

THE Information Advisor's

GUIDE TO INTERNET RESEARCH

Review: Paratext's Government Documents Masterfile

» BY ROBERT BERKMAN, CO-EDITOR

All good researchers are aware of the wealth and diversity of data, reports, and documents published by the U.S. federal government. The government has always been one of the best places to turn for countless business research queries and projects and remains an enormous creator and publisher of information.

The problem has always been figuring out just where to go to perform an effective search of what's available from the U.S. government. Because there is *so much* information, and in so many formats published by so many departments and agencies, it has always been a challenge to know where to find what one needs.

To help researchers and the general public ameliorate that problem, over the years, the federal government has created several "one-stop shopping" sites. Ironically, the fact that there has been a need to create more than a single "one-stop shopping" site points to the fact that none are truly comprehensive. Still, the best of these clearinghouses have helped

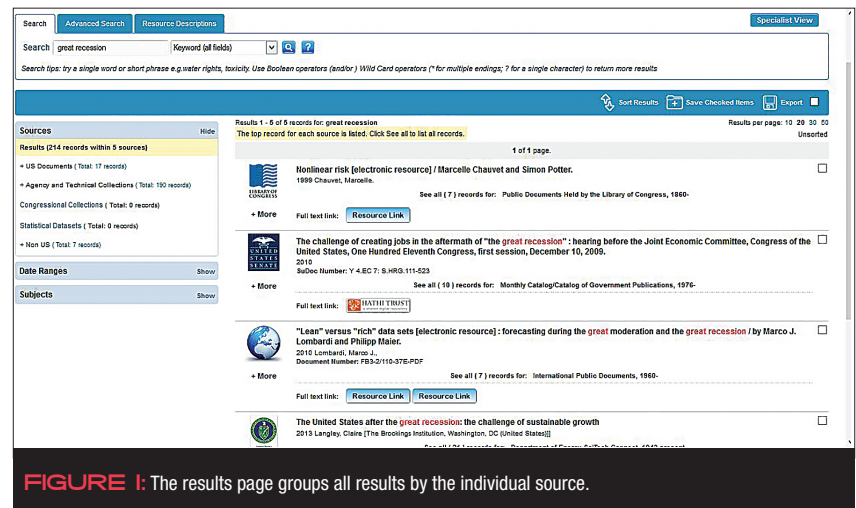


FIGURE 1: The results page groups all results by the individual source.

researchers search across a wide range of diverse types of information published by multiple agencies.

Some of our favorite government-created clearinghouses include Data.gov; FedStats; FDSys and MetaLib from the Government Publishing Office (GPO); and USA.gov. You can read a full review of FDSys in the July/August 2013 issue of *The Information Advisor*; and a roundup

of other clearinghouses of government information appeared in the July/August 2012 issue.

Another option to find what you need from the federal government is to search Public Documents Masterfile, a fee-based database that allows researchers to find and discover more than 12 million items published by the federal government.

CONTINUED ON PAGE ► 2

IN THIS ISSUE

- 1 REVIEW: PARATEXT'S GOVERNMENT DOCUMENTS MASTERFILE
- 1 BBC: BEYOND THE NEWS
- 4 ROUNDUP: NEW AND INTERESTING BUSINESS RESEARCH VENDORS AND PRODUCTS
- 7 THE SCHOLAR'S WEB: MEET ME AT THE PUBAG
- 8 BIG DATA: HEALTHCARE ANALYTICS

BBC: Beyond the News

» BY RAN HOCK

Numerous news resources contain an immense amount of content that is easily missed when we, out of habit, just go for the headlines or other specific sections of the resource. One of the most highly regarded news sources worldwide is BBC News (bbc.com/news), and BBC is an epitome of this situation of easy-to-miss, but often very useful content. This article is going to assume that most readers are

familiar with BBC News (and maybe Weather and Sports) and will address what BBC has to offer beyond those, the information that can be found in other sections and in the nooks and crannies of the site. Above all, BBC offers a tremendous amount of nonheadline-news information that can serve as valuable background for various kinds of research, both professional and personal; this background can be especially

CONTINUED ON PAGE ► 5

About half of those documents, which are retrieved from 26 primary government collections, link to the full text.

This database is a product of Paratext (paratext.com), a Virginia-based library services company founded in 1993. Paratext is known by many academic librarians for its various historical indexes and citation databases. In this issue, we review its Public Documents Masterfile of government information, and evaluate how it performed and where it fits for the researcher who needs to locate that needle of information in the giant federal government information haystack.

WHAT'S INSIDE

First let's take a look at what kinds of information Public Documents Masterfile contains. The database draws from 26 primary government collections, specifically, the following:

- American State Papers, 1789–1838
- Annals of Congress, 1789–1824
- Catalogue of the Public Documents of the 53rd Congress and of All Departments of the Government, 1893–1895
- Checklist of United States Public Documents, 1789–1909
- Compilation of the Messages and Papers of the Presidents, 1789–1897
- Comprehensive Index to the Publications of the United States Government, 1881–1893 (the Ames Index)
- Congressional Globe, 1833–1873
- Cumulative Title Index to United States Public Documents, 1789–June 1976
- Data-Planet Statistical DataSheets
- Department of Energy SciTech Connect, 1943–present
- Descriptive Catalogue of the Government Publications of the United States, September 5, 1774–March 4, 1881 (Poore's Catalogue)
- Educational Resources Information Center (ERIC), 1965–present
- *Farmers' Bulletin*, 1889–1984
- International Public Documents, 1960–present
- Journal of William Maclay, 1789–1791
- Journals of the Continental Congress, 1774–1789
- Monthly Catalog of United States Government Publications (Hickcox), 1885–1894
- Monthly Catalog Subject Indexes, 1895–1976

- Monthly Catalog/Catalog of Government Publications, 1976–present
- Popular Names of U.S. Government Reports, 1821–1981
- Public Documents Held by the Library of Congress, 1860–present
- Public Documents of the First Fourteen Congresses, 1789–1817
- Register of Debates, 1824–1837
- Sketches of Debate in the First Senate of the United States, 1789–1791
- U.S. Congressional Serial Set, 1818–1994 (Readex)

A couple of these sources are worth calling out. The GPO, or Government Publishing Office (which until January 2014 was called the Government Printing Office), is the primary general source for disseminating government information. Its official charge, according to its website, is to be the “primary centralized resource for producing, procuring, cataloging, indexing, authenticating, disseminating, and preserving the official information products of the U.S. Government in digital and tangible forms.” The GPO publishes more than 800,000 titles—that's more than any other official government source—but still does not include *all* of the data and reports created by the federal government.

Another of Paratext's suppliers, Data-Planet, makes more than 150,000 metadata records representing authoritative publicly and privately sourced datasets of statistical information. (We covered Data-Planet's own online service in detail in our May 2015 issue, reviewing and evaluating its database of statistical data series.) Those records are drawn from 425 databases provided by 78 different publishers, from organizations such as the British Bankers' Association, International Monetary Fund, National Bureau of Economic Research, National Center for Education Statistics, Standard & Poor's, United Nations Economic Commission for Europe, U.S. Census Bureau, the World Bank, and others. Note that in order to access the full text of Data-Planet's statistical records, Paratext users must also be subscribers to Data-Planet. If not, the searcher will only be able to link to a citation for the retrieved records.

Finally, the International Public Documents database is a file that Paratext tells us it is phasing out, as its goal is to only provide access to U.S. federal data.

According to Aaron Koepper, Paratext's media and data coordinator, the database

is updated on a quarterly basis, with an additional lag time of 2 to 3 weeks for processing.

SEARCHING THE DATABASE

The way users obtain access is by supplying Paratext with their server's public IP address. Upon going to the link (documents.paratext.com), the user is authenticated as a subscriber.

At the homepage, the user can immediately run a keyword search across the entire collection, either by using the standard basic search box or by clicking the advanced search link to get a few more search options. Note that keywords are not searched against the entire full text of all documents accessible by Paratext, but by the associated MARC citation data, which includes elements such as a Library of Congress title and subject heading, language, and other fields, depending on the source and document. Note as well that the MARC subject heading in each record is hotlinked, offering the searcher the option to click on the subject directly and rerun a search on that specific term.

When running a basic search, the user can enter keywords, wildcard characters (a question mark for a single placeholder and an asterisk for multiple places), apply quotation marks for phrases, and use Boolean operators AND, OR, and NOT. Entering multiple words defaults the search to an AND operator. We found that using quotation marks incorrectly retrieved records when the words appeared anywhere in the citation, not just as a phrase.

Clicking on the Advanced Search link also provides the user with the ability to limit the search by date range, author, title, SuDoc (Superintendent of Documents) number, and specific source. (The user may choose one source only, not multiple sources.)

When displaying results, the database organizes results so that the left-hand panel identifies how many items were retrieved, and each one is grouped under a few major source type categories. (See Figure 1 on page 1.) The center panel displays a short citation of one item from each source with the user's keywords highlighted in red. The sample items are selected by their relevancy (which includes the number of occurrences of the search terms in the indexed fields), as well as those which include a full-text link. The items are ranked by relevancy by default, though the user can adjust the listing to be ranked by date, author, or other criteria.

At this point, the searcher can examine the documents retrieved from one of the

providing sources by reviewing one of the five umbrella categories on the left-hand panel (U.S. documents; agency and technical collections; congressional collections; statistical datasets; and non-U.S. documents). Clicking on the plus sign next to an umbrella category displays a list of the specific government agencies, providers, etc., grouped within that category, and then clicking on that particular provider retrieves just those documents retrieved from that source. Another option on the left-hand panel is to filter results by subject and date range as well. It is not possible to refine results by combining *both* a subject and date filter after the search is run. This can only be done beforehand when entering keywords into an advanced search.

Note that it is not possible to review *all* records retrieved, whether by relevancy or other sorting option, regardless of specific source supplier. A specific source provider must be chosen first.

Also displayed at the bottom of the left-hand panel is a cloud tag image that identifies words and phrases associated most frequently from the documents retrieved from the search. (See Figure 2.) The user can choose to click on any of those to re-run a search on the selected term.

When the searcher is ready to obtain more information from a specific record, or find additional information about a displayed item, he or she can simply click the “more” link associated with the record, which will then display additional information about that item, and then view the full MARC record. (See Figure 3.) Finally, when available, the user can click another link to view the full text. (See Figure 4.)

Other features available in the database include the ability to email a link, export results, and see results in a “Specialist View.” The Specialist View displays results in the older format that longtime users had been accustomed to viewing.

PRICING

According to Koepper, pricing for academic libraries is determined by Carnegie Category, which specifies whether an institution is categorized as doctoral, master’s, baccalaureate, etc. For public and special libraries, pricing is determined by the size of the population served. No other pricing information was made available.

EVALUATION

We liked Paratext’s Public Documents Masterfile, and believe it provides another

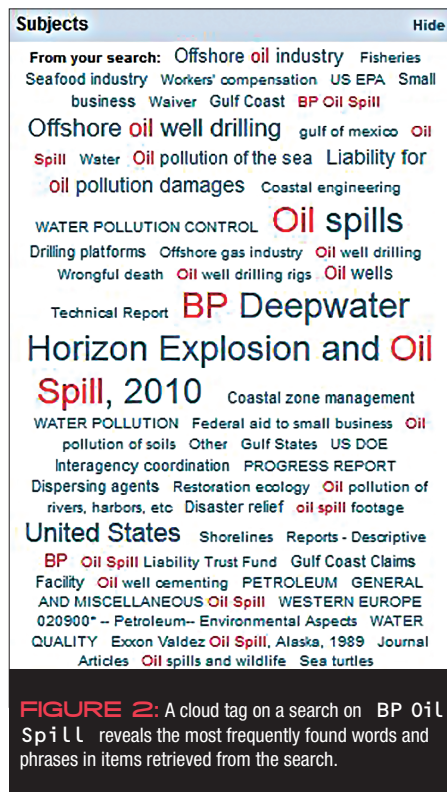


FIGURE 2: A cloud tag on a search on BP Oil Spill reveals the most frequently found words and phrases in items retrieved from the search.

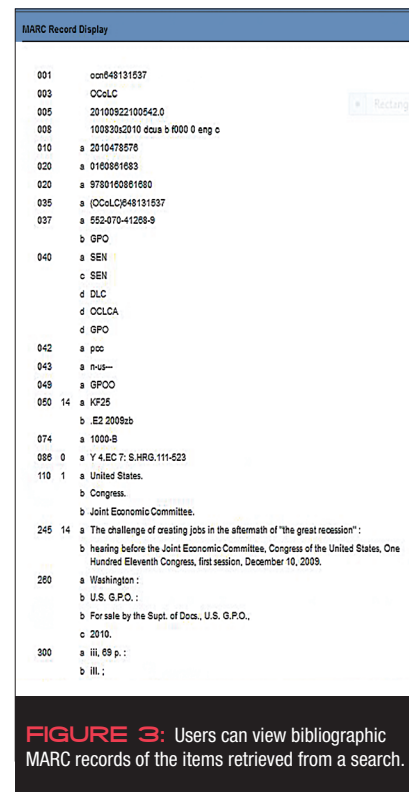


FIGURE 3: Users can view bibliographic MARC records of the items retrieved from a search.

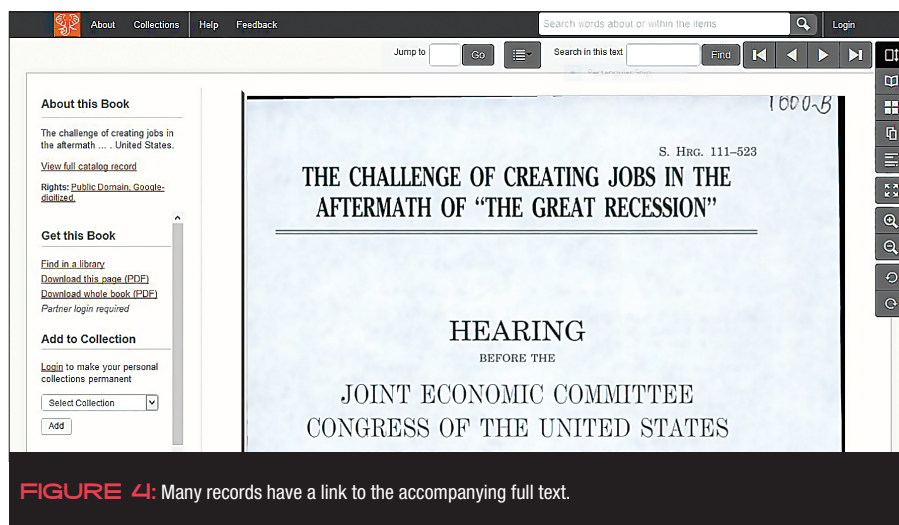


FIGURE 4: Many records have a link to the accompanying full text.

very good option for searching huge amounts of government information, supplemented by other public sources such as Data-Planet and Readex. We also like the fact that the database integrates such a credible and exact indexing method, which can make searching more precise compared to searching only on the full text.

However, we would like to see even more advanced search options to give the researchers more control over what he or she turns up. For example, we would like to see the option to build a longer sophisticated Boolean

search string; the ability to include more than one source as well as the ability to exclude one or more sources; and the option to perform post-search filtering by more than one element at a time. And while it can be convenient to view the results grouped by source, there should also be an option to view all records, ranked by relevance or date, regardless of the provider.

Finally, because we were unable to obtain more specific pricing information from the company, it is difficult to assess its overall value.

Roundup: New and Interesting Business Research Vendors and Products

» BY ROBERT BERKMAN, CO-EDITOR

One of the best ways to discover the latest research sites, software, and products is finding out what is being exhibited at key library and information conferences. Events that we find to be particularly useful for learning about new offerings include the Special Libraries Association (SLA) annual conference as well as Computers in Libraries, Internet Librarian, ACRL, Data Summit, and the SIIA.

There are, of course a lot of familiar “usual suspects” exhibiting at these events, but we decided to take a look at who was exhibiting within the past year (including at the June 2015 SLA conference) to identify just those vendors and products that 1) we were unfamiliar with and 2) looked to be of potential interest to business researchers.

We have listed our favorites here. For now we have excerpted each vendor's own description of its products; however, we plan to go through this initial listing and select several for in-depth reviews during the course of the year. If you spot any you would personally like us to review, please feel free to email one of the editors to let us know.

ACI INFORMATION GROUP

aci.info

Founded in 2004 as Newstex, ACI Information Group is the world's leading aggregator of editorially selected and curated social media and blog publications, with more than 10,000 scholarly blogs and more than 500,000 scholarly blog posts in its indices.

ALEXANDER STREET PRESS

alexanderstreet.com

By partnering with thousands of the world's leading content providers, Alexander Street Press provides high-quality resources in disciplines including history, literature, music, health sciences, business, and engineering. It is also the only database with the full text of CBS' *60 Minutes*, a video resource titled *60 Minutes: 1997–2014*.

ALLIS INFORMATION MANAGEMENT HOMEPAGE

allisinfo.com

This business intelligence and research firm gathers and analyzes business information for competitive and crisis management teams.

ARTICLE GALAXY WIDGET–REPRINTS DESK, INC.

info.reprintsdesk.com/widget101

Article Galaxy Widget simplifies lowest-cost access to full-text scholarly papers via PubMed, Google Scholar, and more than 50 popular search and discovery tools.

AUDIT ANALYTICS: AUDIT REGULATORY AND DISCLOSURE INTELLIGENCE

auditanalytics.com

Audit Analytics provides detailed research on more than 150,000 active audits and more than 10,000 accounting firms.

BMI RESEARCH

bmiresearch.com

BMI Research links macroeconomic, industry, and financial market analysis to provide insight across 24 industries and 200 global markets via forecasts, data, and analyses for multinationals, governments, and financial institutions to guide critical strategic, tactical, and investment decisions.

CANBACK & COMPANY

canback.com

This is the world's leading management consulting firm leveraging predictive analytics. It also offers analytic services with the Canback Global Income Distribution Database (C-GIDD) as a flagship product.

CHINA DATA CENTER

chinadatacenter.org

The China Data Center at the University of Michigan is an international center designed to advance the study and understanding of China. A primary goal is the integration of historical, social, and natural science data in a geographic information system, where spatial and temporal references are maintained through a relational database.

DIGIMIND: SOCIAL MEDIA MONITORING & ANALYTICS SOFTWARE

digimind.com

Digimind provides software to monitor unlimited social media and web sources and sort mentions by their reach.

DISCOVERY SEARCH ENGINE–TRANSPARENSEE SYSTEMS

transpareensee.com

Transpareensee provides an integrated enterprise search platform with fuzzy matching, structured data navigation and visualization, and free text search. The technology provides near and exact search matches, a recommendation engine, and widgets and sliders to enhance the user experience. An in-memory index specializes in understanding the relationships among objects in a dataset across hundreds of metadata points.

EAST VIEW INFORMATION SERVICES

eastview.com

Whether from Russia, China, Egypt, or elsewhere, East View identifies, sources, and distributes materials in the sciences, humanities, and social sciences in regions where information is difficult to find.

ECOAMÉRICAS

ecoamericas.com

EcoAméricas provides a monthly report on development and the environment in Latin America for an international audience of businesses and public agencies. EcoAméricas was the winner of the 2013 SIPAwards for Best Newsletter–Public Sector/Government.

ECOSYSTEM HUB PLATFORM–GUST

gust.com

Gust is a central site for a regional entrepreneurial ecosystem that brings together every startup, angel investor, venture capital fund, event, job, course, workspace, blog, video, news story, and business startup resource in the region.

EXCHANGE DATA INTERNATIONAL LTD

africadata.com

African Financial and Economic Data (AFED) from Exchange Data International provides definitive economic intelligence on all 56 African economies and their financial market information.

FDI INTELLIGENCE

fdiintelligence.com

fDi Intelligence is a specialist division from The FT Ltd. providing industry insight

and analysis on cross-border expansion and greenfield inward investment.

GLOBAL FINANCIAL DATA

globalfinancialdata.com

Global Financial Data contains financial and economic data on more than 60,000 companies covering 300 years and appends current data to the historical data; it offers complete, unabridged data series.

INFOMINEO: BUSINESS RESEARCH SERVICES

infomineo.com

Infomineo provides business research services in Africa, the Middle East, Europe, and all over the world. Services include market research, outsourcing, KPO, and BPO.

INFONGEN: DISCOVERY ENGINE FOR BUSINESS

infongen.com

InfoNgen is a discovery engine for business professionals.

OPINIONARCHIVES

opinionarchives.com

OpinionArchives is the sole provider of the complete full-text electronic archives of *The New Yorker*, *Harper's*, *Weekly Standard*, *Washington Monthly*, *The Nation*, *National Review*, *The New Republic*, *Dissent*, *Orion*, *Commentary*, *American Spectator*, *NACLA*, *The Progressive*, *Commonweal*, *NYRB*, *Moment*, and *New Leader*.

POLICY AND REGULATORY REPORT (PARR)

parr-global.com

PaRR delivers global intelligence, analysis, and proprietary data on competition law, IP and trade law, and sector-specific regulatory change. It provides subscribers with inside knowledge, analysis, and comment, focusing on seven areas that directly impact development of competition laws around the world: Merger Review; Cartel Investigations; Intellectual Property; Abuse of Dominance; International Trade & Dispute; State Aid & Public Programs; Other Legislations.

PREQIN

preqin.com

Preqin is a leading source of information for the alternative assets industry, providing data and analysis via online databases, publications, and bespoke data requests. It covers private equity, real estate, infrastructure, and hedge funds.

QUADRANT: GLOBAL ECONOMIC AND FINANCIAL DATA

quadrant.io

Quadrant finds and charts economic and financial data.

SIMPLYMAP

geographicresearch.com/simplymap

SimplyMap provides information mapping and data visualization software for examining and analyzing complex data sets. It combines a robust catalog of more than 75,000 variables with a streamlined interface that helps researchers, analysts, and strategists make better-informed decisions.

SINEQUA

sinequa.com

Sinequa offers real-time search and deep analytics to extract significant value from all kinds of enterprise data: structured and unstructured, internal and external.

SMARTLOGIC

Smartlogic.com

SmartLogic's Semaphore is a content intelligence platform. It captures the context of a business and analyzes content to identify and present its value, enabling better access to corporate knowledge, improved risk management and compliance, superior customer service and relationship management, enhanced information findability for key audiences, and an improved ability to monetize information.

SPORTS BUSINESS RESEARCH NETWORK: SBRNET

sbrnet.com

SBRnet serves any individual or organization with an interest in the business of sports with "one-stop" access to multiple resources.

USPTO PATENT AND TRADEMARK RESOURCE CENTERS (PTRCS)

uspto.gov/learning-and-resources/support-centers/patent-and-trademark-resource-centers-ptrcs

PTRC library staff are information experts trained on how to use search tools to access patent and trademark information.

VISIBLE (CISION)

cision.com/us/social-software

Visible focuses on three core capabilities: social media monitoring, enriched data analytics and insights, and customer engagement. All three are integrated into one platform, which is directly accessible through a browser.

BBC • CONTINUED FROM PAGE ► 1

useful to researchers who find themselves researching areas outside their usual areas of expertise.

One way to begin to understand the extensiveness of the BBC site is to ask Shakespeare's question, "What's in a name?" Though the phrase "British Broadcasting Corporation" is not visible on the BBC's homepage (but is found in the page's metadata), "British Broadcasting" quite accurately describes what BBC does and what the site is about. The founding father of BBC, John Reith, had, according to the website, a vision "to educate, inform and entertain the whole nation." In the BBC's earliest days—it began broadcasting in 1922—the broadcasting consisted of plays, concerts, talks, and variety programs, with news only occurring in the evening. The BBC website today even more emphatically educates, informs, and entertains with the primary intended audience still being British, but with a geographic "sun never sets" reach and utility.

EXPLORING THE BBC WEBSITE

Finding what you want on BBC is not particularly difficult. In addition to browsing by means of menus, there is a site search box provided next to the main menu. However, as with many other information sources, until one understands the kinds of information present, one may not even think of going to that resource to do a search there—one might not know what to want or expect from a resource.

Start your own exploration with the main menu found on the very first line of BBC pages. Skip the News, Weather, and Sports options and go to Earth, Future (maybe stop by Shop, Radio, and TV), and especially the More option. Under More, you will find links to 17 sections (though on different pages you may find fewer or more), plus a link to the "Full A-Z" index. (These categories appear at the top of BBC pages, as well as at the bottom of most pages.) The rest of this article focuses on a few of the 17 sections, some in detail and some more briefly, to point out specific, unique, perhaps unexpected, and hopefully interesting content.

As you explore the BBC site, you encounter pages with this message: "This page has been archived and is no longer being updated." Though this may sound worrisome, it's actually a good thing. Its main point is that users should be aware that the content of that page may be out-of-date and/or may contain something no longer accurate because of its age. The message is also reflective of a project at

CONTINUED ON PAGE ► 6

BBC in which millions of pages are being moved to a new location but will be securely saved and accessible for many years.

SECTIONS OF THE BBC WEBSITE

As you browse through these sections, note that a very large portion of the website's content relates to TV and radio programs currently or recently produced by the rather numerous BBC radio and television channels. Also note that most of the TV programs from BBC are only available by means of the BBC iPlayer App, which is only available in the U.K., and those programs are only downloadable in the U.K. There is an iPad/iPhone app that, with a subscription, makes BBC TV content available in 17 countries, but not the U.S. This restriction is not connected to the events of 1776 or 1812 or the tea tax or the Stamp Act. The situation is related to the U.K. "television license," which is required of British citizens who wish to watch TV at home and which provides the major funding for the BBC. Nevertheless, the good news is that access to BBC Radio programs is not restricted. Take advantage of the availability of BBC Radio, which has a very broad-ranging programming content; in particular, check out BBC World Service (bbc.co.uk/worldserviceradio) if you think there might be any political partisanship in U.S. news stations. BBC is widely regarded for its more "international" perspective.

Two of the 17 sections, Earth and Future, are on the main menu. Earth is very much an online "nature" magazine, with photos, stories, and videos on earth, sea, and sky creatures and astronomy, often addressing scientific topics, but written at a "popular" level. Content is arranged under categories labeled A Colourful Life, Strange & Beautiful, Amazing Moments, and Bizarre Cosmic Objects. Future will take you to articles largely about where we are headed, particularly in the areas of technology, science, and health.

One of the richest and most extensive sections on the BBC site is Learning. On the main Learning page, you see subsections for Adult Learners, Schools, Parents, and Teacher Packs. On that page you also find, for navigation purposes, a Browse by subject section. Go to the Adult Learners page to find links to a variety of online courses, arranged in categories for Skills, General Interest, and Performing, Sports, and Media. Skills courses include those for using the web, literacy skills, languages, gardening, cooking, history, ethics, singing, sports, and more. The Languages section is particularly notable, offering online

courses for seven U.K. languages and 40 other languages. The BBC language courses provide videos, transcripts, guides, worksheets, and other tools at an introductory level as well as teacher guides and other resources.

The Schools section of Learning offers learning and teaching resources at both the primary and secondary levels, while the Parents section offers guidance and aids for parents, including information on child development, safety, special education needs, etc. The Teacher Packs section of Learning includes age-specific lesson plans and other curriculum resources. On the main Learning page, explore the subject guide and also the link to BBC's new iWonder section. The iWonder section is actually a BBC education brand currently located under BBC's "old" URL at bbc.co.uk/iwonder. iWonder "provides thought-provoking answers to the questions sparked in your everyday life." Articles include "Sun-nis and Shias: What's the Story?" and "Can Classical Music Make Us Smarter?" Typical of the convenient interconnectedness of BBC content, you will also find iWonder articles popping up in several other sections.

The Capital section of the website is about money, primarily personal financial matters and related topics. Main subcategories are Columns, Personal Finance, and Careers. The first subcategory, Columns, includes 22 regular columns, including 90 Second Financial Fix ("Life's financial questions and issues ..."), Business Traveler ("Everything you need to know about the most-travelled work cities"), CEO Guru Shorts ("The world's chief executives on leadership and management"), and Expat Guide ("Exploring the best and worst aspects of cities popular with expatriates"). The Personal Finance subsection of Capital includes articles such as "How to Go Back to School at 40" and "Why Does Shopping Feel So Good?" Some pieces, such as a video on getting the best mobile phone plans, have a British slant, but most are more universal. The Personal Finance and Careers subcategories are very similar, containing many of the same stories, but chosen with a different emphasis. For easy navigation, beneath each story are topic links (e.g., networking, psychology, managing, corporate responsibility, etc.), each leading to a page of stories relating to that topic.


BBC's Travel section can perhaps be best thought of as an online travel "magazine" and/or travel guide, with articles and videos on destinations, interests, and other travel topics. The Destinations link on Travel's menu leads to pages on more than 80 countries; each page has multiple articles. The In-

terests link on the menu takes you to a page for 34 travel "interests," including camping, casinos, cruises, diving, ecotourism, festivals, golf, national parks, photography, and snow and skiing. Each of these pages typically contains more than a dozen BBC articles on that area of interest. The Video menu choice gives access to episodes from 10 series of short videos, most from 3 to 5 minutes in length.

The main page of BBC's History section contains a variety of articles and video, much from BBC TV and consequently not available to the U.S. However, scroll down the page to find links to subsections for British History, Ancient History, Family History, History for Kids, Historic Figures, History on BBC Radio, and History Highlights. The last is material available through iPlayer and so not available in the U.S. Content for the other sections is available without such restriction. Those other sections contain, in well-organized subsections, hundreds of history-related articles, plus video, audio, timelines, and guides.

Your Paintings (bbc.co.uk/arts/yourpaintings), done in partnership with the U.K.'s Public Catalogue Foundation, is a section that can remind you there are times when you will definitely want to use the A-Z index or the search box for finding a specific topic. This robust feature is not listed in the categories under More or at the bottoms of pages, nor is it linked to from the Culture page. Your Paintings aims to show the entire U.K. national collection of oil paintings, the stories behind the paintings, and "where to see them for real." It contains photos of more than 2,000 paintings and thousands of artists (not just British). For 67 artists, there is a slideshow, a biography, and additional links.

Using BBC's A-Z Index (found by a link on most pages) is a great way to more fully understand and utilize both the breadth and the specificity of BBC content. The link provides a lot of content, particularly special features (such as Your Paintings) and content related to local, national, and international news, that will not easily be noticed even when browsing by section. Pick any letter ("T" works nicely) and scan the list of topics found there. The A-Z Index page carries the alphabetic index as well as boxes for browsing the sections just discussed.

No matter how you get there, you will find something new or of interest on the BBC website. 

RANDOLPH (RAN) HOCK, Ph.D., specializes in customized seminars for effectively using the internet and is the author of *The Extreme Searcher* book series. He can be reached at ran@onstrat.com.



Meet Me at the PubAg

» BY NICK TOMAIUOLO

Earlier this year, Ran Hock wrote about AgWeb (agweb.com), a portal for agricultural news and reports. As Ran said, it's crammed with real-time information for anyone connected with the farming industries. The article spun me off on another angle: specifically, which databases support the scholarly agricultural literature? As it happens, a new Department of Agriculture resource has cropped up, and it's called PubAg (pubag.nal.usda.gov). Calling to mind the kerfuffle surrounding the Department of Energy's dumping PubScience (former OSTI director Walter Warnick deemed it "superfluous"), let's remember what the feds give, the feds can take away. But let's analyze PubAg with hopeful skepticism.

Our search for cheap alternatives to subscription databases seems perpetual, yet we're wary of free databases. We all love a gush of content, but if we are personally going to use a free database or offer it to others, some selection ground rules must exist. The guidelines from Los Rios Community College District's Collection Development Policy (crl.losrios.edu/services/library/about/collection/electronic#Criteria) include ease of use, favorable reviews, and relevance to curriculum. I'll add that the site should be free of advertising, sponsored by a recognized entity, and include an editorial policy.

PubAg is a specialized resource that may seem far out of scope for non-STEM organizations and places that prepare graduates for careers in business, criminal justice, or education. My experience with PubAg confirmed that it's useful for multiple subject areas and offers some interesting features.

COVERAGE AND RELEVANCE

According to Emily Marsh, digital librarian at the National Agricultural Library (NAL), PubAg's earliest record goes back to 1996, but she adds that's only true "at the moment," because NAL is "working to add records at a rate of 20,000 per month on a continual basis." PubAg includes 450,000 citations, and the USDA's own staff has authored and stocked the database with more than 40,000 peer-reviewed articles (in traditional journals published by Elsevier and Wiley, etc.). In fact, the majority of the full text you'll find within the resource is authored by USDA staff, and out

of a 12-month embargo period (others originate from open-access journals). It covers the agricultural sciences including nutrition, food safety, food quality, sustainable agricultural systems, and animal and crop production and protection, but also deals with natural resources, rural development, economic and policy issues, imports and exports, and agricultural statistics.

My searches found information on various topics relevant to undergraduate/graduate students (and faculty), including sustainable tourism, highway construction, DNA sequencing, smartphones, management information systems, anatomy and physiology, global warming, and beer.

SEARCHING AND RETRIEVAL

Having met the selection tests of editorial policy, sponsorship, and relevance, PubAg needed to be tested for ease of use. But because search interfaces and post-retrieval options have become so bromidic, I was prepared to skip describing these aspects of PubAg entirely. But what initially appears to be a "same old" one-line search box on the homepage is the tip of an iceberg where the approach to knowledge discovery is both interesting and idiosyncratic.

Words entered in the homepage's basic search box are immediately matched (if possible) with terms in the NAL Thesaurus (NALT). Plus PubAg will search it as a text word and every other "equivalent" NALT term. For example, I enter **dogs**. Because "dogs" is a NALT term, PubAg searches for "dogs" as a thesaurus term, then ORs that with equivalent terms in NALT, including "canine," "canis familiaris," and "domestic dogs." This is confirmed by information shown (in Boolean language) under "Search Details" on the retrieval screen. Beneath the "Search Details" box you will see "Refine my search." Clicking that link will bring you to a screen where you may select/deselect more broad and/or narrow terms such as "therapy dogs," "dog diseases," "mammals," and "canidae," and add them to (or subtract them from) your search. And if you're uncertain as to whether you should use any of these terms, you may always consult the NALT thesaurus (agclass.nal.usda.gov/dne/search.shtml). There, you'll see the scope for

each term as well as its relationship to other subject headings.

For example, entering **working dogs** into the homepage basic search box retrieves 82 items; entering "**working dogs**" (note quotation marks) retrieves 33. Click Advanced Search, and you may enter **working dogs** and use a drop-down menu to search a specific field (the choices are Author, Title, Abstract, Journal, All Fields, Full Text, Year). In the Advanced Search multiple-word input, "working dogs" is already searched as a phrase—and you'll never be obligated to deal with the NALT.

Not all dot-gov databases are created equal. Truncation is not available in PubAg, but is assumed in NAL's Agricola. PubMed allows saved searches and email updates. Science.gov permits several methods for delivering alerts (including RSS). PubAg doesn't offer these features. PubAg has "Print" (sends brief citation to browser) and "Export" (to an Excel spreadsheet); that's all you'll get when it comes to manipulating retrieval output.

PubAg lists retrieval in the form of a brief citation. Clicking the title brings up the abstract, hotlinked NALT terms, hotlinked DOI, a link to the free full text, or (beware, paywall ahead) a link to the publisher's site. If you are accessing the database from an institution that has subscriptions to articles not residing within PubAg as full text, you should be able to view/print/download the documents without charge.

HARVEST FESTIVAL

Here's how to get the most bang for *no* bucks. Start with AgWeb for the news on your topic. Hop over to PubAg to find scholarly journal articles. Then complement your search by checking the University of Minnesota's open access repository called AgEcon Search. AgEcon is somewhat different in that it encourages researcher participation and invites uploads of content from conferences or journals if sponsored by an organization. Implicit is the fact that while individual resources may fulfill their own missions, our awareness of sources helps us fulfill ours. Welcome PubAg.



NICK TOMAIUOLO (tomaiulon@ccsu.edu) works at the reference department of Elihu Burritt Library at Central Connecticut State University.



BIG DATA

Healthcare Analytics

» BY JACQUELINE CAROL SHANE

Healthcare has the potential to build a feedback mechanism like no other professional field because the practice creates so much data and relies so heavily on information and smart decisions. The knowledge-base of clinical medicine has the potential to grow each time medical clinicians observe and record data. If enough attributes are captured and fed back into data repositories, it may help to streamline decision skills. The pay-off for medical practice is the opportunity to apply a dynamic wealth of scientific knowledge in order to close gaps in care or locate opportunities for cost-savings. The drawback is that these data warehouses require management and are vulnerable to cybersecurity breaches that may potentially hinder the privacy of individuals. Healthcare analytics uses statistical and qualitative analysis software in order to make predictive healthcare models.

HEALTHCARE INFORMATION AND MANAGEMENT SYSTEMS SOCIETY (HIMSS)

himss.org

Though physically based in Chicago, HIMSS represents 52,000 people, most of whom work in healthcare, government, or nonprofits around the globe. HIMSS is a nonprofit society whose purpose is to optimize healthcare outcomes through information technology. It provides health IT education, market research, and events, such as the HIMSS Annual Conference and Exhibition. HIMSS also offers several virtual seminars each month, such as "Preventing a Healthcare Data Breach," which focuses on HIPAA compliance and information flow.

HIMSS has a Resource Library filled with original reports and articles. I recommend reading "Health IT Body of Knowledge" (himss.org/library/health-it-body-of-knowledge), a good introduction for health IT beginners. This Health IT 101 document primarily covers concepts and definitions, such as patient-centered systems and workflow. The section on workflow has a brief description plus a link to five other articles for discussion. Find it via the Resource Library at the top and scroll down to the Featured Collections. For those interested in creating a career in healthcare IT, HIMSS has a "jobmine,"

virtual seminars, podcasts, and online publications, all of which can be found in the Resource Library.

HIMSS publishes the *Journal of Healthcare Information Management* (himss.org/jhim). Although there are about 20 core journals in healthcare administration, *JHIM* is to date the only online, all-digital, peer-reviewed journal focusing on healthcare IT.

OPEN CLINICAL

openclinical.org/home.html

Open Clinical is a British website targeted toward healthcare professionals and managers, as well as computer scientists and people who work in the field of medical informatics. Its aim is to be a "one-stop shop" for tracking developments in advanced management technologies for healthcare, such as guidelines for clinical workflow and point-of-care decision support. Like HIMSS, Open Clinical is a nonprofit organization, receiving support from Cancer Research UK. It provides background, research, clinical expertise, and links to commercial suppliers. Its Public link provides some international best practice examples of e-Health initiatives, such as PQL, or Pure Quality Life, used as a stress management tool in Sweden, and an online "diabetes diary" in France. Users are encouraged to register, which enables updates on topics of interest and entitlement to publish one's own work. It promises an anti-spam policy and privacy security.

While Open Clinical, along with HIMSS, does offer background material on the topic of Healthcare IT, most resources appear to be targeted toward advanced practitioners. Its research links to guideline modeling, past and current projects, ethics and safety, and downloads for shareware. The biggest difference between Open Clinical is its emphasis on decision support and artificial intelligence, as opposed to IT and analysis.

HEALTH CATALYST

healthcatalyst.com

Headquartered in Salt Lake City, Health Catalyst is a commercial company that offers healthcare IT solutions for the healthcare industry. Health Catalyst specializes in developing healthcare data warehouse solutions that enable faster analytics than mere databases. It claims its Catalyst technology enables faster and more capable analysis of healthcare data than traditional methods. The reason I selected this site is for its discussions on the type of data collected and

how providers can use it. Data can consist of patient records and conditions, comorbidities, and "changes in scientific flow." (Comorbidity is when two or more chronic conditions exist simultaneously, yet independently, in one patient.)

The best reason to look at the Health Catalyst page is for its Knowledge Center. Though there is a heavy bias toward using data warehouses rather than databases for healthcare analytics, the site does provide good background on the elements involved. The Knowledge Center contains original articles, success stories, some executive papers and reports, videos, and ebooks. Health Catalyst provides a My Folder mechanism that enables saving and sharing files seamlessly. It reportedly offers data warehouse and platform analytics platforms to 1,900 hospitals and clinics, including such clients as Stanford Hospitals & Clinics.



JACQUELINE SHANE, former reference librarian and associate professor at the University of New Mexico, is an information consultant and writer in Santa Fe, N.M.

THE Information Advisor's
GUIDE TO INTERNET RESEARCH
WWW.INFORMATIONADVISOR.COM

(ISSN: 2333-2336) is published 10 times per year by Information Today, Inc.
143 Old Marlton Pike, Medford, NJ 08055

Editors > Robert Berkman and Barbara Quint
Email > iag@infotoday.com

Contributing Editor > Mary Ellen Bates
Senior Managing Editor > Lauree Padgett
Senior Graphic Designer > Jackie Crawford

The views in this publication are those of the authors and do not necessarily reflect the views of Information Today, Inc. (ITI) or the editors.

No part of this publication may be reproduced or transmitted in any form without permission from the publisher.

SUBSCRIPTIONS

Domestic: 1 year (10 issues) \$199.95
2 years (20 issues) \$377
3 years (30 issues) \$578

The Information Advisor's Guide to Internet Research is available on a site-license basis!

Available for use on your company intranet:
\$595 per year for 6–20 users;
\$895 for up to 100 users;
more than 100 users, please inquire.

SUBSCRIPTION INQUIRIES

Email > iasubs@infotoday.com

Phone > (609) 654-6266

Fax > (609) 654-4309