Knowledge of the Law in the Big Data Age
G. Peruginelli and S. Faro (Eds.)

doi:10.3233/FAIA190014

Tools for Discovery: Opening Doors to Legal Research

Ginevra PERUGINELLI and Sebastiano FARO
Istituto di Teoria e Tecniche dell'Informazione Giuridica – ITTIG-CNR (Italy)

Abstract. Discovery tools are specialized portals for bibliographic research widely used in libraries with heterogeneous collections of electronic and digital resources. The Chapter provides an overview of the library resource discovery environment, explaining how these technologies, methodologies, and products might be able to adapt to changes in the evolving information landscape in scholarly communications. This Chapter also attempts to explore which are the effects of discovery tools on legal research.

Keywords. library resource discovery services, digital library, bibliographic legal research

1. Introduction

In recent years, the library world has witnessed the emergence of new bibliographic research tools, including Discovery Tools (DTs). These tools provide a single window approach to the resources subscribed by the library. This includes the Online Public Access Catalogue (OPAC), e-resources subscribed by the library, institutional repositories, open access content and many more. Similar to Google, or any other general search engine, DTs are built using a pre-harvested central index of data. Internet search engines rely on open access and public domain data to populate their central index, which can sometimes lead to broken links, inaccessible sources and dubious quality of content. By contrast, DTs use data supplied by libraries and publishers, resulting in more reliable search results with stable, direct links to licensed, full text articles and digital content [1]. Therefore, they are created to respond to users' needs of managing the new emerging class of electronic resources and of providing patrons with simpler, and web-based research services [2]; [3]; [4].

From an evolutionary point of view, they represent the effect of a long digital revolution that, starting from the 1980s, has affected the library world. Over the decades, OPACs have gone through several generational phases; it is with the birth of the Web that they acquire visibility outside the physical space of the library and become effective information retrieval tools. The Web has not only played a determinant role in the technological arena, leading a relevant number of users to prefer the use of search engines as main tools for finding information resources, but also permeating our social life [5]. The user is no longer satisfied with a traditional use of information materials, rather he/she needs to be able to move and interact with library catalogs with the same autonomy and independency used on the Web.

Gallacher [6] has highlighted a deep cultural conflict between traditionalists and the so called Google generation[1], explaining the tension between the advocates of traditional bibliographic research and those who fully rely on online research. Furthermore, the work of Gunter, Rowlands and Nicholas [7] analyzes in detail the question of a break between the pre and post Google generations. To characterize the Google generation, these authors make specific hypotheses which fully describe its bibliographic research approach: (1) prefers visual information to textual information; (2) wants a variety of learning experience; (3) has definitely adopted the digital modes of communication; (4) is multitasking; (5) is impatient and does not tolerate any delay; (6) considers its peers more credible than other points of authority; (7) needs to feel permanently connected to the Web; (8) learns more through action and knowledge; (9) prefers information in small quantities, easy to digest, rather than reading the full text; (10) has a poor understanding and lack of respect for intellectual property; (11) is not interested in format or container issues; (12) tends to put virtual reality at the same level of experience.

In the law environment, this clash is very noticeable also with respect to teaching legal research to today's law students and young lawyers. The traditionalist view of legal research is essentially based on the belief that law research based on traditional bibliographic sources such as books, commentaries, print journals is superior to online research, at least as a first step in research. On the other hand, it has been observed that more and more current legal generation is so online-oriented that it has learned to rely on powerful search engines to provide answers to complex questions. Young jurists refer more and more to search engines to conduct complex and comprehensive search pointing to different resources (not just bibliographic) with one stop search. The easy-to-use appeal of the Web for legal research is attracting more and more followers fascinated by the idea of finding the solution to a practical legal case. Furthermore, most students entering universities are more comfortable working on a keyboard and reading from a computer screen rather than using paper in their hand.

In this context and in the other sciences as well, DTs carry out their task reconciling different needs and becoming the undisputed leader in the field of bibliographic research [8]. On the one hand, DTs are closely related to the evolution of libraries and to the quality services they offer; on the other, they are meeting habits of the Google generation.

This Chapter provides an overview of the actual resource discovery environment with a particular focus on the advantages and disadvantages of these tools in the library world. Authors conclude with some reflections on possible developments and effects of these tools on legal bibliographic research.

2. Origin of Discovery Tools

In recent years two innovative types of bibliographic research tools have been developed: the Next Generation Catalog (NGC) and the DT. They represent an evolution with respect

[1]The 'Google generation' is an expression referring to a generation of people, born after 1993, growing up in a world dominated by the Internet. The expression has entered popular usage as a shorthand way referring to a generation whose first port of call for knowledge is the Internet and a search engine, Google being the most popular. This is used in contrast to earlier generations who gained their knowledge through books and conventional libraries.

to OPACs, which since the beginning of the catalog automation era have been the main bibliographic research tools for information retrieval of library collections.

OPACs have been evolved over time: from simple information retrieval tools allowing only a few types of research, to third-generation OPACs, enhanced in functionality and usability, expanded in indexing, data records and collection coverage and extended through links and networks, acting as a gateway to additional collections.

After the mid-2000s, a strong dissatisfaction with OPACs began to appear. In that period the Web evolved at technological and functional levels. The user becomes accustomed to new search engines, also utilized as information retrieval tools for bibliographic research.

A first response was the NGC, which has been defined as an application that still uses data in traditional form, allowing users to perform a simplified search based on friendly interfaces similar to the Web while integrating external resources. The use of these tools spread between 2005 and 2007. The aim was to achieve greater integration within the Web, trying to provide ease of use, better ways to communicate, search and use information. The new features and services introduced by NGC concerned data, search, information retrieval and displaying of results.

Starting from 2007 the NGC begun to lose their primacy in favor of DTs for bibliographic research. In 2010, Marshall Breeding announced that NGCs would become unsuitable and obsolete instruments and, for this reason, they would have been overthrown by DTs [9]. At their origin, there is an increase in digital publishing and the consequent need to manage the huge amount of electronic resources that lead libraries to adopt a tool capable of adequately representing the new type of electronic resources such as e-books, e-journals and databases, while offering a simultaneous access to different types of bibliographic resources.

3. Structure, Components and Functionalities

The structure of DTs is organized into three parts [8]: (i) index, (ii) search interface, and (iii) link resolver.

The *index* includes metadata and full text resources resulting from agreements with commercial publishers, to which are added metadata and open access resources also contained in research repositories. The central index is the main element of competition among the producers of this kind of tools [10]. The amount of indexed metadata depends on existing agreements and licenses between libraries and digital content providers. It is worth mentioning that access rights to metadata and full texts are governed by two types of agreements, independent of each other: those stipulated by the institution and those stipulated by the DT producer adopted by the library.

The *search interface* is similar to search engines' user interface. Navigation is allowed without the need for special instructions, starting from a unique Google like string, or with the possibility of articulating multiple search keys in an advanced search modality; the results are then presented in a short or complete form. Ranking of results by relevance is a common feature of the various DTs and is combined with other sorting criteria, which, however, vary according to the product chosen by the library. The algorithms that allow sorting by relevance of search results are owned by the DTs' vendor and are not always made public [11].

Link resolvers are software that use OpenURL standard[2] and a knowledge base to connect a searcher from a citation to the item held in the library or to online full text content. If the content has been subscribed by the library, the link resolver will directly connect the searcher to the subscription content site. Link resolver software and their associated knowledge bases are essential technologies for modern academic libraries. The success of link resolvers is dependent on (i) complete, consistent, accurate citation metadata, (ii) well-defined knowledge base holdings, and (iii) accurate link syntax as generated by the software. Selecting link resolver and knowledge base software has become more complicated than ever, due to the increasing need for system interoperability. Libraries' expectations for clean metadata following professional and industry standards have correspondingly increased [13].

As a matter of fact DTs are index-based systems: the content of databases, local and remote, is re-indexed by these tools. During the indexing process, the system should treat all content equally. It is not clear how content is indexed and, therefore, recovered in the research phase: there are no standards regulating this process [8] and problems possibly arising from reindexing derive from the different quality and completeness level of metadata that these discovery systems receive from different sources, where metadata can already vary in terms of quality and quantity [14]; [15].

On the basis of this structure, the core features of these tools are represented by the content and technology used. Content, as knowledge base, includes: journals in any format, books, databases, aggregators' content, open source materials, newspapers, local indexes (library catalogue, institutional repositories). Technology includes, among others: harvester (OAI/PMH etc.), automated transfer routines, metadata mapping, indexing technology, de-duplication algorithms, link resolvers, relevancy algorithms, interface technologies.

These components allow DTs to provide users with a great array of functionalities [16] which are briefly described in Table 1. Of course the functionalities showed are included in the majority of most popular library discovery services software available on the market today. A list of selected proprietary and open source software is shown in Table 2.

These are only some of the functionalities that libraries look for. Other important components include ease of implementation, price, vendor support and estimated content coverage.

It is to be noted that the availability of open source solution affects the price charged by commercial vendors. At the same time, the implementation cost of open sources software must be taken into consideration as it requires a lot of expertise on the part of libraries which want to implement it or decide to depend on external service providers for implementing it.

4. Pros and Cons of Discovery Tools

Ease of use seems to be one of the main advantages of DTs and appears as the answer to Breeding's call for a seamless experience presenting a consistent interface, despite the

[2]As stated in [12] OpenURL framework provides "a standardized format for transporting bibliographic metadata about objects between information services". NISO standardized the OpenURL protocol in 2004 as ANSI/NISO Z39.88 and many vendors developed and released their own link resolver software.

Table 1. Advanced features library looks for in DTs

Functionality	Description
One stop search	All library resources in one search
Modern design web interface	A design similar to general search engines
Enriched content	Book, images, reviews, user driven input, such as comments, descriptions, ratings, and tag clouds covered
Faceted navigation	Users are allowed to narrow down search results by categories, also called facets (location, publication date, author, format. . .)
Simple keyword search box	A simple keyword search box of a Google like type is offered. An advanced search is also provided
Results ordered by relevance	Relevance algorithms are applied to the list of results
Recommendations/related materials	Suggestions like 'readers who borrowed this item also borrowed the following ... or a link to recommended readings are offered
Integration with social networking sites	Users can share links to library items with their friends on social networks
Persistent links	A stable URL capable of serving as a permanent link to the record is available

Table 2. Selection of proprietary and open source software

Proprietary tools	EBSCO Discovery Service (EDS); Summon (ProQuest); Axiell Arena 3.1. (Axiell); BiblioCore (Biblio Commons); Primo and Primo Central Index (ex Libris); WorldCat Local (OCLC); OvidDiscovery; Inspire™ Discovery (Innovative Interfaces); Endeca (Oracle); Enterprise (Sirsi Dynix)
Open source tools	Blacklight (University of Virginia Library); VuFind (Villanova University); eXtensible Catalog/XC (University of Rochester); LibraryFind; Franklin; SOPAC

use of multiple technology and content products behind the scenes [17]. According to Gross and Sheridan's studies and usability tests [18], students certainly find DTs an easy way to get results, probably easier than the various options they were confronted before.

In addition, users will be searching a much larger data set than previous databases were able to offer. As Vaughn points out, these new services, with hundreds of millions of items, many of them full text, previously housed in dozens or hundreds of individual silos [19], make it easier for users to find content that they would otherwise miss.

Furthermore, as stated by Way [20] in its usage statistics, after the implementation of a DT, the sharp drop in single database usage, associated with high increase in the number of full-text downloads and resolver clicks of links, suggest that such tool has a considerable impact on users' search behavior and on use of library collections.

However, there are also some relevant disadvantages [21] which are very considerable when libraries choose to rely on DTs.

The number of results and varied formats that DTs return to users is overwhelming, especially on simple, non specific searches [22]. The huge number of results coupled with the increasing amount of available object types and formats can make for a confusing jumble of results. One of the fields in which DTs find greater use are university libraries where there is a greater need to offer a unique search point that coordinates access to materials such as electronic resources and databases through a single tool, capable of managing the authentication to resources reserved for the users of an academic institution

[23]. In such a context, most users need assistance to refine their search. This massive number of records is also a concern for specialty librarians. In many cases, users are interested in a specific set of resources, and the use of DTs does not fit their information needs. Moreover, from an empirical research it has emerged that the students "once they had chosen [the web scale discovery tool] ... stayed with it even when an alternative pathway may have proved more fruitful" [18].

Cost issues are also a concern, as DTs, both from a monetary and staff time perspective, require lots of resources. Initial design and setup of the system, testing, and implementation require time and a specialized skill set.

Another disadvantage regards unrealistic user expectations. Several librarians expressed concern about the possible creation in students' mind that these tools set "the expectation that everything is available online in full text". It was also found that some users expressed frustration and disappointment when the tool pointed them to a physical book, located in the stacks at their library [24].

Finally, in order for content to appear in DTs, it must be licensed by both the library and the DT vendor. This leads to situations where only portions of a library's holdings are searchable via DTs. Users may still need to be directed to conduct their search in topic-specific databases. For this reason DTs require a coordination between content producers, resource discovery service providers and libraries.

Therefore, it is up to libraries, in collaboration with their partners, to set up the central index (which resources will be indexed, etc.), the link resolvers to make the best use of the resources available to their users, and integrate the authentication systems in a transparent manner. It follows that the value of the service will largely depend on the coordination work between content producers, software solution providers, and libraries.

5. Possible Developments

To reflect on what is missing and on the opportunities that DTs can take, the work of Marshal Breeding, helps us [8]. We resume, from his contribution, some of the features not fully realized in the current generation of DTs.

DTs well represent relevant material, but omissions in coverage remains. As an example, resources redacted in non-English languages are not covered in an optimal way. In particular, in the field of law, some of these developments would be very useful for improving legal bibliographic research. Articles, books, journals in many areas of law refer to specific national legal orders and have been published in the language of the jurisdiction analyzed by the authors. Multilingual search represents a crucial issue for the future development of DTs so that the content represented in the discovery index becomes more and more heterogeneous by language.

Advanced and precision searching remains a very important functionality for DTs and a consistent management of metadata should be a major step towards enhancement of these components.

Also the exposition of open access materials from a variety of sources is a big challenge. These resources are very numerous and DTs often provide duplicate and unclear answers. In fact, the original metadata of open access resources often do not follow any standard: DTs do not worry about deduplication [10].

Furthermore, DTs should strongly consider legal blogs and social media, able to quickly summarize the development and evolution of legal science. Many journal arti-

cles mention as bibliographic references blogs of scholars and professionals. The community of legal professionals is getting closer and closer to these containers that are quickly spread for their simplicity of implementation and immediacy of communication. Some bloggers with time have become opinion makers and influencers, constituting an alternative source to traditional legal ones, with which they sometimes collaborate, while maintaining the dialogic character with readers.

Relevancy ranking is another key issue to consider for the developers of DTs. However, how relevancy works is ignored by librarians and users.

Enhanced discoverability through non-textual associations is also desirable. Clustering technologies may be able to produce facets based on the content of resources retrieved to guide the user towards the ones that match his/her interests. Progress has been made, but there is still much room for improvement.

Nevertheless, Breeding continues by listing future enhancements that should be made in response to requests from libraries and users. In particular, some of these improvements are related to social features, analytics, altmetrics. Opportunities to enable social interaction would depend on standardized mechanisms that enable interoperability between the ecosystems of discovery services and those of external social networks. Furthermore, libraries and publishers have considerable interest in the ability to measure the performance of their discovery services and which resources have been retrieved. Finally, alternative measures relating to the description of the impact of scholarly resources and the performance of academic libraries are undoubtedly necessary. It can be discussed to what extent they can become part of the discovery ecosystem and whether they can be used in relevance algorithms to help identify materials of higher interest or quality.

6. The Impact of Discovery Tools on Bibliographic Legal Research

DTs have determined a paradigm shift. They give greater importance to discovery than to simple search typical of information retrieval techniques on which previous tools were based [25].

DTs offer a different approach to information, aiming at providing awareness, rather than a specific response, by aggregating content related to a particular area of interest.

In such a way, the path of discovery follows a more explorative approach, often driven by a generic need, or at least less explicitly stated. Actually, DTs are born with the goal of providing the certainty of not missing the most important information on a specific topic.

In the legal domain, this approach is not always the correct one. As a matter of fact, legal users are of different types and often DTs are not appropriate for all these kind of users. If we think of a scholar expert in a specific field and legal concepts, he/she has a different approach of searching that does not necessarily require the support of a DT. More advanced users may find discipline-specific databases still useful in providing a better search performance. Instead, a student approaching a general topic for the first time, needs DTs to get a rough outline of the topic. However, centralized index-based DTs can obscure the complexity of the information retrieval process because of the immediacy in retrieving the resources proposed. Immediate access to full text may prompt to download a result in full-text just because it is available [26]; [27]. This also applies if we move on the types of libraries. A university library, if equipped with a DT will of-

fer to its users a useful service for legal research; whereas in specialized libraries which cover specific areas of law, the library catalog and the resources selected by librarians (databases, ejournals ...) could fully satisfy a legal expert.

Therefore, the exploitation of DTs is certainly only one phase of the legal research process, which is a complex task that is not resolved in the exploration phase of the resources, but requires the evaluation, interpretation and connection with other sources relevant to the study of law. The real solutions to complex legal issues and problems require analysis and thoughtful conclusions. Each research question has a different starting point, process, and conclusion. Of course, it includes false starts, dead ends, and revisions. Most important, legal research is never 'stopped' but the skilled researcher recognizes when to finish. As Felix Frankfurter stated [28]: "Research requires the poetic quality of the imagination that sees significance and relation where others are indifferent or find unrelatedness; the synthetic quality of fusing items theretofore in isolation; above all the prophetic quality of piercing the future, by knowing what questions to put and what directions to give to inquiry".

A successful researcher is one who understands how to use the many resources that are available in a flexible and efficient manner.

One last point concerns the difficulty to predict the future of DTs as libraries continue to struggle to find their path in the actual shifting environment of information provision. However librarians are the main actors for undertaking action and have a fundamental role to play in selecting, implementing, and evaluating the appropriate DTs for specific contexts, as well as in training users to exploit these tools effectively, helping them to interpret the results obtained. Linking these tools to the library computing environment is also a crucial point in their implementation and a prerequisite for their proper functioning. For sure, in the law domain information professionals are uniquely placed to shape and lead all necessary changes needed in bibliographic legal discovery.

References

[1] Wang, Y. & Mi, J. (2012). Searchability and Discoverability of Library Resources: Federated Search and Beyond. *College & Undergraduate Libraries, 19*(2-4), 229-245.

[2] Morgan, E. L. (2006). A "Next-generation" Library Catalog - Executive Summary. *LITA Blog: Library Information Technology Association*, 7 July 2006, http://litablog.org/2006/07/a-next-generation-library-catalog-executive-summary-part-1-of-5/.

[3] Nagy, A. (2011). Defining the Next-generation Catalog. *Library Technology Reports, 47*(7), 11-15.

[4] Yu, H. & Young, M. (2004). The Impact of Web Search Engines on Subject Searching in OPAC. *Information Technology and Libraries, 23*(4), 168-180.

[5] Coyle, K. (2007). The Future of Library Systems, Seen from the Past. *Journal of Academic Librarianship, 33*(1), 138-140.

[6] Gallacher, I. (2006). Forty-Two: The Hitchiker's Guide to Teaching Legal Research to the Google Generation. *Akron Law Review, 39*(1), https://ideaexchange.uakron.edu/akronlawreview/vol39/iss1/5.

[7] Gunter, B., Rowlands, I. & Nicholas, D. (2009). *The Google Generation. Are ICT Innovations Changing Information-seeking Behaviour?* Chandos Publishing, 299 ff.

[8] Breeding, M. (2015). *The Future of Library Resource Discovery*. NISO, http://groups.niso.org/apps/group_public/download.php/14487/future_library_resource_discovery.pdf.

[9] Breeding, M. (2010). *Next-gen Library Catalogs*. Neal-Schuman.

[10] Trombone, A. (2019). Formare e gestire collezioni con i discovery tools. *Biblioteche oggi, 39*, 13 and 18.

[11] Breeding, M. (2013). Next-generation Discovery: An Overview of the European Scene. In Chambers, S. (Ed.). *Catalogue 2.0. The Future of Library Catalogue*. Facet Publishing, 37-64.

[12] Van de Sompel, H. & Beit-Arie, O. (2001). Open Linking in the Scholarly Information Environment Using the OpenURL Framework. *D-Lib Magazine, 7*(3).
[13] Chisare, C., Fagan, J. C., Gaines, D. & Trocchia, M. (2017). Selecting Link Resolver and Knowledge Base Software: Implications of Interoperability. *Journal of Electronic Resources Librarianship, 29*(2), 93-106.
[14] Somerville, M. M. (2013). Digital Age Discoverability: A Collaborative Organizational Approach. *Serials Review, 39*(4), 234-239.
[15] Biagetti, M. T., Schaerf, M., Iacono, A. & Trombone, A. (2017). Verifica della disponibilità delle monografie attraverso i cataloghi delle biblioteche (VerDiMAC). *ANVUR Working Papers No. 4*, http://www.anvur.it/wp-content/uploads/2017/06/WPS%20201704-VerDiMAC.pdf.
[16] Yang, S. Q. & Li, L. (2015). *Emerging Technologies for Librarians: A Practical Approach to Innovation*. Chandos Publishing, 1st ed., 31-32.
[17] Breeding, M. (2010). The State of the Art in Library Discovery 2010. *Computers in Libraries, 30*(1), 31-35, http://www.librarytechnology.org/ltg-displaytext.pl?RC=14574.
[18] Gross, J. & Sheridan, L. (2011). Web Scale Discovery: The User Experience. *New Library World, 112*(5-6), 236-247, https://doi.org/10.1108/03074801111136275.
[19] Vaughan, J. (2008). Investigations into Library Web-Scale Discovery Services. *Information Technology and Libraries, 31*(1), 32-82, https://doi.org/10.6017/ital.v31i1.1916.
[20] Way, D. (2010). The Impact of Web-scale Discovery on the Use of a Library Collection. *Serials Review, 36*(4), 214-220, http://dx.doi.org/10.1016/j.serrev.2010.07.002.
[21] Hoy, M. B. (2012). An Introduction to Web Scale Discovery Systems. *Medical Reference Services Quarterly, 31*(3), 323-329.
[22] Machetti, C. (2016). Biblioteche e discovery tool: il caso OneSearch e l'ateneo di Siena. *AIB studi, 56*(3), 395.
[23] Johnson, M. (2013). Usability Test Results for Encore in an Academic Library. *Information Technology and Libraries, 32*(3), 59-85, http://ejournals.bc.edu/ojs/index.php/ital/article/view/4635.
[24] Howard, D. & Wiebrands, C. (2011). *Culture Shock: Librarians' Response to Web Scale Search*, http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=7208&context=ecuworks, 9.
[25] Marchitelli, A. (2014). Il catalogo connesso. *Biblioteche oggi, 32*(6), 13.
[26] Breitbach, W. (2012). Web-Scale Discovery: A Library of Babel? In Pagliero Popp, M. & Dallis, D. (Eds.). *Planning and Implementing Resource Discovery Tools in Academic Libraries*. IGI Global, 637-645.
[27] Guthrie, A. & Mccoy, R. (2014). A Glimpse at Discovery Tools within the HBCU Library Landscape. In Spencer, J. S. & Millson-Martula, C. (Eds.). *Discovery Tools: The Next Generation of Library Research*. Routledge/Taylor & Francis, 177-191.
[28] Frankfurter, F. (1930). The Conditions for, and the Aims and Methods of Legal Research. *Iowa Law Review, 15*, 124.