholar's search results page. Date of access: March 29, 2021. [Copyrighted material removed]

Figure 5.4 A record of an article as displayed in Baidu Scholar. Date of access: March 29, 2021. [Copyrighted material removed]

Scientific publishers are interested in signing agreements with Baidu Scholar, which would allow their publications to be discovered and accessed by a wider audience. An article, posted on Wiley's web site ("Discoverability in China: Why Baidu Scholar is good news for researchers") discussed the importance of Baidu Scholar for researchers (and for publishers!) (Eassom, 2021).

360 Search**.** The most challenging competitor to Baidu is Qihoo 360—originally, an anti-virus company. Its search engine, *360 Search* (with the domain *so.com*), quickly became the second most popular search engine in China. The interface of *360 Search* can be translated into English (Figure 5.5). When I bravely entered my name in a search box on an interface that I could not understand, I was up for a surprise. The search engine of 360 Search (so.com) retrieved even some obscure abstracts of posters that I had presented at my university years ago. It also found an article that I had forgotten that I had even written. The search engine parses extensively institutional repositories, web pages, LibGuides, professional listservs (it even showed an email that I had sent to a professional listserv with an announcement for a job position).

Figure 5.5 The translated interface of the Chinese search engine "360 Search" with the domain "so.com." Date of access: March 28, 2021. [Copyrighted material removed]

Bibliographic indexing databases

Bibliographic indexing databases are organized digital collections, which include references (bibliographic records) of literature. The references can be of journal articles, conference proceedings, patents, and other documents. A bibliographic database can be of interdisciplinary or discipline-oriented type. Examples of interdisciplinary bibliographic databases are Scopus and Web of Science. The discipline-oriented databases are covering content specific for a particular discipline and for other disciplines related to it. PubMed/MEDLINE (the largest database for biomedical literature) and the Chemical Abstracts database (CAplus) are examples of such domain-specific bibliographic indexed databases. Depending on the indexing practices of the particular database, the indexing can be performed at different levels such as title, abstract, full text, references, figures, tables, and images. The elements that are indexed usually include articles, conference proceedings, graphs, tables, images, DOIs, authors names, ORCID, ResearcherID, and other unique identifiers.

Although most of the scientific bibliographic databases include articles, rather than monographs and monograph chapters, there are databases that also cover books, book chapters, clinical trials, conference proceedings, dissertations, patents, preprints and other types of documents. Web of Science even covers articles retracted from journals. Some bibliographic databases have transformed into digital libraries, providing full text of the indexed documents. Others aggregate many different databases that can be searched either individually or together.

In the past, when searching in a library catalog, you would usually find books. Users can now type keywords, titles, and authors' names in the search box of a library catalog—and find articles and other documents that they were previously able to find only when searching a bibliographic indexed database. This technology still needs improvement, because searching for articles in a library catalog is not always working well. Many of the bibliographic databases are proprietary and do not share their information. They provide more sophisticated "smart" tools to find articles, conference proceedings and other documents than the library catalogs.

Although the general patterns of searching and refining search results have common features, each database also has specific ways and tools for performing a search. Knowing how to use the capabilities and the features of the resources would make a big difference with respect to what information you discover. Some of the general rules for searching for scientific information are listed below:

- Use reliable and authoritative sources for finding information.
- Use appropriate terms (keywords, phrases, natural language, document identifier to describe what you are looking for.
- Use natural language, when this option is available. This provides some context for what you are looking for.
- Arrange the search terms by their significance for the search. The order in which they are entered may affect the results.
- Look at the search terms suggested by the search engine or indexing service.

After performing a search, it is very important to narrow down (filter) the results. Depending on the resource, users are provided with different options for searching and refining search results—for example, searching for literature by using keywords, phrases, index terms, or natural language. Not all databases allow using natural language. When searching for literature in a bibliographic indexed database, we can use the following options:

- *Research Topic:* key words, phrases, or natural language. Words such as “and,” “of,’ “for” (called stop words) are ignored by indexers.
- *Author Name:* Last Name, First name, and Middle Initial. Look for alternate spellings of author’s name. Unique author identifiers.
- *Company Name:* Name of a company, university, organization, or institution
- *Document Identifier:* DOI, Publication Year, Volume, Issue
- *Source (Journal) Title:* Full journal name or its abbreviation, Title Words, Author Name, Publication Year.

Searches can be analyzed and refined using specific criteria. For an unfamiliar topic, it would be useful to refine search results to review articles. Citations included in the review articles provide a quick and easy way to find the most important literature published on the topic. Strategies for finding scientific information using the most common science information resources are described in more detail in another publication (Baykoucheva et al., 2016).

Some of the most important bibliographic indexing databases are presented below.

BioMed Central (part of Springer Nature) is a pioneer in Open Access publishing. It is a platform with some 300 peer-reviewed journals in science, technology, engineering and medicine. Authors retain copyright of their work through a Creative Commons attribution license that allows readers to copy, distribute, and use their research free of charge.

BioOne Complete (Published by the nonprofit publisher BioOne) is a database of more than 200 subscription-based and open-access peer-reviewed journals in the biological, ecological, and environmental sciences. It provides libraries with cost-effective access to high-quality curated scientific information on a single platform.

CERN Document Server (from CERN, the European Organization for Nuclear Research) is a database produced by one of the world's largest and most respected centers for scientific research. It indexes journal articles, technical reports, preprints, books, software, and multimedia content in the field of particle physics and related fields. Materials date back to the mid-1900s with full coverage from 1980 to present. Some of the material is available in full text.

Crossref (Crossref, 2021) interlinks millions of various content types, including journals, books, conference proceedings, working papers, technical reports, and datasets. The content includes materials from scientific, technical, medical, social sciences and humanities. Crossref provides the infrastructure for this reference linking using Digital Object Identifiers (DOIs). It also manages the deposit and query service for DOIs.

Dimensions is part of Digital Science (*Dimensions*, 2021). It provides a collection of linked data on a single platform. All publications and citations in it are freely available for personal, non-commercial use, together with rich contextual information. The moto of Dimensions that we see on its front page says: "Linked research data from idea to impact." Its creators have created an open infrastructure, which resulted in a database with a huge collection of linked data (from grants, journal articles, datasets, clinical trials, patents, and even policy documents)—on a single platform. Dimensions is changing our understanding of what a database should be, transforming how research is discovered, accessed, and evaluated. The types of linked resources in Dimensions are shown in Figure 5.6.

Figure 5.6 Types of linked resources in Dimensions (Date of access: March 20, 2021). [Copyrighted material removed]

EBSCO Information Services is a multidisciplinary platform (published by EBSCO), which aggregates many bibliographic indexed databases from nearly every area of academic study. It includes titles compiled by the company as well as journals from other databases/publishers such as MEDLINE and EconLit. The company also maintains Academic Search Complete, which covers the full text of many journals and includes a large collection of peer-reviewed full text articles. A list of the databases on the EBSCO platform is available on the publisher site (EBSCO, 2021).

Energy Citations Database includes scientific and technical research results in disciplines of interest to the US Department of Energy such as chemistry, physics, materials, environmental science, geology, engineering, mathematics, climatology, oceanography, and computer science

Entrez (From the National Library of Medicine (NLM)) is a large information retrieval system that includes many databases. Its user interface allows rapidly searching sequence and bibliographic data. Precomputed similarity searches for each record create links to related records in other Entrez databases, facilitating integrated access across the various databases. Results may be viewed in various formats and a graphical interface provides visualization of full genomes or chromosomes, as well as biological annotation on individual sequences. Entrez also allows downloading large volume of search results at the same time.

Faculty of 1000 (F1000) is a publisher of services for life scientists and clinical researchers with a unique publication model. ***F1000Research*** is an open access, open peer-review scientific publishing platform focused on the life sciences. Articles are peer-reviewed after they are published. The reviewers' names and their comments are made public. F1000 was acquired by Taylor & Francis Group in 2020.

High-Energy Physics (HEP) Literature Database is an open access resource based primarily on pre-prints but also links to the final journal article citation and full text, when it is available.

IEEE Xplore is a digital library for accessing scientific and technical content. It covers books, conferences, courses, journals, magazines, and standards. IEEE Xplore is published by the IEEE (Institute of Electrical and Electronics Engineers).

MDPI Open Access journals are published by MDPI (Multidisciplinary Digital Publishing Institute), the largest open access publisher in the world. Founded as a chemical archive, it now has over 200 journals of broad scope. The number of published papers has been growing significantly in the last decade. Most of the journals are Open Access.

New Journal of Physics is an important open access platform that provides content in all areas of physics. It was co-founded by the Institute of Physics and Deutsche Physikalische Gesellschaft.

Nuclear Science References (from the Brookhaven National Laboratory, U.S. Department of Energy) indexes nuclear physics articles containing defined nuclear structures or reaction data. The database is open access and contains references to journal articles published since 1910. It currently covers about eighty journals.

PLoS is a platform of the Public Library of Science—a nonprofit, open access publisher for peer-reviewed journals in science, technology, and medicine.

PubMed (from the U.S. National Library of Medicine of the National Institutes of Health) is a free biomedical resource. It is based on the MEDLINE database. PubMed is the largest database for biomedical literature. It is a major component of the NLM suite of resources. The database underlying PubMed is MEDLINE (i.e., MEDLINE is a subset of PubMed), which uses a controlled vocabulary thesaurus (MeSH Index) to index documents. The "Advanced" search option in PubMed provides "smart" indexing tools to retrieve only articles that are directly related to the topic of the search, rather than only mentioned in a document. Using the MeSH indexing system to perform a search may not find all relevant results, if the index term used by the authors is not included in the thesaurus.

MEDLINE is a biomedical database with a controlled vocabulary system (a thesaurus) called the Medical Subject Headings (MeSH). PubMed includes more content than MEDLINE, including many articles from generalist journals like *Nature* and *Science*. Users can search for literature by just typing keywords. This general type of search will retrieve all documents that have just mentioned the search terms. To avoid getting too many results and retrieve only documents that are especially devoted to the topic of interest, it would be better to use the Advance search option, which is based on the MeSH indexing capabilities of MEDLINE, the indexed segment of PubMed. The "Advanced" search option in PubMed provides "smart" indexing tools to retrieve only articles that are directly related to the topic of the search, rather than only those that just mention the keywords entered in the search box on the main interface.

PubMed Central is produced by the National Library of Medicine (NLM). It is a database of citations, abstracts and full text of publications in the Life Sciences. It is a post-print repository for publications of research funded by the U.S. National Institutes of Health.

Paperity is a multidisciplinary aggregator of Open Access journals and papers.

Science.gov is one resource to access millions of papers and reports from several federal agencies. It bundles and offers free access to search results from more than 15 U.S. federal agencies. It allows searching the resources from these agencies at the same time, providing abstracts and, for some databases, links to full text.

ScienceDirect is a peer-reviewed, full text database from Elsevier, which covers electronic books, book chapters and journal titles in all the major areas of science, including computer science and technology, biology, medicine, environmental science and social sciences, and other areas of academic research. It has an extensive coverage of the physical and biological sciences. ScienceDirect is among the most trusted databases for scientific research.

Scitation is the electronic journal publishing platform of the American Institute of Physics (AIP). It includes a database of citations to AIP publications (journals, conference proceedings, and standards) and links to the actual articles, which are available to subscribers). It also includes the references to all articles included in the database and links to the online versions of those publications. It covers peer-reviewed journals, including a growing number of open access titles from all areas of the physical sciences.

Scopus (published by Elsevier) is the largest indexed database for peer review and other literature. It covers significantly more scientific journals than other indexed scientific databases(Baykoucheva, 2010a, 2015b; Huang, 2018; Martin-Martin et al., 2021; Pranckutė, 2021). Scopus indexes journals, conference papers, trade publications, book series, patents from the fields of science, technology, medicine, social sciences, and arts and humanities.

Scopus also provides research analysis and tracking tools. It displays "Mendeley readership statistics" for any article that was downloaded to the bibliographic management program Mendeley. The statistical information includes demographics (e.g., countries), disciplines, and the career levels of the people who have downloaded the article. Scopus Journal Analyzer ranks scientific journals using different criteria using four types of quality measure for each title: the h-Index, CiteScore, SJR (SCImago Journal Rank) and SNIP (Source Normalized Impact per Paper). Scopus also offers author profiles which cover affiliations, number of publications and their bibliographic data, references, and details on the number of citations each published document has received.

SpringerLink is published by Springer Nature. It is a platform that carries online journals in Biomedical and Life Sciences, Chemistry and Materials Science.

Springer Open includes more than 200 peer-reviewed open access journals in all areas of science.

The Lens is a search engine that is collecting patents, articles, books, and theses from different places and presents the search results in a unified system, while preserving their original location (The Lens, 2021). It is using new technologies and efficient filters, facets, visualization tools and export features, which allow examining and analyzing the data in a more complex way. The goal of The Lens is to make the information freely available and become a global infrastructure for a more efficient information discovery. Some see The Lens as a future competitor to the "Big Ones"—Scopus and Web of Science.

Web of Science (WoS) is published by Clarivate Analytics. It is a very large interdisciplinary database covering the journal literature of the sciences. It is also an online subscription-based scientific citation indexing service originally produced by the Institute for Scientific Information (ISI), later owned by Thompson Reuters, and currently provided by Clarivate Analytics (previously, the Intellectual Property and Science business of Thomson Reuters). WoS includes literature databases that cover the sciences, social sciences and arts & humanities. It is one of the largest and most widely used databases for peer reviewed scientific literature (Baykoucheva, 2015b; Pranckutė, 2021).

WoS is based on the Science Citation Index, a multidisciplinary index covering the journal literature of the sciences, including agriculture, astronomy, biochemistry, biology, chemistry, computer science, materials science, mathematics, medicine, neuroscience, physics, plant sciences, and zoology. It is the best database for finding which papers have cited other papers. Cited references can be traced forward in time. It provides access to the international journals, open access resources, books, patents, proceedings, and web sites.

The Science Citation Index (SCI) was created by the Institute for Scientific Information (ISI) in the 1960ies (Baykoucheva, 2019). It is a selective subset of journals, which are high impact titles in many scientific disciplines. Other products included in Web of Science that are based on the Science Citation Index are Journal Citation Reports and Essential Science Indicators. Chapter 7 and Chapter 8 present the history of the SCI and WoS in greater detail.

The Web of Science Core Collection includes the Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Emerging Sources Citation Index, Book Citation Index, and Conference Proceedings Citation Index. Web of Science is also a platform that includes other databases such as Essential Science Indicators, and Journal Citation Reports. Web of Science covers all important chemistry journals. The chemistry databases *Current Chemical Reactions* and *Index Chemicus* are part of the Web of Science Core Collection. A complete list of the resources on the Web of Science Core Collection can be found elsewhere (*Web of Science Core Collection*, 2021). Figure 5.7 shows the consecutive steps of searching for literature in Web of Science and refining results.

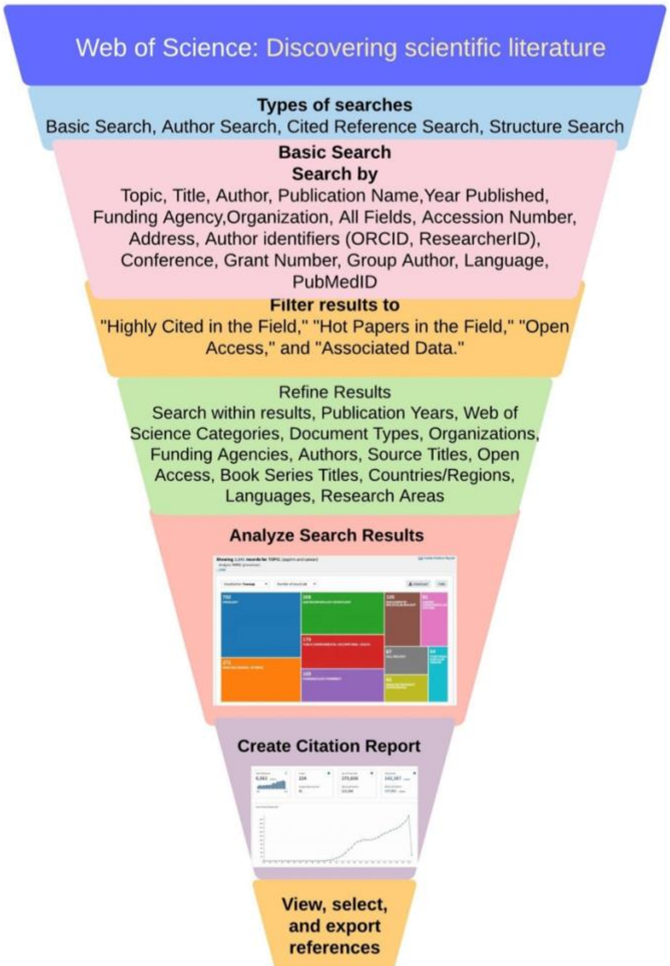

Figure 5.7 Performing a literature search in Web of Science, refining search results, and analyzing the output.

Wiley Open Access Library is a multidisciplinary collection of online resources covering life, health, and physical sciences; social sciences; and the humanities.

zbMATH Open is an open access database (Formerly, Zentralblatt MATH), published by FIZ Karlsruhe – Leibniz Institute for Information Infrastructure GmbH, Berlin, Germany. Contains reviews or abstracts from over 3,000 journals and serials, 180,000 books from all areas of pure mathematics, as well as applications in natural sciences, computer science, economics and engineering.

Finding chemical literature and properties of chemical compounds

> *There is one aspect of chemistry that makes the information-seeking activities of chemists quite unique. That is the use of the chemical structure as a universally recognized search key. Much of the organization of chemical information is predicated on the similarities of properties among chemical substances with similar structures. It was only natural that ways would be devised to search the primary chemical literature and secondary databases by structure*.
>
> Gary Wiggins (Baykoucheva, 2015c)

Chemists can search for properties of chemical compounds by drawing molecular structures and reactions. Interactive tools such as structure editors (or "drawing editors") allow users to draw chemical structures and reactions in a database and search for properties of chemical compounds. Search results can be visualized, downloaded, and deposited.

Books and articles discuss different aspects of discovering chemical information (Baykoucheva, 2015b; Baykoucheva et al., 2016; Buntrock, 2016; Currano & Roth, 2014; Engel et al., 2018; Krallinger et al., 2017; Milman & Zhurkovich, 2020). An issue of the ACS *Journal of Chemical Education* was specifically devoted to this topic (Baysinger, 2016). The ACS Guide to Scholarly Communication is a valuable resource for questions related to chemical information (Banik et al., 2021) .

The most important databases for finding properties of chemical compounds are ChemSpider, CRC Handbook of Chemistry and Physics, PubChem, Reaxys, SciFinder, and The Merck Index. ChemSpider (from the Royal Society of Chemistry) and PubChem (from NCBI) are free resources, but the other resources require paid subscriptions. Chemistry is closely related to other disciplines such as physics, the life sciences, and engineering, and researchers in these fields benefit from using such integrated platforms as SciFinder, rather than individual specialized databases. The literature in chemistry is extensively covered not only in SciFinder, but also by PubMed, Scopus, and Web of Science. The Web of Science Core Collection platform includes the chemistry databases *Current Chemical Reactions* and *Index Chemicus*. A list of the most important chemistry resources is shown below:

ChemSpider is a free database for properties of chemical compounds. It is published by the Royal Society of Chemistry (RSC).

CRC Handbook of Chemistry and Physics **(**published by the Taylor and Francis Group) provides Information about the chemical, physical and thermodynamic properties of chemical compounds (Baykoucheva, 2009).

Current Chemical Reactions covers new synthetic methods published in leading journals and patents. Records show the overall reaction flow for each method, accompanied with a detailed graphical representation of each reaction step. It can be accessed from the Clarivate Analytics' platform Web of Science Core Collection.

Index Chemicus contains structures and data of new organic compounds reported in leading international journals. It is a very important source of new information on biologically active compounds and natural products. Index Chemicus can be accessed from the Clarivate Analytics' platform Web of Science Core Collection.

The Merck Index is an online reference work for information on chemicals, drugs and biological materials (Baykoucheva, 2010b). It is published by the Royal Society of Chemistry (RSC).

PubChem is a free database for properties of chemical compounds and biological activities of small molecules. It includes three linked databases: PubChem Substance, PubChem Compound, and PubChem BioAssay, which provide substance information, compound structures, and bioactivity data, respectively. It is produced by the National Center for Biotechnology Information (NCBI) at the NIH's National Library of Medicine .

Reaxys is a web-based platform that includes two important chemistry databases—the Beilstein Handbook of Organic Chemistry and Gmelin's Handbook of Inorganic and Organometallic Chemistry. The platform also carries the US Patent and Trademark Office database and the World and European Patents (Alexander et al., 2014; Buntrock, 2013). The Beilstein database was originally based on the Beilstein Handbook of Organic Chemistry, created by the Russian chemist Friedrich Konrad Beilstein in 1881 (Luckenbach, 1981). It is the largest database in the field of organic chemistry, covering the scientific literature from 1771 to 1979.

Reaxys is an extensive repository of experimentally validated data, including structures, reactions and physical properties of chemical compounds. It also contains pharmacological, eco-toxicological and toxicological data, specific bioassay results, and toxicity information. It allows searching for substances by chemical name or by structure and reactions. Reaxys is produced by Reed Elsevier Properties SA.

SciFinder is a resource for finding chemical, life sciences, and biomedical information. It covers many document types such as journal articles, conference papers, books, dissertations, and patents. SciFinder is a platform that carries two very large literature databases—the Chemical Abstracts Database (CAPLUS) and MEDLINE. The two databases can be searched simultaneously (default option) or individually (Baykoucheva, 2011; Baykoucheva et al., 2016).

SciFinder also provides access to the CAS Registry File, the largest property information database. For every new chemical compound reported in the literature, CAS creates a unique number (called CAS Registry Number). Searching for properties of a chemical compound using the chemical name is sometimes difficult, as the names of chemical compounds can be presented in many different ways. The CAS Number of a chemical compound is its unique identifier—its "Social Security Number." SciFinder is produced by the Chemical Abstracts Service, a division of the American Chemical Society (ACS).

Searches in SciFinder can be performed through three main pathways: topic, structures, and reactions. Figure 5.8 shows the steps for performing topical searches to find literature in chemistry, life sciences and the biomedical field. When searching on a topic, SciFinder allows using keywords, phrases, and natural language.

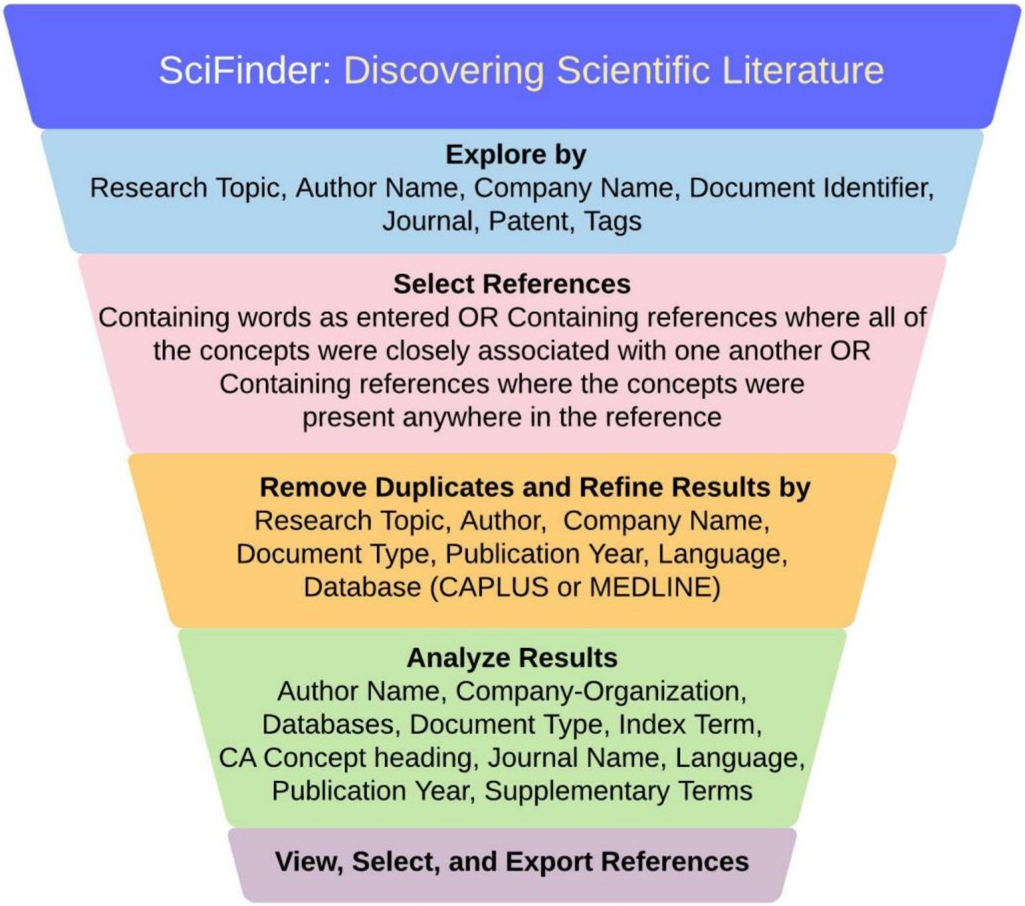

Figure 5.8 Performing a literature search in SciFinder and refining search results.

Searching for properties of chemical compounds can be performed using a chemical name, a CAS Registry Number, or drawing a molecular structure or a reaction. Besides the CAS Registry Number, there are other unique identifiers of chemical compounds. The International Chemical Identifier (InChI) is a textual identifier for chemical substances to encode molecular information and to facilitate the search for such information in databases and on the Web (Heller et al., 2013). It was initially developed by IUPAC (the International Union of Pure and Applied Chemistry) and NIST (the National Institute of Standards and Technology) (http://nist.gov). SMILES (Simplified Molecular Input Line Entry System) is another chemical nomenclature and data-exchange format that is used to retrieve properties of chemical compounds.

Managing search results

Discovering scientific information is only one of the challenges for researchers. After performing a search, users need to export all or selected references to bibliographic management programs, which allow storing, organizing, and citing these references. Some of the bibliographic management programs allow users to search directly library catalogs, and indexed databases from within the program. Others, such as Mendeley, support collaboration (Murphree et al., 2018). The most common bibliographic management programs are shown in Figure 5.9.

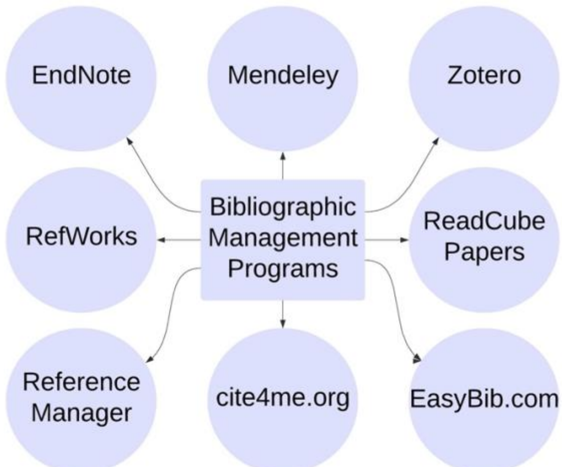

Figure 5.9 Bibliographic management programs

Citation management tools are discussed in more detail elsewhere (Murphree et al., 2018). Selecting a bibliographic management tools depends on the needs of users and their patience to deal with technical problems. It is often a matter of personal preference which program to use, but EndNote is considered the most robust one for researchers. It is a very complex system, though, that may betray its users. The most used bibliographic management programs are shown below:

EndNote

EndNote is a bibliographic management program from Clarivate Analytics, which is available as EndNote Online and EndNote Desktop versions (Baykoucheva et al., 2016). EndNote Online is free with institutional license, with limited storage. The software version (EndNote Desktop) is for a fee, with no limitation for storage. EndNote offers the highest number of citation styles and provides excellent live technical support.

Mendeley

Mendeley is a free desktop and web-based bibliographic management program from Elsevier for managing and sharing research papers, discovering research data and collaborating online. It is free with limited web storage and can be upgraded for a fee. It has simple interface; retrieves metadata for retrieved PDFs. Mendeley performs also as a social media and collaborative tool.

Zotero

Zotero is a free open-source program, which is an easy-to-use tool for collecting, organizing, citing, and sharing references. It has a browser add-on for Firefox, Chrome, and Safari. It is free for limited web storage, but it is upgradable for a fee. Zotero has excellent user guides. It can manage a variety of formats and has good functionality. Downloading, saving, and inserting citations is easy with Zotero.

Preprint servers

Preprint servers are online archives, or repositories, containing works or data associated with scholarly papers before they are peer reviewed or accepted by traditional academic journals. Many researchers are now posting their papers on preprint servers before submitting them to a scientific journal. Preprints offer a way to freely disseminate research findings while a manuscript undergoes peer review. This early dissemination of a paper allows authors to obtain feedback about their paper before its official publication. This also promotes it, increasing the possibility to be cited (Fu & Hughey, 2019). Authors can submit revised versions of their papers to the preprint server at any time.

Since the early 1990s, there has been a movement to make papers available online by posting them on preprint servers, to facilitate the dissemination of scientific results. The first preprint server was ArXiv, which was used by researchers in physics. ArXiv later started to include preprints from researchers in computer science, mathematics and other science disciplines. It now accepts e-prints in the fields of physics, mathematics, computer science, quantitative biology, finance, statistics, electrical engineering, and economics. Biologist are using bioRxiv (Fu & Hughey, 2019), while chemists post their papers mainly on ChemRxiv (Pagliaro, 2021). A list of selected preprint servers is shown below:

- ArXiv
- Authorea
- BeilsteinArchives
- ChemRxiv
- Cognet
- OSF Preprints
- PeerJ Preprints
- RepEc
- Preprints
- SSRN
- ResearchSquare
- ResearchGate
- TechRxiv
- Zenodo

Finding research data

Data are generated through experiments, statistical procedures, or other ways and include experimental data, computer simulation, and geospatial data. In the social sciences, data are usually numeric data, but they could also be in other formats such as be video, audio, and other digital content. Sharing research data and depositing them in publicly available curated databases increases their discoverability and allows scientists from all over the world to use this information in research. Sharing and making data widely accessible is not only expected within many disciplines, but also required by federal granting agencies and many scientific journals (Culley, 2017). A roadmap for implementing data citation practices provides guidance for achieving human and machine accessibility of deposited data (Cousijn et al., 2018). The suggested practices, if implemented, will significantly improve verification, validation, reproducibility and re-use of scholarly/scientific data.

Reproducibility and reusability of research results has become a great concern in scientific communication and science policy, especially in view of the dramatic increase of retractions of papers from even reputable journals (Baffy et al., 2020). Retraction Watch, a website dedicated to the preservation of scientific integrity, monitors retractions of papers and helps draw attention to how researchers, journals, and institutions should correct the scientific record (*Retraction Watch*, 2021). Retraction Watch has been building a comprehensive and freely available database of retractions. In an interview, the founders of this service talked about the important role played by Retraction Watch in promoting transparency and integrity in science and disseminating best practices in scientific publishing (Markovac et al., 2018).

Being able to compare datasets or interpret a particular dataset requires knowing a broad array of information about the data. Having more "raw" data for research linked to a published article is helpful in many ways. For example, it would help readers who want to look at data from a different perspective than the author did or who want to replicate an experiment. As part of the publishing process, automated methods that help in detecting error, fraud, or plagiarism will be essential for the integrity of data. For authors, data offer an opportunity to be recognized and cited in new ways.

Nature has published a list of recommended dataset repositories (Nature, 2021). Additional information about storing, managing, citing and finding data is available elsewhere (Baykoucheva, 2015a; Cousijn et al., 2018). Some of the resources for datasets are presented below (Baykoucheva, 2015a).

DataCite is a data registry, which allows finding a repository to deposit and/or find data. Assigns DOIs for datasets. Incorporates Databib (a catalog, registry, directory, and bibliography of research data repositories).

Data Citation Index has a single search interface for data repositories worldwide.

Data.gov **i**s U.S. Government's site for open data, which includes tools and resources to conduct research, develop web and mobile applications, and design data visualizations.

Digital Curation Centre (DCC) is a leading international center for data management located in the United Kingdom.

DataONE (Data Observation Network) provides access to data across multiple member repositories, supporting discovery of Earth and environmental data.

Dryad Digital Repository is a curated resource that provides a home for a wide diversity of datatypes and makes scientific data discoverable, reusable, and citable.

figshare is a repository for many different kinds of files (e.g., figures, datasets, media, papers, posters, and presentations), which can be uploaded and visualized in a browser.

Google Dataset Search enables users to perform a simple keyword search to find datasets stored in thousands of repositories across the Web. It accompanies the search engine Google Scholar.

Harvard Dataverse Network is a repository for long-term preservation of research data, which provides permanent identifiers for datasets. It is open to all researchers worldwide to publish research data across all disciplines.

Mendeley Data allows storing, sharing, publishing, and finding research data, Includes millions of datasets from domain-specific and cross-domain repositories.

National Technical Reports Library is a data publishing and discovery platform of the National Data & Surveying Services (NDS) Consortium hosted at the University of Illinois.

NIH Data Sharing Repositories is the National Institute of Health-supported data repositories that make data accessible for reuse.

OpenAIRE is the Open Access Infrastructure for Research in Europe. It is a large-scale shared archive network of aggregated digital repositories for datasets and other kinds of scientific outputs from many disciplines.

OpenDOAR is a global directory of Open Access repositories and their policies.

RDA (Research Data Alliance) is an International organization, which connects researchers and enables open sharing of data across technologies, disciplines, and countries.

re3data.org is a global registry of research data repositories that covers research data repositories from different academic disciplines. It presents repositories for the permanent storage and access of datasets to researchers, funding bodies, publishers and scholarly institutions; funded by the German Research Foundation (DFG).

Zenodo is an open-access repository developed from the European OpenAIRE program and operated by CERN. It allows researchers to deposit research papers, datasets, research software, reports, and any other research related digital objects. A persistent digital object identifier (DOI) is assigned to each new submission, which allows citing the stored items.

There are significant differences between researchers in their practices of sharing data (Baykoucheva, 2015a). They depend on the discipline, the culture of the lab, and whether there is a potential for patenting the results. An example of a culture of research where sharing is a common practice is the area of bioinformatics, an area of research that uses interdisciplinary techniques and processes large volumes of diverse data into more manageable and useable information. The handling and analysis of data in this discipline are performed through annotation of macromolecular sequences and by classifying sequences or structures into groups. Researchers can interact with these databases and manipulate them, which is of critical importance and significantly benefits users.

Finding patents and trademarks

Patents are an important document type for the sciences. A list of databases providing access to patents is presented below:

- DEPATISnet
- Derwent Innovations Index
- Espacenet (European Patent Office - worldwide patents)
- Freepatentsonline.com
- Global Brand Database(World Intellectual Property Organization (WIPO))
- Global Design Database (World Intellectual Property Organization (WIPO))
- Google Patents
- Nexis Uni Academic Patent Search
- The Lens
- Patentscope (World Intellectual Property Organization (WIPO))
- PIUG
- Reaxys
- ScienceDirect
- SciFinder
- Science Citation Index Expanded (Web of Science)
- USPTO (United States Patent and Trademark office)

Social media as an information environment

Social media have become an important resource for discovering scientific information (Boudry & Durand-Barthez, 2020; Ebrahimzadeh et al., 2020; Manca, 2018; Nicholas et al., 2017; Nicholas et al., 2015). Academic sites such as Academia.edu and ResearchGate provide access to full text articles. Some publishers have taken actions toward preventing ResearchGate from providing access to full text articles that are protected by copyright (Jamali, 2017; Van Noorden, 2017).

Figure 5.10 summarizes the features of Academia.edu, ORCID, ResercherID, and ResearchGate.

	ORCID	ResearcherID	ResearchGate	Academia.edu
Accessibility	Non-profit	For-profit company	For-profit company	For-profit company
	Access without restriction	To get an identification number, a minimum of one publication in Web of Science is mandatory	Access without restriction	Access without restriction
	Free of charge	Free of charge	Free of charge	Free of charge with paid version offering more features
Institutional use and career links	In practice, mandatory in most institutional approaches	Highly recommended in most institutional approaches besides ORCID	No	No
			Recommendations, job opportunities	Grants, job opportunities
Interoperability with author identifiers	Link/exchange of data with ResearcherID and Scopus ID	Link/exchange of data with ORCID	ResearcherID and ORCID optionally added by the researcher in the CV but are not search keys	ResearcherID and ORCID optionally added by the researcher in the CV but are not search keys
Interconnection /researcher	No	No	Internal mailing	Internal mailing
			Followers/Following	Followers/Following
Researcher CV	Yes	Yes	Yes	Yes
Publications	Diverse sources admitted (Crossref, ResearcherID, Scopus. . .) including personal. No full text upload / share by user.	ResearcherID links researcher's publications across all Web of Science Group products. Addition of external items admitted, including personal. No full text upload/share by user.	Full text upload / share by user, if applicable. Copyright infringement recorded in the past	Full text upload/share by user, if applicable. Copyright infringement recorded in the past
Citations	No	Web of Science Core Collection Citations	Yes	Yes in paid version
Analytics	No views and/or download stats	No view and/or download stats	Views (publications), proprietary metrics (e.g. RG Score, Research Interest)	Views (publications and profile) and downloads (full texts) up to 6 months (more on paid version)

https://doi.org/10.1371/journal.pone.0238583.t001

Figure 5.10 Comparison of the characteristics of Academia.edu, ORCID, ResearcherID, and ResearchGate (Boudry & Durand-Barthez, 2020).

Academia.edu and ResearchGate serve as sources of bibliometric and Altmetrics indicators such as publication counts, reads, number of downloads, citations, and profile views. They automatically alert users about the addition of new publications considered to be of interest to particular researchers (Boudry & Durand-Barthez, 2020). Social media significantly increase the visibility of scholarly publications. These sites have only a very basic operation, such as performing a search on an author's name.

ResearchGate is mostly oriented towards science, technology, engineering, and medicine. The social sciences, arts & humanities are also present, but to a much lesser extent. While ResearchGate offers complete services at no cost, Academia.edu charges for some additional services (for example, finding out who has cited or read your work). ResearchGate offers specific metrics, such as the RG Score, the relevance of which is still not very clear and has been criticized. It is based on the volume of a researcher's' publications as well as the extent of participation with other researchers.

ResearchGate has its own citation index by extracting citations from documents uploaded to the site and reporting citation counts on article profile pages. Since authors upload preprints to ResearchGate, it may use these preprints to show new papers. An article assessed whether the number of citations found for a set of recent articles was comparable to other citation indexes. The results show that ResearchGate found fewer citations than did Google Scholar, but it retrieved more citations than both Scopus and Web of Science. The citations retrieved by ResearchGate correlated with Google Scholar citations (from the same journals), indicating that they both are tapping a similar source of data. Preprint sharing in ResearchGate is significant, which makes it a serious resource for finding recent articles (Thelwall & Kousha, 2017).

Academia.edu is used by scholars in the social sciences and the arts & humanities. It offers all the standard services provided by academic social networks, but it also offers a paid version that provides additional services such as direct and complete access to citations. The benefit of Academia.edu is that members could add alternative spellings and forms of their names.

ORCID is a unique author identifier used by researchers to manually create their author profiles, which include affiliation, employment, funding, publications, and peer reviews. Contrary to other social sites such as Academia.edu and ResearchGate, ORCID is interoperable with numerous organizations, allowing automatic updating of information from other sites. As the creation of profiles is not overseen or controlled, researchers can create multiple profiles, leading to duplication. ORCID does not provide metrics and does not allow researchers to upload full texts of their publications.

ResearcherID can be used to find the works of a particular author. Registered users create their profiles, and their publications are automatically added to their profile from the Web of Science database, if they were published in journals that are indexed in this database. Authors can also add their publications from other databases, from DOIs, or by uploading files using BibTex or RIS formats. It is not possible to upload full texts on ResearcherID. Authors can add information about their affiliations, add keywords to describe their research field, add reviews, view their citations (based on publications included in Web of Science), and view different metrics. ResearcherID, now on Publons.com, is provided by Clarivate Analytics through its Web of Science Core Collection. ResearcherID and ORCID automatically exchange publication records. Searching ORCID or ResearcherID using author's name will retrieve the publications of this author. These unique identifiers can also be used as search keys in some of the databases.

Conclusion

The science information discovery field is going through a major transition, as an entirely new generation of indexing and analytical services is emerging. Based on Open Science and using modern business models, newcomers such as Dimensions , eLife (*eLife*, 2021), and The Lens, are reshaping the discovery process, offering innovative solutions for a smoother discovery experience. Services such as CrossRef, which are publicly available, have started challenging the existing discovery services that have dominated the research environment for decades. Integrating bibliographic management programs in the discovery process will allow a smoother retrieval and processing of scientific information.

The current revolution in Internet, mobile, and machine-to-machine (M2M) technologies are expected to bring together diverse technologies and enable more efficient discovery of information. With access to so much information and technology, it is researchers need to pay attention to new information channels for discovering scientific information.

References

Baffy, G., Burns, M. M., Hoffmann, B., Ramani, S., Sabharwal, S., Borus, J. F., Pories, S., Quan, S. F., & Ingelfinger, J. R. (2020). Scientific Authors in a Changing World of Scholarly Communication: What Does the Future Hold? *Am J Med*, *133*(1), 26-31. https://doi.org/https://doi.org/10.1016/j.amjmed.2019.07.028

Banik, G., Baysinger, G., Kamat, P. V., & Pienta, N. J. E. (2021). *The ACS Guide to Scholarly Communication (ACS Publications)*. American Chemical Society. Retrieved March 22 from https://pubs-acs-org.proxy-um.researchport.umd.edu/doi/book/10.1021/acsguide

Baykoucheva, S. (2009). ATG Interviews David R. Lide, Editor-in-Chief, "The CRC Handbook of Chemistry and Physics": A Mountain, a Cathedral, a Battleship, or ... an Iceberg? [Interview]. *Against the Grain*, *21*, 50-56.

Baykoucheva, S. (2010a). Selecting a Database for Drug Literature Retrieval: A Comparison of MEDLINE, Scopus, and Web of Science. *Science & Technology Libraries*, *29*(4), 276-288 http://drum.lib.umd.edu/handle/1903/11411 DOI: 11410.11080/0194262X.0192010.0522946. https://doi.org/10.1080/0194262X.2010.522946

Baykoucheva, S. (2010b). The Merck Index, an Encyclopedia of Chemicals and Natural Products Interview with Maryadele O'Neil. *Chemical Information Bulletin*, *62*(3), 5-9. http://acscinf.org/content/merck-index-encyclopedia-chemicals-and-natural-products.

Baykoucheva, S. (2011). Comparison of the contributions of CAPLUS and MEDLINE to the performance of SciFinder in retrieving the drug literature [Article]. *Issues in Science & Technology Librarianship*, *66*(Summer), http://www.istl.org/11-summer/refereed11.html DOI: 10.5062/F5042Z5013FT.

Baykoucheva, S. (2015a). Coping with "Big Data": eScience. In S. Baykoucheva (Ed.), *Managing Scientific Information and Research Data* (pp. 71-84). Chandos Publishing. https://doi.org/https://doi.org/10.1016/B978-0-08-100195-0.00008-1

Baykoucheva, S. (2015b). Finding and managing scientific information. In S. Baykoucheva (Ed.), *Managing Scientific Information and Research Data* (pp. 33-41). Chandos Publishing. https://doi.org/https://doi.org/10.1016/B978-0-08-100195-0.00005-6

Baykoucheva, S. (2015c). The complexity of chemical information: interview with Gary Wiggins. In *Managing Scientific Information and Research Data* (pp. 97-102). Chandos Publishing. https://doi.org/https://doi.org/10.1016/B978-0-08-100195-0.00010-X

Baykoucheva, S. (2019). Eugene Garfield's ideas and legacy and their impact on the culture of research [Review]. *Publications*, *7*(2), 43. https://doi.org/10.3390/publications7020043

Baykoucheva, S., Houck, J. D., & White, N. (2016). Integration of EndNote Online in information literacy instruction designed for small and large chemistry courses. *Journal of Chemical Education*, *93*(3), 470-476 http://drum.lib.umd.edu/handle/1903/19141. https://doi.org/10.1021/acs.jchemed.5b00515

Baysinger, G. (2016). Introducing the Journal of Chemical Education's "Special Issue: Chemical Information" [10.1021/acs.jchemed.6b00113]. *J. Chem. Educ.*, *93*(3), 401-405. https://doi.org/10.1021/acs.jchemed.6b00113

Boudry, C., & Durand-Barthez, M. (2020). Use of author identifier services (ORCID, ResearcherID) and academic social networks (Academia.edu, ResearchGate) by the researchers of the University of Caen Normandy (France): A case study. *PLoS One*, *15*(9), e0238583. https://doi.org/10.1371/journal.pone.0238583

Buntrock, R. E. (2016). Using citation indexes, citation searching, and bibliometrics to improve chemistry scholarship, research, and administration [10.1021/acs.jchemed.5b00451]. *Journal of Chemical Education*, *93*(3), 560-566. https://doi.org/10.1021/acs.jchemed.5b00451

Cousijn, H., Kenall, A., Ganley, E., Harrison, M., Kernohan, D., Lemberger, T., Murphy, F., Polischuk, P., Taylor, S., Martone, M., & Clark, T. (2018). A data citation roadmap for scientific publishers. *Scientific Data*, *5*(1), 180259. https://doi.org/10.1038/sdata.2018.259

Crossref. (2021). Crossref. Retrieved April 14 from https://www.crossref.org/

Currano, J. N., & Roth, D. (2014). *Chemical information for chemists : a primer*. Royal Society of Chemistry.

DataPlanet. (2020). *Data Planet*. Sage Publishing. Retrieved december 9, 2020 from https://www.data-planet.com/data-planet-statistical-datasets

Dimensions. (2021). Digital Science. Retrieved April 14 from https://link.springer.com/content/pdf/10.1007/s11192-019-03217-6.pdf

Eassom, H. (2021). *Discoverability in China: Why Baidu Scholar is Good News for Researchers*. Wiley. Retrieved March 29 from https://www.wiley.com/network/researchers/promoting-your-article/discoverability-in-china-why-baidu-scholar-is-good-news-for-researchers

Ebrahimzadeh, S., Rezaei Sharifabadi, S., Kamran, M., & Dalkir, K. (2020). Triggers and strategies related to the collaborative information-seeking behaviour of researchers in ResearchGate. *Online Information Review*, *ahead-of-print*. https://doi.org/10.1108/OIR-12-2019-0380

EBSCO. (2021). *EBSCO Research Databases*. Retrieved April 14 from https://www.ebsco.com/products/research-databases

eLife. (2021). Retrieved April 15 from https://elifesciences.org/

Engel, T., Gasteiger, J., John, W., & Sons. (2018). *Chemoinformatics : basic concepts and methods* (First edition. ed.). Wiley-VCH Verlag GmbH & Co. KGaA.

Fu, D. Y., & Hughey, J. J. (2019). Meta-Research: Releasing a preprint is associated with more attention and citations for the peer-reviewed article. *Computational and Systems Biology*, *8*, e52646. https://doi.org/https://doi.org/10.7554/eLife.52646

Huang, P. C. C. (2018). Citation Indexes: Uses and Misuses. *Modern China*, *44*(6), 559-590. https://doi.org/10.1177/0097700418796778

Jamali, H. R. (2017). Copyright compliance and infringement in ResearchGate full-text journal articles. *Scientometrics*, *112*(1), 241-254. https://doi.org/10.1007/s11192-017-2291-4

Krallinger, M., Rabal, O., Lourenco, A., Oyarzabal, J., & Valencia, A. (2017). Information retrieval and text mining technologies for chemistry [Review]. *Chemical Reviews*, *117*(12), 7673-7761. https://doi.org/10.1021/acs.chemrev.6b00851

Leung, B. T. H., Xie, J., Geng, L., & Pun, P. N. I. (2019). Search Engines. In B. T. H. Leung, J. Xie, L. Geng, & P. N. I. Pun (Eds.), *Transferring Information Literacy Practices* (pp. 163-174). Springer Singapore. https://doi.org/10.1007/978-981-13-7743-3_7

Luckenbach, R. (1981). The Beilstein Handbook of Organic Chemistry: the first hundred years. *Journal of Chemical Information and Computer Sciences*, *21*(2), 82-83. https://doi.org/10.1021/ci00030a006

Manca, S. (2018). ResearchGate and Academia.edu as networked socio-technical systems for scholarly communication: a literature review. *Research in Learning Technology*, *26*(0). https://doi.org/10.25304/rlt.v26.2008

Markovac, J., Kleinman, M., & Englesbe, M. (2018). Publishing Ethics: An Interview With the Founders of Retraction Watch. In J. Markovac, M. Kleinman, & M. Englesbe (Eds.), *Medical and Scientific Publishing* (pp. 179-186). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-809969-8.00019-X

Martin-Martin, A., Thelwall, M., Orduna-Malea, E., & Lopez-Cozar, E. D. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations [Article]. *Scientometrics*, *126*(1), 871-906. https://doi.org/10.1007/s11192-020-03690-4

Milman, B. L., & Zhurkovich, I. K. (2020). Big Data in Modern Chemical Analysis. *Journal of Analytical Chemistry*, *75*(4), 443-452. https://doi.org/10.1134/s1061934820020124

Murphree, L., White, M., & Rochen Renner, B. (2018). Reference Managers that Support Collaborative Research: Dreaming of the Perfect Fit. *Med Ref Serv Q*, *37*(3), 219-233. https://doi.org/10.1080/02763869.2018.1477706

Nature. (2021). *Recommended Data Repositories | Scientific Data*. Macmillan Publishers. Retrieved March 13 from https://www.nature.com/sdata/policies/repositories

Nicholas, D., Boukacem-Zeghmouri, C., Rodríguez-Bravo, B., Xu, J., Watkinson, A., Abrizah, A., Herman, E., & Świgoń, M. (2017). Where and how early career researchers find scholarly information. *Learned Publishing*, *30*(1), 19-29. https://doi.org/10.1002/leap.1087

Nicholas, D., Herman, E., Jamali, H., Rodriguez-Bravo, B., Boukacem-Zeghmouri, C., Dobrowolski, T., & Pouchot, S. (2015). New ways of building, showcasing, and measuring scholarly reputation. *Learned Publishing*, *28*(3), 169-183. https://doi.org/10.1087/20150303

Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The Titans of Bibliographic Information in Today's Academic World. *Publications*, *9*(1), 12.

Randall, L. (2005). *Warped passages : unraveling the mysteries of the Universe's hidden dimensions* (1st ed. ed.). Ecco.

Retraction Watch. (2021). Retrieved April 15 from https://retractionwatch.com/

The Lens. (2021). Retrieved 14 Feb 2021 from

Thelwall, M., & Kousha, K. (2017). ResearchGate versus Google Scholar: Which finds more early citations? *Scientometrics*, *112*(2), 1125-1131. https://doi.org/10.1007/s11192-017-2400-4

Van Noorden, R. (2017). Publishers threaten to remove millions of papers from ResearchGate. *Nature*. https://doi.org/doi:10.1038/nature.2017.22793

Web of Science Core Collection. (2021). Clarivate Analytics. Retrieved April 14 from https://images-webofknowledge-com.proxy-um.researchport.umd.edu/WOKRS533JR18/help/WOS/hp_database.html