



Revista Interamericana de Bibliotecología
ISSN: 0120-0976
Escuela Interamericana de Bibliotecología

Gasparotto, Melissa
Using Google's Custom Search Engine Product to Discover
Scholarly Open Access and Cost-Free eBooks from Latin America
Revista Interamericana de Bibliotecología, vol. 41, no. 2, 2018, May-August, pp. 153-166
Escuela Interamericana de Bibliotecología

DOI: <https://doi.org/10.17533/udea.rib.v41n2a04>

Available in: <https://www.redalyc.org/articulo.oa?id=179056423004>

- How to cite
- Complete issue
- More information about this article
- Journal's webpage in [redalyc.org](https://www.redalyc.org)

The logo for Redalyc, featuring the word 'redalyc' in a stylized font with a red dot above the 'i' and a red line connecting the 'r' and 'd'.

Scientific Information System Redalyc
Network of Scientific Journals from Latin America and the Caribbean, Spain and
Portugal
Project academic non-profit, developed under the open access initiative

Using Google's Custom Search Engine Product to Discover Scholarly Open Access and Cost-Free eBooks from Latin America*

Abstract

Many Latin American scholarly monographs are available for free to read and download in a scattered fashion across the web, hosted on educational, institutional and government websites as well as commercial websites and publishing platforms. There is as of yet no single way to identify all of this content at once, but web-based discovery leveraging existing search engine indexing would seem to be a likely option. This case study suggests and evaluates one such method for discovery of open access and other cost-free scholarly monographs produced in Latin America. One possible configuration of Google's Custom Search Engine product is proposed and evaluated, and findings suggest its usefulness for a variety of applications, including for collection development, the preparation of thematic research guides with open content, and the enrichment of existing lists of open access eBook sources from Latin America. Unlike existing open access eBook portals, which search across known collections of such materials, search portals such as the one proposed allow users to search across the entire web to uncover scholarly free eBook sources that were previously unknown to them alongside known content sources, a key advantage to this method of discovery. The results further suggest the importance of pursuing discovery of these monograph titles outside established known collections, as an astonishing 45 % of all monographs identified through the Custom Search Engine portal were not discoverable in any edition, print or electronic, through WorldCat, and only 27 % were indexed by Google Books. Additionally, the low number of these eBook titles hosted in preservation-worthy repositories raises cause for concern about their long-term digital availability.

Keywords: Open Access, eBooks, scholarly monographs, search engines; Latin America.

Melissa Gasparotto

Assistant Director for Research Services
at The New York Public Library Research
Libraries. Master of Science in Library
and Information Science from Long Island
University and Master of Arts in Latin
American and Caribbean Studies from New
York University.

melissagasparotto@nypl.org

orcid.org/0000-0002-3048-0282

Cómo citar este artículo: Gasparotto, M. (2018). Using Google's Custom Search Engine Product to Discover Scholarly Open Access and Cost-Free eBooks from Latin America. *Revista Interamericana de Bibliotecología*, 41(2), 153-166. doi: 10.17533/udea.rib.v41n2a04

Recibido: 2017/08/06 / **Aceptado:** 2017/12/11

* This project was completed with support from the Aresty Research Center at Rutgers University. The author gratefully acknowledges research assistant Harry Nicolau for his help with data collection. .



Uso de la plataforma Google's Custom Search Engine para descubrir fuentes académicas de acceso abierto y libros electrónicos gratuitos de América Latina

Resumen

Muchas monografías académicas de América Latina están disponibles para leer y descargar gratuitamente, dispersas en la web, alojadas en sitios educativos, institucionales y gubernamentales, así como en sitios web comerciales y plataformas de publicación. Aún no existe una forma única de identificar todo este contenido a la vez, pero el descubrimiento basado en la web para potenciar el uso de la indexación de los motores de búsqueda existentes parece ser una alternativa. Este estudio de caso sugiere y evalúa uno de estos métodos para el descubrimiento de acceso abierto y otras monografías académicas gratuitas producidas en América Latina. En este trabajo, se propone y se evalúa una configuración posible de la plataforma Google's Custom Search Engine, y los hallazgos sugieren su utilidad para una variedad de aplicaciones, incluido el desarrollo de colecciones, la preparación de guías temáticas de investigación con contenido abierto, y el enriquecimiento de listas existentes de libros electrónicos de acceso abierto de América Latina. A diferencia de los portales de eBook de acceso abierto, que buscan en colecciones conocidas dichos materiales, las plataformas de búsqueda como la que se propone en este trabajo permiten a los usuarios buscar en toda la web para descubrir fuentes académicas gratuitas de libros electrónicos, que antes eran desconocidas, además de buscar en las fuentes de contenido conocidas, una ventaja clave para este método de descubrimiento. Los resultados sugieren, además, la importancia de buscar estos títulos monográficos fuera de las colecciones establecidas ya conocidas, pues un asombroso 45 % de todas las monografías identificadas a través de la plataforma Custom Search Engine no fueron detectables en ninguna edición, impresa o electrónica, a través de WorldCat, y solo 27 % fueron indexadas por Google Books. Además, el bajo número de estos libros electrónicos alojados en repositorios dignos de preservación plantea motivos de preocupación sobre su disponibilidad digital a largo plazo.

Palabras clave: acceso abierto, libros electrónicos, monografías académicas, buscadores, América Latina.

1. Introduction

The idea of open access monograph discovery and collection building in Latin American and Area Studies is hardly new. As digital library and cultural heritage institution repositories have grown in size and prominence, so too has the interest in pulling them together for ease of discovery alongside related collections. In Spain, the impressive Biblioteca Virtual Miguel de Cervantes includes dozens of thematic collections of full-text scholarly Latin American and Iberian content including articles, books, and recordings. The CLACSO and FLACSO digital libraries pull together open access scholarly monographs produced by multiple regional affiliates under a single portal. Europe-wide efforts include Europeana, which includes digital collections from dozens of libraries, museums, and other national institutions. HathiTrust contains a sizable subset of Latin American and Iberian content. And of course, the Directory of Open Access Books (DOAB) and the OAPEN Library are important aggregators of scholarly open access eBooks, although they remain small in size. All of these projects include known OA titles currently available in existing collections, begging the question: what unaggregated content available in digital form might be missing from such compilations?

A growing trend has drawn attention to an additional source of digital scholarly monographs in need of discovery: publications made available for free reading, and often for free download, produced by Latin American independent and university publishers, cultural heritage institutions and government agencies and secretariats. Because the number of publishers participating in these activities is large when viewed across the region, web-based discovery would appear to be a likely solution for these monographs. However, search engines generally lack the user-controlled refinements necessary to restrict results to quality open scholarly monograph output, and searching without these refinements results in too many unrelated websites to be useful for eBook discovery.

This paper proposes options for Latin American OA and cost-free¹ scholarly monograph discovery utilizing a Google product known as Custom Search Engine (CSE) to create customized search portals. Unlike the digital collections mentioned above, which search across discrete existing known collections, search portals and techniques such as those proposed allow users to search across the entire web to uncover scholarly free eBook sources that were previously unknown to them alongside known content sources, a key advantage to this method of discovery.

2. Library Collections and the Latin American eBook Landscape

While the availability of scholarly eBooks published in Latin America has been expanding steadily, there remains substantial room for improvement in their availability for library collection development. As with titles published by larger global eBook providers, Latin American eBooks may be available as part of subscription or purchased packages from vendors, as well as on an individual title basis from a variety of providers large and small.² Individual publishers, organizations and retail bookstores offer eBooks for sale through their own websites and third-party platforms, with a variety of DRM and DRM-free options. However, these eBooks pose challenges for libraries due to their spread-out nature, incompatibility with library digital platforms, and publisher inflexibility around licensing for institutional access, among other issues. Working with the more established eBook vendors is also a challenge. As Dracine Hodges (2015) notes, the four largest global e-content vendors, Elsevier, Wiley, Springer and Taylor & Francis, provide substantially more material in journal format than eBook format. Further, they all publish content that skews heavily toward the sciences, leaving “the rich cultural scholarship of the social sciences

and humanities in other world regions unaddressed” (p. 174). Lourdes Gutiérrez-Palacios (2012) also observes the science-heavy focus of major commercial eBook vendors, adding that these aggregators provide access to primarily English-language content (p. 9). Suzanne M. Ward, Robert S. Freeman and Judith M. Nixon (2015) concur, noting that these issues together pose a barrier to building global academic eBook collections (p. 4). For the purpose of Latin American eBook collection development, then, one major challenge is the lack of available content for acquisition through standard library-vendor routes.

Given the challenges associated with insufficient representation of global publishers in current eBook aggregators, librarians must investigate alternate means to provide access to the growing scholarly eBook output outside the global north. This should include a coordinated effort urging vendors to be more inclusive in their content development (see Terence Huwe’s 2017 article “The Long and Winding Road of Ebooks” for a good summary of library-vendor dialog around eBook access), but may simultaneously involve investigating all appropriate models of access over ownership.

3. Developing a Web Portal for Latin American OA and Cost-free eBooks

Given the limited distribution of Latin American printed editions in library holdings outside the region, and the previously mentioned variation in both OA and purchased eBook distribution models, the interest in developing a web portal to provide aggregated access to open eBook content from the region is clear. Print runs of quality scholarly monographs by Latin American university presses, government agencies, and institutes may be quite small, and distribution handled informally within the region at book fairs and events, especially in those countries with smaller publishing industries. The formal distribution of any OA surrogates that may exist is rarely attempted, although some efforts may be made via social media (Domínguez & Ovadia, 2011; Scott, 2016). These eBooks may often be found only on the website of the publisher, or hosted on free publishing platforms such as issuu.com by smaller publishers who lack the infrastructure to undertake

1 Cost-free is used here to distinguish those eBook titles made available for free without the related rights for reuse assigned to open access works. Cost-free eBooks may also lack preservation-worthy hosting and permanent access options that are frequently available for works designated as open access.

2 JSTOR has begun offering a wider selection of Spanish-language scholarly monographs in recent years, joining e-Libro and Digitalia, two early eBook vendors with representation of academic publishers from Spain and Latin America.

digital distribution on their own.³ Government-financed titles may be available for free on cultural ministry and non-profit organization websites, as well. For area studies collection development purposes, it can be challenging to identify a vendor who can secure the printed edition of small print run titles, making discovery of cost-free digital surrogates an enticing proposition. Additionally, some Latin American government-financed organizations and ministries are releasing books solely in a freely downloadable digital format.⁴ These titles are distributed outside existing models that might ensure their inclusion in WorldCat. While Google Scholar indexes many institutional repositories and has begun indexing OA and cost-free eBook titles on a limited basis, the majority of such book-length works fall outside its inclusion. Further, Google Scholar cannot be customized to deliver only books instead of articles.

The convenience factor of eBook access is a further rationale for increasing discovery of open Latin American monograph content through a web portal, coupled with a high potential for use by academics. A study by David Nolen (2014) found that, in a single year, nearly 38 % of all references to scholarly monographs in three top Latin American Literary Studies journals were to foreign university, academic, and governmental publications. This suggests that there may be a motivation for research libraries to enhance discovery of and access to the OA book production from these types of publishers.

Cost to libraries is another factor. As an ever larger percentage of library budgets goes to serials, the attractiveness of high quality OA monographs is clear. Additionally, expanded discoverability for freely available digital scholarly monographs can help ins-

titutions with smaller budgets more adequately serve patron needs: even a small library can give access to targeted collections of digital titles to supplement its collection. This is not to say that a search portal for OA and cost-free academic books can ever replace substantive budgetary commitments to specialized collections, particularly for regions of the world that remain print-heavy due to lack of a strong market for eBooks. However, research libraries can complement print holdings by enhancing discoverability of those electronic materials available at no cost.

Finally, digital preservation is of growing concern to librarians, and the spread-out nature of OA and cost-free monographs hosted in disparate repositories and websites only adds to this problem (Crossick, 2016). How can preservation be assured when discovery is not even possible? How can one even determine the scope of eBook material that may be at risk? A single search point for these kinds of materials would greatly aid efforts to quantify the nature and extent of the OA scholarly monograph preservation puzzle.

These challenges emphasize the importance of efficient discovery strategies for Latin American OA scholarly monographs. However, there is still no easy way to simultaneously search across all possible sources without being overwhelmed by out of scope content.

4. Using Google's Custom Search Engine Product for OA and Cost-Free Scholarly Monograph Discovery

Google released its Custom Search Engine (CSE) product in late 2006 (Hagner, 2006). The product allows developers to specify with relative granularity several search parameters targeting as much or as little of the entire indexed web as they define, making for a unique and highly tailored thematic search portal.

The use of a private sector no-cost product as a platform for any library tool has risks. Google has regularly made changes to and discontinued features of the CSE product, and can choose to cease support at any time. Therefore, other models should simultaneously be a focus of ongoing work in OA monograph discovery.

3 For example, Mexican independent publisher Ediciones La Rana and the Museo de la Palabra y la Imagen in El Salvador are among the many Latin American publishers distributing OA eBooks through the publishing platform issuu.com.

4 For example, this multimedia eBook on traditional textile designs published in 2017 by the Comisión Nacional para el Desarrollo de los Pueblos Indígenas in Mexico is solo disponible para consulta en línea (available only for online consultation): Santos-Concepción, L. (2016). Iconografía tének de origen ancestral. El significado de los bordados. Chihuahua, Mexico: Comisión Nacional para el Desarrollo de los Pueblos Indígenas. Recuperado de <http://www.cdi.gob.mx/iconografia-tenek/>. Accessed February 26 2017.

For example, a search engine built using open source software can still leverage Google's indexing, but it would remain free from many of the uncertainties associated with using products over which libraries have no control and which are subject to change without notice. This would constitute a massive effort by any organization choosing to undertake the work. However, it behooves academic institutions to give it some consideration sooner rather than later. If a not-for-profit does not enter this arena, one likelihood is that commercial entities may develop similar resources to enhance their own discovery products, potentially leading to an increase in their cost to libraries. Further, it can be argued that libraries and other academic entities have a mission to lead in open access rather than leaving such innovations to the private sector.

Despite these concerns about using Google CSE for a project of this nature, the problem of OA scholarly monograph discovery urgently needs attention. A proposed alternative is to use currently available tools as a starting strategy to discover OA and cost-free scholarly monographs, and to develop a clearer picture of current eBook distribution patterns.

Although CSE has been available for nearly ten years, a survey of the literature shows relatively limited use of the tool in libraries, and the number of articles documenting tips for utilizing the service outnumber those articles documenting existing projects and their development. *The Unabashed Librarian* (2007) was the very first library journal to include a list of the potential uses of the CSE product to enhance public service ("Impress patrons...", p. 25). The following year, three very brief articles appeared in *Computers in Libraries* suggesting practical tips for creating a CSE (Notess, 2008; Pretlow, 2008a; Pretlow, 2008b). Hennesy and Bowman (2008) documented the development of an arts research-related portal with CSE in a case study also published in *Computers in Libraries* in the same year. In 2011, Do Pazo-Oubiña, Calvo Pita, Puigventós-Latorre, Periañez-Párraga and Ventayol Bosch developed a pharmacological research personalized search engine which they described in *Farmacia Hospitalaria*, and the following year Schmick, Johnson, Scoville and Vaduvathirayan (2012) described their work creating a tailored search portal for foreign language health information in *Journal of Consumer Health on the Internet*. Throughout this

period, several additional applications of the service in libraries were developed, including DRAG-NET, a legal research database, and these are documented in Terry Ballard's 2012 book *Google This!: Putting Google and Other Social Media Sites to Work for Your Library*. Gavali (2015) published a case study in DESIDOC documenting the use of Google CSE to develop a discovery portal for engineering and technology research, and Giovanna Badia (2015) included three paragraphs on using Google CSE for search filtering in her chapter in the edited volume *The Complete Guide to Using Google in Libraries*. Perhaps the most visible use of a Google Custom Search Engine in libraries is Keely Wilczek's Think Tank Search hosted at the Harvard Kennedy School Library website, which allows users to search across the web publications of several hundred think tanks at once.

The applications above work by explicitly including a selected list of individual websites to search, rather than leveraging the custom features to search across the entire web and as Ballard (2016) notes in a article bemoaning the limited adoption of the product in libraries, some of these useful projects have already ceased to exist (33). The list above is almost certainly incomplete. However, the use of the CSE product in libraries does not appear to be widespread or long-term. No study or documented project could be identified that considered the use of CSE for open access eBook discovery, or for federated discovery of web-based literature published in specific countries or regions. One possible reason for its low adoption rate as a library tool may be the maintenance required to ensure searches continue to deliver adequate results: results delivered through any custom search engine will change slightly each time Google pushes an update to its ranking algorithm.

Additional scenarios in which a thoughtfully configured CSE could prove helpful, but which are not yet mentioned in the literature, include the development of textual corpora by digital humanists, and the identification of newly published books from certain countries on a particular theme for the purpose of collection development.

Outside libraries, one area where custom search engines have been developed with positive results is in pdf file discovery. Numerous pdf search engines exist on the web, and many of these utilize Google's CSE product

to good effect. However, these search engines include all pdf files regardless of document type, audience, or content. They do not attempt to discriminate between monographs published in pdf format and any other type of content (articles, slides, technical reports, etc.).

The Google Custom Search guide (Google LLC, 2017) provides detailed information on customization along with selected examples of how the product has been used, and points toward a way in which the service can be mobilized for open access eBook discovery. Custom search parameters include structured metadata fields permitting filtering by format, language, region, and date, among other criteria from schema.org. Searches can be limited to specific portions of a site by including one of 50 available URL pattern specifications, allowing for the deliberate inclusion of only the portion of a university press website that contains its eBook collections. The engine can be customized to search across a list of existing bookmarks using the Link Hub feature, eliminating the work of adding websites individually, meaning that those librarians who already maintain a list of OA eBook repositories can quickly create a federated search of those collections. Keywords can be appended to all searches in a sophisticated manner, leveraging the specialized search techniques of which so few users are aware, including Boolean operators and image search refinements, and keyword synonym search can be activated as well. These factors can each be assigned a different weight in the results ranking, by using custom XML annotation files.

One of the more useful customizations for researchers in any area of global studies is the ability to use wildcards in the initial CSE setup to specify a subset of the open web at the top-level domain such as *.mx, *.edu or *.edu.mx. This particular feature does not appear to have been utilized in any of the projects documented in the library literature. The use of this feature in combination with the other variables specified above enables more sophisticated searching for data gathering across a region.

5. Customizing a CSE for eBook Discovery

What constitutes a scholarly eBook to a search engine? What constitutes open access? How is Latin America defined in a digital environment? To give the CSE sufficient information about what types of results to deliver, machine-readable markers for these criteria were defined. There is no perfect way to accomplish this in the absence of a highly functional semantic search, however the following parameters are suggested as a starting point. Each of the identified parameters has significant drawbacks when used on its own. However, when used together they result in a custom search engine that discriminates with higher accuracy.

5.1. File Format

eBook file formats are multiple, and can include PDF, Epub, HTML and mobi, among others. Stating file formats for inclusion creates at least one complication, however, as a search engine configured in this way is incapable of identifying those files that are displayed on the publisher site, but hosted elsewhere. Of the list of file formats relevant to monographs, Google permits only the filtering of pdf and HTML files using the file format parameter, limiting its utility to a search. However, initial tests using pdf as a file format filter along with "epub" as a keyword automatically added to the searches performed on the CSE, helped to expand monograph discovery. File formats can also be specified for exclusion, for example to exclude non-book multimedia content.

5.2. Keywords

Any query already permitted by Google's web search may be automatically added to every search performed through a CSE. Adding the Spanish-language keywords "*editorial*" (publisher) and "*ediciones*" (editions) to user queries can help restrict results to files consisting of published monographs, although this may exclude cultural heritage organization and government ministry publications. Adding the keywords "eBook," "*e-libro*" or "*libro electrónico*" is an additional way to refine results from the open web, however this may exclude relevant results that do not use formal terminology to describe an eBook file hosted on the website.

Keyword exclusion is one method of eliminating eBooks available for sale from search results, for example by excluding the words “*precio*” (price), “*tienda*” (store), and their synonyms and translations into other Latin American languages. Excluding journals can partially be accomplished by excluding the keyword “*issn*” from all searches. Excluding words such as “*revista*” (magazine/journal) is not recommended, as this will eliminate monographs whose bibliographies cite articles published in journals with titles containing these words. Keyword exclusion techniques are extremely important in helping to remove journal content from searches. There is, unfortunately, no way to limit results to pdf files over a certain page length, which would make this the best alternate option.

5.3. Language

Restrictions by language, or setting a preference for one language over the others, is possible in CSE. However, these language preferences can only be set one at a time from the database editing interface, meaning that the CSE cannot easily simultaneously limit to Spanish, Portuguese, and multiple Latin American indigenous languages. This problem can be sidestepped if the custom XML annotations file is edited manually. However, coding in language restrictions may prevent discovery of English or other language monographs that meet all other criteria.

5.4. Website Domain

Country-level domain specifications are one way to limit to content hosted by regional websites. Top-level domains such as .edu and .org include whole swathes of the web with potential content of interest. However, because setting these domains as the only sites to search could inadvertently exclude relevant content hosted on non-regional domains, it is also possible that country, .org, .gov, .gob, and .edu domains simply be given added weight in the CSE custom xml file rather than defined as the only domains to be searched.

Domains may also be excluded. Excluding the .com domain is one way to begin eliminating commercial websites. However, that solution causes numerous additional problems. For example, this strategy will exclude some commercial publishers who also offer

cost-free items. Additionally, excluding .com eliminates all of the large commercial document repositories and publishing platforms that many academic presses and government ministries use to distribute their OA monographs: *issuu.com* is an important example. A partial solution may be to exclude the .com domain on a blanket basis and manually include individual .com sites known to be important as exceptions. This is not wholly thorough, but it may ensure higher quality results with fewer false positives.

Individual websites may also require explicit exclusion. For example, the large number of *Worldcat.org* and Google Books records for non-OA eBook content, coupled with the sites' high authority in Google's ranking algorithm make them likely candidates for exclusion. *Wikipedia.org* and *Academia.edu* are also good candidates for exclusion given the volume of out-of-scope article content hosted on these platforms.

5.5. Schema

Schema from *schema.org*, such as “Book” or “Creative Work” may be included in the CSE search configurations. A major drawback to using this structured metadata is that the convention is not widely adopted by Latin American webmasters and therefore a CSE using filters based on this metadata may be excluding a high number of relevant websites providing access to content of interest. Initial testing yielded very few search results when stipulating this parameter.

5.6. Creative Commons Status

Google CSE permits filtering by Creative Commons license type. Initial experimentation with this filter yielded very few search results, indicating limited implementation of this metadata convention by Latin American publishers.

6. Configuring the CSE

Using the above considerations as a starting point for performing the initial CSE configurations, a CSE was developed, tested, and assessed through a quality analysis of results. The final custom search engine for OA and cost-free Latin American eBooks can be acces-

sed at the following url: goo.gl/agah2k. The following parameters were specified:

- **File Format:** pdf
Query Addition: (pdf OR epub) AND libro -issn -abstract
 Search these sites: *.org *.edu *.gob *.ar *.bo *.br *.bz *.cl *.co *.cr *.cu *.do *.ec *.gt *.hm *.hn *.mx *.nl *.pe *.pr *.pt *.py *.sv *.uy *.ve
- **Domains and URL Patterns Excluded:**
 *.org/*artic*
 *.edu/*artic*
 *.org/*revista*
 *.edu/*revista*
 *.gob/*artic*
 *.gob/*revista*
 Academia.edu
 wikipedia.org

All of the Latin American country domains were individually specified for inclusion through the use of a wildcard symbol, as well as .org, .edu, and .gob. See Figure 1. This strategy sidesteps the problem with default CSE parameter options which permit only the prioritization of a single country domain at a time using a dropdown menu.

Sites selected for exclusion included large highly ranking sites such as Wikipedia.org and Academia.edu. URL patterns including journal article-related keywords were also selected for exclusion (see Figure 2).

The screenshot shows a search engine configuration interface. It has two main sections: 'Sites to search' and 'Sites to exclude'. Each section has a table with columns for 'Add', 'Delete', and 'Filter'. The 'Sites to search' table lists various country domain wildcards: *.bo, *.br, *.bz, *.cl, *.co, *.cr, *.cu, *.do, *.ec, *.gt, *.hm, *.hn, *.mx, *.nl, *.pe, *.pr, *.pt, *.py, *.sv, *.uy, *.ve. The 'Sites to exclude' table lists various URL patterns and domains: *.org/*artic*, *.edu/*artic*, *.org/*revista*, *.edu/*revista*, *.gob/*artic*, *.gob/*revista*, Academia.edu, and wikipedia.org. There are also search filters and a 'Search only included sites' checkbox.

Figures 1 and 2: Including and Excluding Domains and URL patterns.

Additional words were specified as automatic additions to user queries and these included words targeted for inclusion and exclusion to help identify monograph content, such as “libro”, “- issn” and “-abstract.”

Language was not specified in the configuration; rather it is assumed by using the Spanish-language keyword “libro” appended automatically to each query.

In addition to the above criteria, labels were created to permit the tabbed filtering of the full results list. The labels are Government, which prioritizes those results on .gob. and .gov websites, Academic and University, which displays primarily .edu results, and Organizations, which displays primarily .org results.

7. Assessing the Custom Search Engines

For the formal assessment, five searches were conducted in the custom search engine in February 2017, and the first 100 relevancy ranked results were evaluated based on scope, format and holdings criteria. The analysis is limited to the first 100 results as Google CSE does not display results beyond this number. Search keywords were chosen for broad relevance to scholarly and non-scholarly materials, to help determine how well the search engine configurations selected for scholarly materials.

7.1. Search Strings

In Search 1 the keyword string *arte feminista durante la dictadura* (feminist art during the dictatorship) was chosen to help identify scholarly works produced not only by university presses but also museums and cultural institutions, as well as popular feminist press eBooks.

In Search 2 the keyword string *cocina indígena* (indigenous cooking), was chosen because it could match on cookbooks and other non-scholarly materials as well as scholarly studies of cooking as it relates to history and culture. It was supposed that a search of this nature would help determine the accuracy of the search engine configuration in distinguishing between popular and scholarly monographs.

In Search 3 the keyword string *producción de quesos* (cheese production) was chosen to help determine the CSE's ability to provide access to technical monographs.

In Search 4 the keyword string *patrimonio natural México* (natural patrimony Mexico) was used to help

discover social science eBook content that overlaps thematically with governmental and non-governmental reports.

In Search 5 the keyword string *Rubén Darío crítica literaria* (Rubén Darío literary criticism) was chosen to help evaluate the CSE’s suitability for humanities-based eBook discovery (see Table 1).

Table 1: Search Strings

Search 1	Arte feminista durante la dictadura
Search 2	Cocina indígena
Search 3	Producción de quesos
Search 4	Patrimonio natural México
Search 5	Rubén Darío crítica literaria

8. Search Results

A list of free and open access eBook titles identified through test searches conducted through this CSE in February 2017 may be found at the following url: goo.gl/2h9zsi. Note that searches using the same keyword strings will not deliver the same eBook results, as newly available content coupled with multiple changes to Google’s algorithm since this date will have impacted the results.

Table 2: Item Types

	Scholarly OA or Cost- Free Monogra ph	OA or Cost- Free Book Chapter	OA or Cost- Free Scholarly Article	Whole journal	Dissertati ons and Theses	Commer cial eBooks for Sale	Other (legal statutes, curricular materials, presentatio ns, etc.)
Search 1 (n=100)	49 % (n=49)	2 % (n=2)	18 % (n=18)	2 % (n=2)	3 % (n=3)	2 % (n=2)	23 % (n=23)
Search 2 (n=100)	64 % (n=64)	3 % (n=3)	8 % (n=8)	1 % (n=1)	1 % (n=1)	2 % (n=2)	22 % (n=22)
Search 3 (n=100)	45 % (n=45)	6 % (n=6)	5 % (n=5)	0 % (n=0)	3 % (n=3)	0 % (n=0)	41 % (n=41)
Search 4 (n=100)	62 % (n=62)	3 % (n=3)	8 % (n=8)	0 % (n=0)	0 % (n=0)	0 % (n=0)	(n=27)
Search 5 (n=92)	29 % (n=27)	8 % (n=7)	29 % (n=27)	2 % (n=2)	3 % (n=3)	0 % (n=0)	28 % (n=26)
All Searches (n=247)	50 % (n=247)	4 % (n=21)	13 % (n=66)	1 % (n=5)	2 % (n=10)	4 (<1 %)	28 % (n=139)

8.1. Item Type

Overall, OA or cost-free monographs comprised 50 % of all results (247 out of the 500 results analyzed), however the percentage of OA and cost-free monographs delivered for each search string varied quite widely. Search 5 returned the smallest number of monographs at 35/100 or 29 % while Searches 2 and 4 each returned over 60/100 monographs, 64 % and 62 % respectively (see Table 2). Generally, searches delivered a higher proportion of monograph to non-monograph results in the first few pages of results when compared to later results pages where more out-of-scope content was delivered. For example, of the first 30 results (3 pages) of Search 1, a full 82 % of links directed to full text copies of in-scope monographs related to the search (n=25). This drops to 71 % of the first 50 (5 pages) of results (n=36), 59.5 % (n=42) of the first 70 results (7 pages) and 49 % (n=49) out of all 100 results. Although the number of in-scope scholarly monographs was fewer towards the end of the results list, the titles remained high quality and thematically relevant.

Out of scope results for all searches included primarily presentation slides, scholarly articles, newspaper articles, dissertations, and blog posts. Very few commercial publisher sites appeared in the results with eBooks for sale (1 % or 4 out of 500 results)(see Table 2). Nearly all out of scope items related to the theme of the search.

8.2. Date Range

The publication dates of all identified OA or cost-free monographs were overwhelmingly within the last 10 years (66 %, 164 out of 247 results). 80 titles (32 %) were published within the last 5 years. There was very little variation between searches in terms of the distribution of publication dates. The results demonstrate that the age of OA and cost-free eBooks discoverable through this CSE portal are well within the expected usable life of a scholarly monograph (see Table 3).

Table 3: Age Distribution of Monographs

Age of Monograph	Search 1 (n=49)	Search 2 (n=64)	Search 3 (n=45)	Search 4 (n=62)	Search 5 (n=27)	All Searches (n=247)
0-5 years old (2012-2016)	35 % (n=17)	36 % (n=23)	38 % (n=17)	29 % (n=18)	19 % (n=5)	32 % (n=80)
5-10 years old (2007-2011)	39 % (n=19)	28 % (n=18)	24 % (n=11)	52 % (n=32)	15 % (n=4)	34 % (n=84)
10-20 years old	12 % (n=6)	17 % (n=11)	20 % (n=9)	6 % (n=4)	19 % (n=5)	14 % (n=35)
Older than 20 years	8 % (n=4)	9 % (n=6)	4 % (n=2)	2 % (n=1)	33 % (n=9)	9 % (n=22)
Undated	6 % (n=3)	9 % (n=6)	20 % (n=5)	6 % (n=4)	0 % (n=0)	7 % (n=18)

8.3. Country of Publication

While publishers located in 23 countries were represented among all identified OA and cost-free monograph results, the overall distribution mapped broadly to print publication trends: inevitably Mexico, Chile, and Argentina were among the most well-represented publisher locations. When searches included a country name, the country distribution among those monographs delivered for the search invariably skewed toward that country. For example, in Search 4 (patrimonio natural México) a full 48 % of monographs were published in Mexico (see Table 4).

Table 4: Distribution of Country of Publication

Search 1	Search 2	Search 3	Search 4	Search 5	Search 1	Search 2	Search 3	Search 4	Search 5
Chile 32.7 % (n=16)	Mexico 28 % (n=18)	Mexico 22 % (n=10)	Mexico 48 % (n=30)	Nicaragua 19 % (n=5)	Dominican Rep. 2 % (n=1)	United States 3 % (n=2)	Bolivia 2 % (n=1)	Bolivia 2 % (n=1)	Honduras 4 % (n=1)
Argentina 30.6 % (n=15)	Chile 14 % (n=9)	Argentina 16 % (n=7)	Chile 13 % (n=8)	Argentina 11 % (n=3)	El Salvador 2 % (n=1)	England 1.5 % (n=1)	Brazil 2 % (n=1)	Brazil 2 % (n=1)	Mexico 4 % (n=1)
Spain 12.2 % (n=6)	Colombia 9.3 % (n=6)	Chile 13 % (n=6)	Argentina 8 % (n=5)	Uruguay 11 % (n=3)	El Salvador 1.5 % (n=1)	Guatemala 2 % (n=1)	Ecuador 2 % (n=1)	Peru 4 % (n=1)	
Mexico 6.1 % (n=3)	Ecuador 9.3 % (n=6)	Peru 11 % (n=5)	Peru 5 % (n=3)	Chile 7 % (n=2)	Nicaragua 1.5 % (n=1)	Uruguay 2 % (n=1)	Spain 2 % (n=1)	Spain 4 % (n=1)	
Uruguay 6.1 % (n=3)	Peru 9.3 % (n=6)	Ecuador 9 % (n=4)	France 3 % (n=2)	Cuba 7 % (n=2)	Panama 1.5 % (n=1)	Venezuela 2 % (n=1)	United States 2 % (n=1)	Venezuela 4 % (n=1)	
Peru 4.1 % (n=2)	Argentina 4.6 % (n=3)	Colombia 7 % (n=3)	Guatemala 3 % (n=2)	Colombia 4 % (n=1)	Spain 1.5 % (n=1)		Unknown 3 % (n=2)		
Bolivia 2 % (n=1)	Costa Rica 4.6 % (n=3)	Spain 7 % (n=3)	Switzerland 3 % (n=2)	Costa Rica 4 % (n=1)	Uruguay 1.5 % (n=1)				
Costa Rica 2 % (n=1)	Bolivia 3 % (n=2)	Costa Rica 4 % (n=2)	United Kingdom 3 % (n=2)	Dominican Republic 4 % (n=1)	Venezuela 1.5 % (n=1)				
					Unknown 3 % (n=2)				

8.4. Presence in WorldCat, HathiTrust, Google Books, and Other Repositories

Availability in WorldCat, HathiTrust and Google Books varied widely between searches (see Table 5). Technical literature discovered through Search 3 (*Producción de quesos*) had the least representation in these databases (27 %) and Search 5 (*Rubén Darío crítica literaria*) had the most (70 %). This makes sense when considering typical library acquisition patterns, as literary criticism is significantly more widely held than special interest agricultural industry resources.

Of the 247 total monograph results, 135 (55 %) were found in OCLC WorldCat in the corresponding printed edition. Of these results, only 19 WorldCat records

(18 %) included links to any open access or free copy of the eBook edition. Occasionally the OA copy cataloged in WorldCat was hosted at a different URL than that discovered through the search portal, and in some of those cases, the URL no longer functioned.

When the link was broken, it was not counted as an OA link in WorldCat.

Thirty-five titles (14 %) were available on a “search inside this title” basis in Google Books, and 17 (7 %) were available on a “search inside this title” basis inside HathiTrust. Thirty-one (13 %) were available in Google Books on a full view basis. None were available full view in HathiTrust (see Table 5).

Table 5: Presence in WorldCat, HathiTrust and Google Books

	Corresponding print edition in WorldCat	Not in WorldCat	Functional OA digital copy linked in WorldCat Record	Google Books (Full View)	Google Books (Search Inside View)	HathiTrust (Search Inside View)
Search 1 (n=49)	63.3 % (n=31)	39 % (n=19)	6.1 % (n=3)	0 % (n=0)	18.4 % (n=6)	8.2 % (n=4)
Search 2 (n=64)	53 % (n=34)	47 % (n=30)	9 % (n=6)	5 % (n=3)	5 % (n=3)	6 % (n=4)
Search 3 (n=45)	27 % (n=12)	73 % (n=33)	7 % (n=3)	18 % (n=8)	13 % (n=6)	0 % (n=0)
Search 4 (n=62)	62 % (n=39)	37 % (n=23)	1.1 % (n=7)	30.6 % (n=19)	19.3 % (n=12)	5 % (n=3)
Search 5 (n=27)	70 % (n=19)	30 % (n=8)	0 % (n=0)	4 % (n=1)	30 % (n=8)	26 % (n=7)
All Searches (n=247)	55 % (n=135)	45 % (n=113)	8 % (n=19)	13 % (n=31)	14 % (n=35)	7 % (n=18)

9. Discussion

The results from the searches in this case study are nothing more than a snapshot of the variety of open access and cost-free monographs published in Latin America. While it cannot be known with certainty what percentage of the OA and cost-free monograph world is represented by this snapshot, results from searches conducted using the portal suggest that a CSE can be a promising starting point for the discovery of OA and cost-free scholarly monographs. Configurations such as those employed can help filter out commercial and non-monograph content, particularly among those items appearing early in results lists. Titles are generally current and thematically relevant to the keyword searches, and an astonishing 45 % of all monographs

identified were not discoverable in any edition, print or electronic, through WorldCat. Even where searches had a low rate of return for monographs, those monographs proved of interest from a discovery point of view due to their sources. For example, although the search for *Rubén Darío crítica literaria* yielded few monograph results, the most well-represented country of publication was Nicaragua, giving high visibility to the open digital output of a country with a relatively small publishing industry. Searches such as this emphasize the usefulness of a web discovery tool to enhance access to materials representing a diversity of quality sources originating in Latin America. However, the low number of titles hosted in preservation-worthy repositories raises cause for concern about their long-term digital availability.

Thematically, the searches with the highest rate of return for eBooks were those academic topics of general public interest - for example, *cocina indígena* and *patri-monio natural México* both did better than a search for *producción de quesos*. Science-related queries did better than expected given the heavily article-driven nature of publications in those fields. The humanities search for literary criticism fared less successfully, mirroring the overall lower eBook production numbers in the humanities when compared to the sciences and social sciences.

One of the obvious drawbacks to searching the web for eBooks is that the accuracy of the results is determined by the quality of the metadata. The full text of books scanned without OCR is unsearchable, and the likelihood that they will be returned for relevant searches is low. On the other hand, quality books whose text is searchable may be returned regardless of low-quality metadata associated with the files. For example, a pdf file of a book may be titled simply “*Edición corregida*” (Corrected edition) or “*Libro pdf*” (pdf book) with no reference to the book title or topic, however as long as the full text of these books is readable to the search engine they can still be returned in the results. This may mean that a results list displaying only file names appears at a glance to be lower quality content. And although the quality of the titles is high and files were found to be hosted primarily on the publishing organization’s website, additional questions remain about the appropriate use of a discovery tool that cannot by itself distinguish between legally and illegally distributed online content.

10. Applications

Once it was established that the search portal could reliably return monographs in search results, applications were explored. These included the preparation of thematic research guides, generation of title lists for collection development, and enrichment of existing lists of OA eBook sources from Latin America.

10.1. Research Guides

Research guides heavy on cost-free content reflect a commitment to broad distribution and access. They can be useful for students and researchers around the world, and help ensure that materials produced in or

about the global south are made discoverable locally via the internet to the communities from whom the source materials come. The portal developed above was used to identify titles to feature in thematic research guides for students. One such guide now includes an entire page dedicated to sources of freely available books from Latin America, subdivided by country, with a single search point for all sources. Other research guides have been enriched with links to open eBooks throughout.

10.2. Collection Development

The search portal can also be used for more traditional collection building efforts. Titles discovered through thematic searching can be cross-referenced against library holdings. Additionally, titles selected as relevant for acquisition, if hosted in preservation-worthy repositories with stable URLs, can simply be cataloged locally as OA eBooks.

Through searches performed in the portal, it became apparent that tourism secretariats throughout the Americas are producing electronically what would in printed form be referred to as “coffee table books.” These high production value works are less frequently acquired by research libraries. However, their composition reflects contemporary images of national heritage and cultural patrimony as used by governments for the purpose of identity consolidation, marketing and economic development. It is an enticing prospect to provide access to these items cost-free without arranging for the international purchase, cataloging, and storage of what would otherwise be considered low-use and large format items.

Some eBooks identified through the portal contained copyright statements explicitly permitting the deposit in and distribution through a repository, and these items can be catalogued and preserved in an institutional repository without securing additional permissions. One such example is the book *Operación Cóndor: 40 años después* (Garzón-Real, 2016), published by the *Centro Internacional para la Promoción de los Derechos Humanos*, which includes the following notably permissive rights statement:

Todos los derechos reservados. Distribución gratuita. Prohibida su venta. Se permite la reproducción total o parcial de este libro, su almacenamiento en un sistema informático, su transmisión en cualquier forma, o por cualquier medio, electrónico, mecánico, fotocopia u otros métodos, con la previa autorización del Centro Internacional para la Promoción de los Derechos Humanos.

[All rights reserved. Free distribution. Sale is prohibited. The total or partial reproduction of this book, hosting on a computer system, transmission in any form through any media, electronic, mechanical, photocopy or other methods, is permitted by previous authorization of the Centro Internacional para la Promoción de los Derechos Humanos.]

As awareness around open access continues to grow, it is to be hoped that more monographs will be published with permissive copyright statements such as this, making those publications more readily available for acquisition, preservation, and use by libraries around the world.

Notably, the fact that the search portal continued to deliver some non-monograph content was helpful for the purposes of collection development. For example, many non-monograph pdf files delivered through the portal were still thematically relevant, sometimes consisting of Latin American book fair programs or book reviews. Resources like these can help librarians augment lists of titles for acquisition.

10.3. OA eBook Repository Lists

The portal and others like it can serve to further enrich existing lists of OA eBook sources. In searching the portal for collection development and research guide preparation, numerous repositories previously unknown to the author were uncovered. These included smaller scope online publishing efforts from individual cities, provincial/state and national agencies, and cultural organizations. Compiling publication sources for these items could be important work for libraries ensuring access to these potentially at-risk born-digital materials.

II. Conclusion

In this case study, a method is proposed for creating a web search portal to identify OA and cost-free Latin American monographs using Google's Custom Search Engine Product, and results show that this is a promising tool for discovery. Open eBook titles discovered through the CSE were generally current and thematically relevant to the keyword searches, and an astonishing 45 % of all monographs identified were not discoverable in any edition, print or electronic, through WorldCat. The broad geographic representation of eBooks delivered through the portal emphasize the usefulness of a web discovery tool to enhance access to materials representing a diversity of quality sources originating in Latin America.

Documenting searches through the portal has given a sense of the breadth and depth of quality scholarly Latin American OA and cost-free monographs available on the web. Several uses for a portal customized for OA monograph discovery are suggested, including user-friendly OA eBook discovery for patrons, research guide enrichment and collection development. Further, in demonstrating that a significant number of freely available eBooks from Latin America are unrepresented in world library holdings, the results emphasize the urgent need to consider open access scholarly monograph preservation and access.

12. Referencias

1. Badia, G. (2015). Googling for answers: gray literature sources and metrics in the sciences and engineering. In C. Smallwood (Ed.), *The complete guide to using Google in libraries*, Vol. 2 (pp. 61-68). New York: Rowman & Littlefield.
2. Ballard, T. (2012). *Google this!: Putting Google and other social media sites to work for your library*. Oxford, England: Chandos Publishing.
3. Ballard, T. (2016). Google's custom search. *Online Searcher*, 40(6), 30-33.
4. Crossick, G. (2016). Monographs and open access. *Insights: the UKSG journal*, 29(1), 14-19. doi:10.1629/uksg.280

5. Do Pazo-Oubina, F., Calvo Pitab, C., Puigventós-Latorrec, F., Periañez-Párraga, L., & Ventayol-Bosch, P. (2011). Desarrollo de un buscador de información farmacoterapéutica no publicada en revistas biomédicas. *Farmacia Hospitalaria*, 35(5), 254.e1-254.e5.
6. Domínguez, D., & Ovadia, S. (2011). Twitter: a collection Development discovery tool for and by the people. *Collection Management*, 36(3), 145-153. doi: 10.1080/01462679.2011.580427
7. Garzón-Real, B. (2016). *Operación Cóndor. 40 años después*. Buenos Aires: Centro Internacional para la Promoción de los Derechos Humanos (CIPDH). Recuperado de http://www.cipdh.gov.ar/wp-content/uploads/2015/11/Operacion_Condor.pdf
8. Gavali, R. (2015). Discovery service for engineering and technology literature through google custom search: A case study. *DESIDOC Journal of Library & Information Technology*, 35(6), 417-421.
9. Google LLC. (2017). *Google Custom Search*. <https://developers.google.com/custom-search/docs/overview>.
10. Gutiérrez-Palacios, L. (2012). *Gestión de la colección de libroselectrónicas en las bibliotecas universitarias*. Presentation delivered at Bibliotecas universitarias: nuevos tiempos, nuevas soluciones. Valladolid, Spain, September 20-21.
11. Hagner, K. (2006, October 24). Like Yahoo, Google adds customized search engine. *New York Times*, C3.
12. Hennesy, C., & Bowman, J. (2008). Curating the web: Building a Google custom search engine for the arts. *Computers in Libraries*, 28(5), 14-15.
13. Hodges, D. (2015). Developing a global e-book collection: An exploratory study. In S. M. Ward, R. S. Freeman, J. M. Nixon (Eds.), *E-books in academic libraries: Stepping up to the challenge* (pp. 171-191). West Lafayette: Purdue University Press.
14. Huwe, T. K. (2017). The long and winding road of ebooks. *Computers in Libraries*, 37(5), 9-11.
15. Impress patrons with your own Google search engines for them. (2007). *Unabashed Librarian*, 143.
16. Nolen, D. S. (2014). Publication and Language Trends of References in Spanish and Latin American Literature. *College & Research Libraries*, 75(1), 34-50.
17. Notess, G. R. (2008). Custom search engines: Tools & tips. *Computers in Libraries*, 28(5), 16-17.
18. Pretlow, C. (2008a). 10 web tools to create user-friendly sites. *Computers in Libraries*, 28(6), 14-17.
19. Petlow, C. (2008b). Custom searching. *Computers in Libraries*, 28(6), 16.
20. Santos-Concepción, L. (2016). *Iconografía tének de origen ancestral. El significado de los bordados*. Chihuahua, México: Comisión Nacional para el Desarrollo de los Pueblos Indígenas. Recuperado de <http://www.cdi.gob.mx/iconografia-tenek/>
21. Schmick, D. D., Johnson, E. D., Scoville C. L., & Vaduvathiriyani, P. K. (2012). Building a Google™ custom search engine (CSE) for foreign language health information: One library's effort to create a new tool for health professionals. *Journal of Consumer Health on the Internet*, 16(1), 27-36. doi:10.1080/15398285.2011.646590
22. Scott, M. (2016). *The Newest novedades: Using social media for collection development*. Presentation delivered at the 61st Seminar on the Acquisition of Latin American Library Materials, University of Virginia, Charlottesville.
23. Ward, S. M., Freeman, R. S., & Nixon, J. M. (2015). Introduction to academic e-books. In S. M. Ward, R. S. Freeman, J. M. Nixon (Eds.), *E-Books in Academic Libraries: Stepping up to the Challenge* (pp. 1-16). West Lafayette: Purdue University Press.